EPIDEMIOLOGIC AND EPIGENETIC BIOMARKERS OF ORAL HUMAN PAPILLOMAVIRUS INFECTION AND OROPHARYNGEAL CANCER

UNIVERSITY OF BIRMINGHAM

OLADEJO OLALEYE

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Institute of Head and Neck Studies and Education
Institute of Genomics and Cancer Sciences
School of Medical and Dental Sciences
University of Birmingham

September 2017

ABSTRACT

Introduction:

High-risk Human Papillomaviruses (HPVs) cause a subset of Oropharyngeal cancer (OPCs), with a rising incidence of HPV-OPC in the developed world. This thesis explores the epidemiology and epigenetics of oral HPV infection and OPC.

Methods and Results:

Systematic Review of 4,365 abstracts and ***Meta-analysis*** of 45 studies: Pooled prevalence for any oral HPV infection was 4.9% (95% CI: 3.7% - 6.3%) and high-risk HPV 1.8% (95% CI: 1.2% - 2.6%), with no difference by gender but variability by geography and sample type.

Prospective Cohort Study in 101 individuals: Oral HPV in oral rinse reduced from 11% before tonsillectomy to 2% afterwards. However, 10% had new infections.

Retrospective Review of 28,846 OPC on HES database: Survival is worse with increasing age, poor socio-economic status and multiple co-morbidities ($p<0.05$).

Genomics and Epigenetics Study on 40 FFPE samples from HPV-OPC: 40% had viral E2 gene disruption detected by PCR, with no differences in methylation at E2-binding sites (E2BS).

Conclusions:

Oral HPV prevalence is low and tonsillectomy may affect its natural history. 60% HPV-OPC have an intact E2 gene with no differences in E2BS methylation. Survival from OPC depends on age, co-morbidities and socio-economic status.

Word count: 192

DEDICATION

''There is a Spirit in man and the inspiration of the Almighty giveth them understanding''

Job 32: 8 KJV

I dedicate this PhD to God Almighty Who gave me His Grace, Wisdom and Strength

in abundance to both begin and complete this amazing journey.

I dedicate this work also to the loving memory of my PhD Co-Supervisor

Prof Ciaran Woodman

The Research on the 'Epigenetics of Human Papillomavirus in Oropharyngeal cancer'

was his idea and I benefitted immeasurably from his insights.

His Legacy lives on.

May his soul rest in perfect peace.

Amen.

ACKNOWLEDGEMENTS

To My Wife – Dr Jaiyeola Adeola Olaleye, I am deeply grateful for your unending support and constant encouragement throughout this academic pursuit. Your love and belief in me made me reach for the stars. Thank you for being there always for both myself and our adorable princesses – Ruby and Lily.

My appreciation to the generous Charities and Institutes that provided the financial support I needed to complete this research. In particular, Prof Mehanna and InHANSE for the first year funding and laboratory support throughout, Mr John Watkinson for his generosity and the Get-A-Head Charity for providing funding and a salary for my second year, the Midlands Institute of Otolaryngology for a research bursary for consumables, and the Institute of Translational Medicine, University of Birmingham for the Fellowship in the final year that paid my salary.

To My Mum – Mrs Durojaiye Olaleye, Your courage in seeking the best for me and my four siblings as a single parent is the single reason why I can stand amongst giants today. Words cannot fully express my gratitude but this achievement is testimony to your unwavering faith in us.

To My Incredible Siblings – Olanrewaju, Oladoyin, Temitope and Olabisi, You guys rock!. Thank you for your encouragement, phone calls and prayers. We all did it together.

To Dr Sarah Leonard, You were my guardian angel. Thank you for finding the strength to pick me up and teach me the depths of laboratory research after Prof Woodman left us unexpectedly. Your gentleness, clarity of thought and purpose was extraordinary. You have given me firm foundations in research and I will always remember your sacrifice.

To the InHANSE crew, You are simply the best!. My appreciation to all of you for the joyful, enabling environment and the incredible team spirit. Thank you for making my studies memorable. Gemma Jones, Nicki Graham, June Jones, Alison Pease, HRH Margaret Hartley, Baksho Kaur, Gordon Ryan, Dr Jill Brooks, Dr Rachel Spruce, Dr Nikos Batis, Dr Jennifer Bryant, Alison Wardle, Dr Paul Nankivell, Dr Neil Sharma, Dr Anthony Kong, Dr Davy Rapozo, Dr Peter Rae, Dr Sandra von Ziedler, Dr Jim Higginson and every student. I enjoyed every moment working with you all!. Thank you

To the HPV Supergroups led by Dr Sally Roberts and Dr Joanna Parish, I absolutely enjoyed working alongside all of you. I learnt incredible science from the many organised teaching sessions. Thank you Dr Ieisha Pentland, Mr Paul McCormack and Dr Karen Campos for helping me navigate the laboratory.

To my incredible friends who encouraged me every step of the way (Dr Oyewole Salako, Dr Oluwasegun Akilapa, Dr Udeme Ekrikpo, Dr Kunle Lawal and Dr Kunle Alonge):
I appreciate you all.

To the Systematic Review and Meta-analysis crew – Ms Jayne Wilson, Mr Piers Gaunt and Dr Esmond Carr, I am truly grateful for your help over months of screening and analysing data from numerous abstracts on oral HPV infection.

To all my trainers and colleagues on the West Midlands Otolaryngology, Head and Neck Training Programme – Thank you for your unwavering support throughout my time out of surgical training for research.

To My Supervisors:

Prof Hisham Mehanna, You have been an inspiration to me as an academic head and neck surgeon. It has been my privilege to witness your focus, determination, energy and desire for excellence. Thank you for guiding me throughout this period of study.

Dr Gosia Wiench, Your approachable style and insight was most reassuring. Thank you for helping me to excel and ensuring I completed the epigenetics project.

Prof Peter Sasieni, Your voice was a compass in navigating the epidemiologic projects. Thank you for every direction and insight. The wealth of experience you shared greatly improved my work.

CONTENTS LISTINGS **Page No**

LIST OF ILLUSTRATIONS: **Page No**

LIST OF TABLES:

LIST OF DEFINITIONS AND/OR ABBREVIATIONS

AMSTAR	Assessment of Multiple Systematic Reviews
APOT	Amplification of Papillomavirus Oncogene Transcripts
BM	Bisulfite Modification
CIN	Cervical Intraepithelial Neoplasia
DEIA	DNA Enzyme ImmunoAssay
DIPS-PCR	Detection of Integrated Papillomavirus Sequences PCR
DNA	Deoxyribonucleic acid
E2BS	E2 Binding Site
FFPE	Formalin Fixed Paraffin Embedded
HES	Hospital Episodes Statistics
HNSCC	Head and Neck Squamous Cell Carcinoma
HPV	Human Papillomavirus
IARC	International Agency for Research into Cancer
ISH	In-situ Hybridization
LiPA	Line Probe Assay
MPTS	Multiplex Type specific System
NATSAL	National Survey of Sexual Attitudes and Lifestyles
NCIN	National Cancer Intelligence Network
OPC	Oropharyngeal cancer
p16 IHC	p16 Immunohistochemistry
PCR	Polymerase Chain Reaction
RNA	Ribonucleic acid
RNA-Seq	RNA Sequencing
SCC	Squamous Cell Carcinoma

EPIDEMIOLOGIC AND EPIGENETIC BIOMARKERS OF ORAL HUMAN PAPILLOMAVIRUS INFECTION AND OROPHARYNGEAL CANCER

PhD CHAPTER 1:

INTRODUCTION and BACKGROUND

1.1 Background and Epidemiology:

1.1.1 Increasing Incidence of Oropharyngeal cancer:

There is an increasing incidence of Oropharyngeal cancer (OPC) in the UK (Shaw et al., 2011), with similar trends reported across much of the developed world and America (Chaturvedi et al., 2013). The National Cancer Intelligence Network in the UK reported the direct standardised incidence rates of OPC in the UK which has doubled from 1.2 per 100,000 population in 1996 to 2.3 per 100,000 in 2006 (NCIN 2010). This increasing incidence of oropharyngeal cancer as demonstrated in Figure 1.1, is contrary to the declining incidence of some head and neck squamous cell carcinomas (HNSCC) such as laryngeal cancers that have the traditional associated risk factors of cigarette smoking and alcohol consumption.

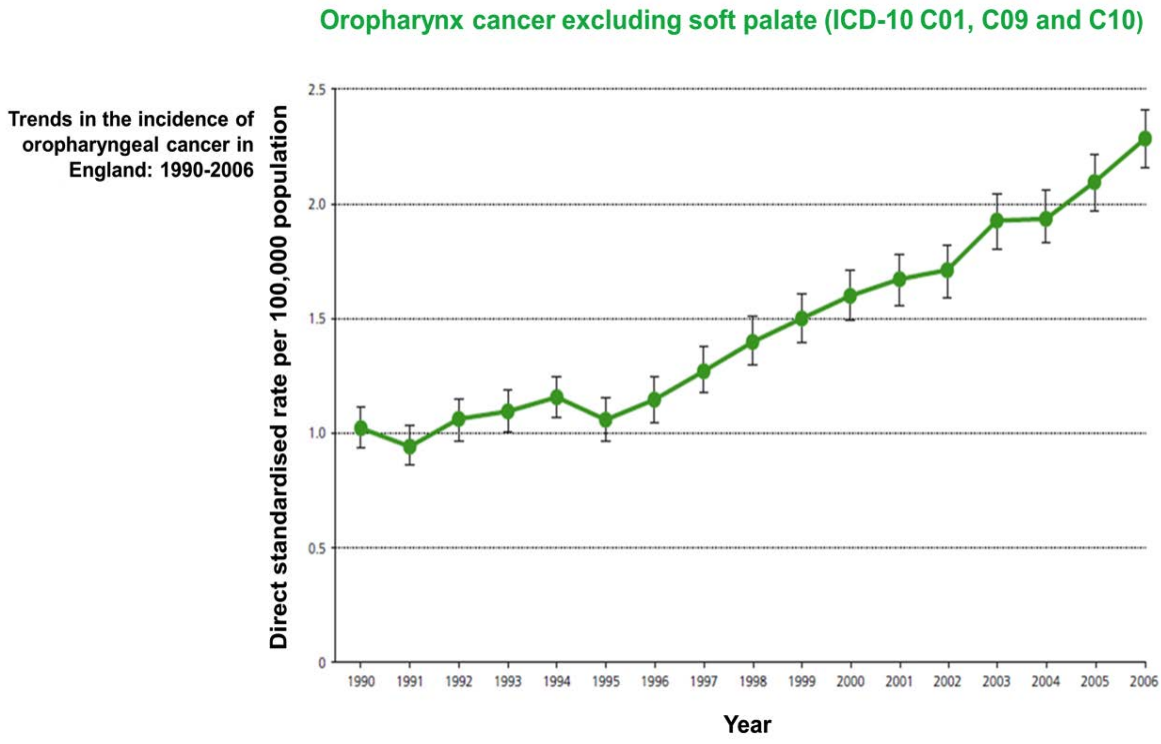


Figure 1.1: Trends in oropharyngeal cancer (OPC) incidence in England (NCIN 2010)

There has been an increasing incidence of OPCs in England in the study period.

This decline in laryngeal cancers and smoking-related cancers such as lung cancer, coincides with the steady decline in smoking rates across England in the last 30 years (ONS 2016). There is therefore an aetiologic driving factor separate to cigarette smoking accounting for the rise in OPC rates.

1.1.2 Human Papillomavirus-associated Oropharyngeal cancer:

The increasing incidence of oropharyngeal cancer across America, Europe, Scandinavia and the developed world has been described as an epidemic (Sturgis and Ang 2011). Oncogenic human papillomavirus (HPV) causes cervical cancer as well as a subset of anal, penile, vaginal and vulvar cancers (zur Hausen 2009, Graham 2010). These are non-enveloped DNA viruses with over 100 HPV strains described however the high-risk HPV genotypes 16 and 18 are responsible for the vast majority of these cancers in humans (Graham 2010). Recent strong epidemiologic and laboratory evidence has shown oncogenic HPV (in particular HPV 16) as an aetiological agent in the causation of a sub-set of Oropharyngeal cancer (OPC) (D'Souza et al., 2007b; Wiest et al., 2002) and is responsible for over 90% of HPV-associated oropharyngeal cancer (Gillison et al., 2000; Schache et al., 2011). A meta-analysis on head and neck cancers indicates strongly that HPV is aetiology in the oropharynx, less so in the oral cavity and no significant causal correlation in the larynx (Hobbs et al., 2006). There is a disparity in the strength of association observed between HPV in oral cavity cancers and in oropharyngeal cancer (Javadi et al., 2017).

Infection with high-risk genotypes of oral human papillomavirus (in particular HPV 16 and 18) therefore partly explains the rising incidence trend in oropharyngeal cancer. A recent meta-analysis showed HPV-OPC has risen from 40.5% to 72.2% in the 10 years spanning

2000 – 2009, with a significantly more rapid increase observed in Europe compared with North America (Mehanna et al., 2013).

Compared with HPV negative OPC, HPV-associated OPC tends to affect a younger demographic who may not have been exposed to cigarette smoking or alcohol consumption. HPV-OPC also appears to respond better to current treatments including chemo-radiotherapy and surgery with better survival outcomes (Fakhry et al., 2008, Shah et al., 2009, Reimers et al., 2007, Weinberger et al., 2006, Rischin et al., 2010, Mellin et al., 2002, Lindel et al., 2001, Licitra et al., 2006). Epidemiologic studies have confirmed risk factors for oral HPV infection to include oral sex and multiple lifetime sexual partners (D'Souza et al., 2009).

HPV status alongside nodal stage and smoking status is now employed as a prognostic classifier in patients with OPC in work done by (Ang et al., 2010), stratifying them into low, intermediate or high risk. HPV-OPC has a 3 year survival benefit following treatment compared with age and sex matched negative OPC (Ang et al., 2010). There are currently large multicentre UK trials on-going with a focus on HPV in OPC, disease progression and modifications to treatment modalities based on HPV status.

1.1.3 Survival from Oropharyngeal cancer:

Disease-free survival following a diagnosis of oropharyngeal cancer is dependent on a number of factors including patient factors, degree of exposure to risk factors, socio-economic factors, tumour factors and presence or absence of transcriptionally active human papillomavirus in the tumour. Survival worsens with increasing age and increasing levels of deprivation. Tobacco smoking and heavy alcohol intake are associated with worse survival outcomes with on-going exposure to these risk factors predisposing patients to higher cancer recurrence rates (Paleri et al., 2016). The toxicity from chemotherapeutic agents is not to be

underestimated especially in these patients with sub-optimal cardiorespiratory reserves. The presence of significant co-morbidities and the complications of surgical treatments and prolonged rehabilitation also take their toll on both the quality of life experienced as well as survival.

This alarming upward trend in OPC incidence in the UK has continued over the last decade and is now projected to be 30 times the current burden by the year 2025 (Louie et al., 2015). This represents a significant public health concern with huge implications for early diagnosis, treatment, resource allocation and long-term rehabilitation efforts which will come at a huge cost (Chesson et al., 2012; Schache et al., 2016).

The key concepts and current literature underlying this thesis will now be discussed including oral cavity / oropharynx anatomy, HPV biology, oral HPV infection and risk factors, oral HPV prevalence and persistence, oropharyngeal cancer, HPV 16 genome and epigenetic regulation.

1.2 Anatomy of the Oropharynx:

The Oropharynx is the anatomic area at the back of the throat and is visible on mouth opening. It consists of the palatine tonsils, lingual tonsils (or base of tongue), soft palate, uvula, posterior and lateral pharyngeal walls (Figures 1.2 and 1.3). Embryologic development of the tonsils occurs during the third month in-utero. Formation is from the lining of the endoderm, mesoderm of the second pharyngeal membrane, second pharyngeal pouch and first two pharyngeal arches. The second pharyngeal pouch epithelium form endodermal buds that develop outwards into the mesoderm. When the core cells in the buds regress, they convert into the tonsillar crypts that become impregnated with lymphoid tissue. Anatomically, the tonsils are bounded anteriorly by the palatoglossus muscles and palatopharyngeus muscles

posteriorly. Superiorly tonsils are related to the lateral aspect of the soft palate whilst inferiorly they merge with the lingual tonsils (tongue base). Blood supply is mainly from the tonsillar artery which is a branch of the facial artery and enters the tonsils from their inferior poles. Other arterial supply is from the dorsal lingual artery, ascending palatine artery, ascending pharyngeal artery, and the lesser palatine artery. Venous drainage is from the peritonsillar venous plexus into the lingual and pharyngeal veins which empty into the internal jugular veins. Nerve supply is from the lesser palatine nerves (branches of the maxillary division of the trigeminal nerves) and tonsillar branches of the glossopharyngeal nerves bilaterally. The oropharynx extends from the hard palate superiorly to the hyoid bone inferiorly, and is divided into anterior wall (tongue base posterior to foramen cecum, vallecular, lingual surface of epiglottis; bounded by the pharyngoepiglottic folds), a lateral wall (palatoglossus, palatopharyngeus, and pharyngeal palatine tonsil), a roof (oral surface of the soft palate), a posterior wall (superior and middle constrictors and buccopharyngeal fascia), and the tongue base (genioglossus muscle) (Homer and Rees 2012).

The palatine tonsils and lingual tonsils (tongue base) are lymphoid tissue in the oropharynx that form the *Waldeyer's ring* alongside adenoidal tissue in the nasopharynx. These play a more prominent role in humoral immunity in childhood but tend to regress both in their relative size and immune function as individuals become adults. Tonsils play a role in providing defence systems against oral pathogens, including bacterial and viral. The tonsils consist of troughs and crests with non-keratinizing stratified squamous epithelium overlying the superficial layer or crests (McClory et al., 2012) and a specialized reticulated epithelium overlying the tonsil crypts or troughs.

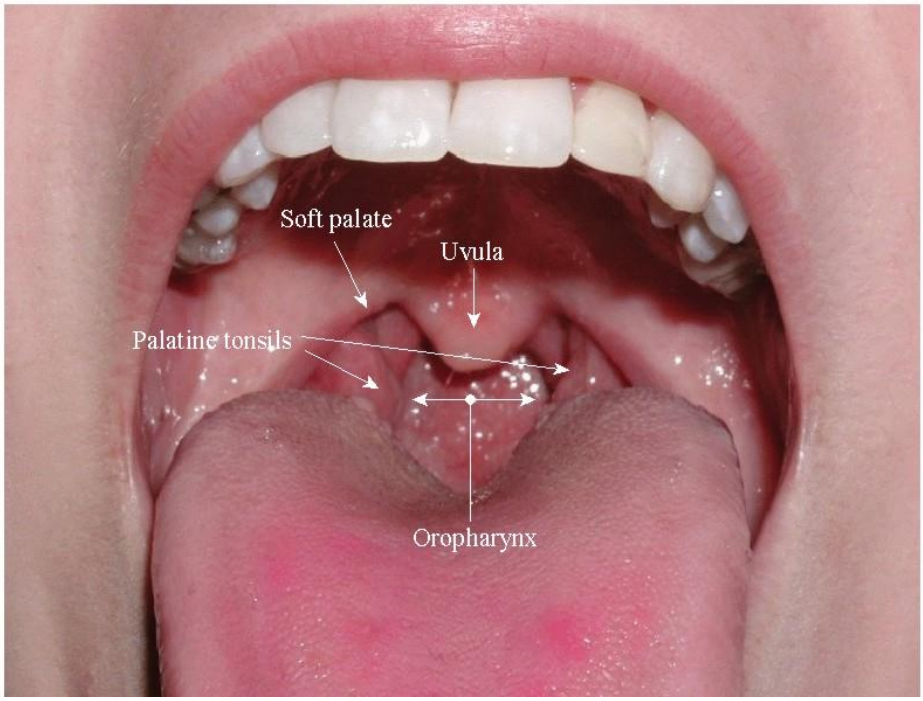


Figure 1.2: Anatomy of the oropharynx (Rayur.com 2012 / image thanks to uc.edu)
The oropharynx is demonstrated with subsites soft palate, uvula, palatine tonsils and posterior pharyngeal wall shown.

Tonsils are infiltrated with lymphocytes and have germinal centres which produce immune cells called memory B cells and an antibody called Immunoglobulin A. T-lymphocyctes or T-cells which are largely produced in the thymus gland and are essential components for immune functioning, have also recently been shown to be produced in the tonsils (McClory et al., 2012). In addition, cytokine-producing cells have also been identified in tonsillar tissue including T-helper 1 and T-helper 2 cytokines and cytokine messenger RNA. The specialized surface of tonsils contains M cells that capture antigens and subsequently trigger a stimulatory immune reaction that activates the B cells and T-lympocytes (Kato et al., 2013).

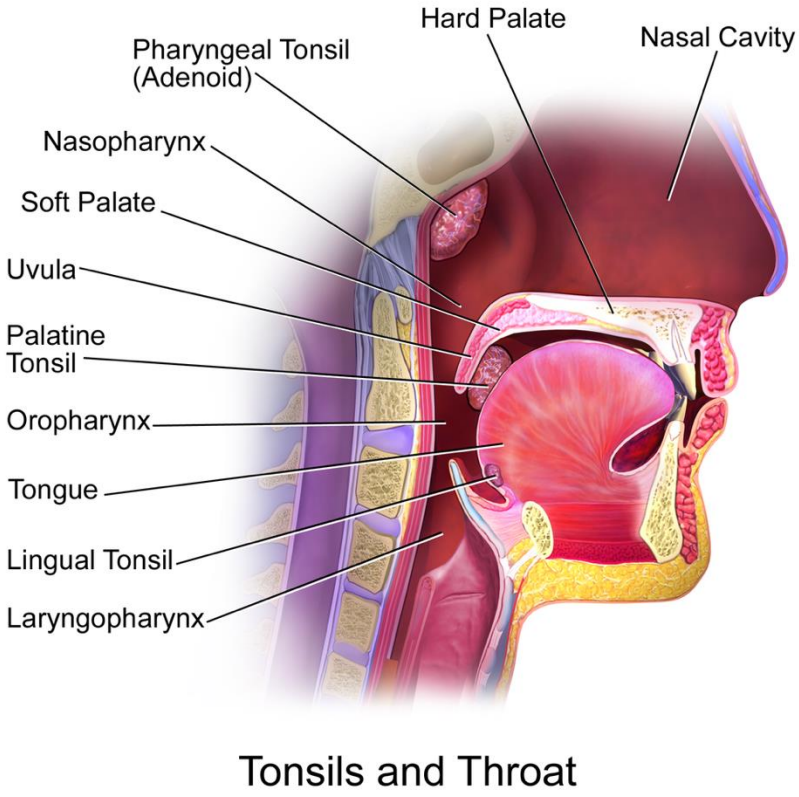


Figure 1.3: Anatomic relations of Oropharynx (sagittal view) Blausen.com staff (2014)

Sagittal section of the face showing the oropharynx along with its boundaries.

1.3 Human papillomavirus biology:

HPV is a common cause of sexually transmitted infections, many of which are transient. The vast majority of the population are exposed to and acquire infection with HPV at some stage in their lives (Baseman and Koutsky, 2005). Over 120 HPV genotypes have been described with mucosal HPV accounting for 40% of the genera (Fernandes et al., 2013) and cause a spectrum of diseases ranging from benign lesions to malignancy. HPV can be stratified depending on their transformative potential for malignancy. Low-risk such as HPV 6 or 11 are responsible for anogenital warts which are benign lesions while high risk such as HPV 16 or 18 have been implicated in cervical, anogenital and oropharyngeal cancer.

HPV is a non-enveloped, double-stranded DNA virus protected by a capsid. The major capsid protein is L1 is present in 72 copies in each virus while L2 protein occurs in variable copies (Longworth and Laimins, 2004; zur Hausen, 2009). Each HPV contains structural proteins L1 and L2 as well as E1, E2, E4 – E8. The structural proteins help viral particle assembly while viral DNA replication and control of transcription is undertaken by E1, E2, E4, E5 and E8 genes. E6 and E7 are involved in maintaining the competence of replication (Bodily and Laimins, 2011). The E4 protein has been shown to be an important regulator of the HPV lifecycle with diverse actions that involve cell organisation, host cellular DNA suppression and cell cycle dysregulation (Roberts et al., 2008). The HPV genome is passed to daughter cells during mitosis and it is able to do this by hijacking the natural cellular mechanisms to maintain stability. In particular, the HPV E2 protein targets several cellular proteins during mitosis to facilitate transfer of genomes to chromosomes ensuring persistence (Feeney and Parish, 2009).

In the cervix, HPV infects the basal undifferentiated cells following trauma. The virus then replicates in tandem with the basal cells. Most high risk HPV infections are cleared by the immune system within 12 – 18 months but sometimes persistent infection occurs (Kero et al., 2012). In 2008, Harald zur Hausen was awarded the Nobel Prize in Physiology or Medicine for his discovery of the causative role of HPV in cervical cancer. Incredibly, the Nobel Prize that year was shared with Françoise Barré-Sinoussi and Luc Montagnier who discovered the human immunodeficiency virus. Today we now know that HPV infection, persistence and carcinogenesis in the cervix, anogenital region or oropharynx are all increased in the presence of immunodeficiency from HIV.

1.4 Oral HPV Lifecycle:

Oral mucosa consists of epithelium and connective tissue separated by basement membrane. Epithelium can be keratinised or non-keratinised. The keratinised type has basal, spinous, granular and squamous layers. The non-keratinized epithelium can be divided into a superficial layer, an intermediate layer and a basal layer. In the cervical epithelium, micro-abrasions or trauma allows human papillomavirus access to the basal layer of cells supported by the basement membrane (Woodman et al., 2007; Doorbar et al., 2006). HPV begins to replicate its viral genome in the basal layer and progresses with exponential multiplication towards the superficial squamous layer where eventual viral release occurs. Once HPV infection has been established, early viral genes are produced (E1, E2, E4, E5, E6 and E7) until late viral genes (L1 and L2) are expressed in the superficial layers with capsid formation (Woodman et al., 2007).

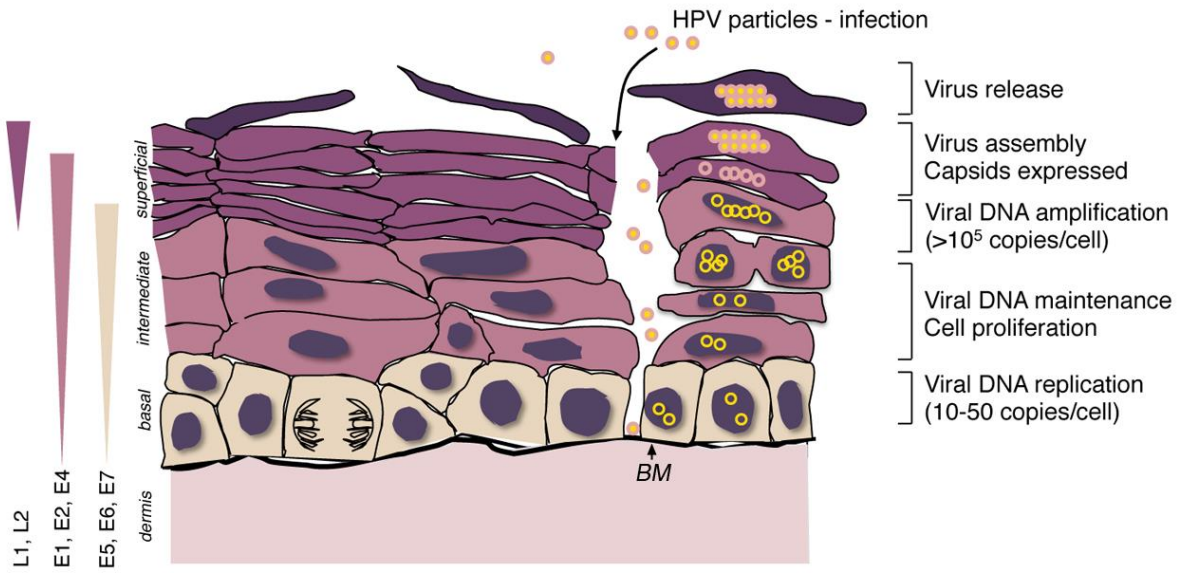


Figure 1.4: HPV Lifecyle in skin (Roberts 2015 – used with permission)
HPV lifecycle in host cells depicted with viral DNA replication commencing in the basal cells with exponential replication and amplification to >10^{5}copies/cell, virus assembly and then release of viral particles at the superficial layer.

The HPV replication machinery is a complex system that is dependent on the replication of the host cell. Viral DNA replication progresses from 10 – 50 copies per basal layer cell, through a proliferative phase and subsequent amplification to $>10^5$ copies / cell in the intermediate layer (Figure 1.4; from Roberts 2015). Virus assembly is completed in the superficial layer with expression of capsids. This phase is modulated by the late viral genes L1 (major capsid protein) and L2 (minor capsid protein). Ultimately release of HPV occurs until the infection clears. The described HPV lifecycle has largely been derived from observations in the cervix. In the tonsils however, HPV has been shown to localise in the specialized epithelium of the tonsillar crypts (Begum et al., 2005). Oral sex has been shown to be an independent risk factor in the acquisition of both oral HPV infection and in case control studies on HPV associated oropharyngeal cancer (Gillison et al., 2012). It is unclear if the theory of micro-abrasions preceding HPV localisation in the basal cell layer holds true in the tonsils. The cells in the oral cavity and oropharynx are rapidly dividing cells and often exposed to various shades of trauma daily from such activities as tooth brushing, eating, and deep kissing, which could precipitate micro-abrasions. The lymphoid nature of the tonsils also means that micro-organisms such as viruses and bacteria get trapped and elicit inflammatory reactions in such established conditions as tonsillitis and glandular fever. Digital transmission of HPV into the oral cavity is also an infrequent route of transmission.

1.5 HPV Genotypes detected in normal oral mucosa:

HPV can be classified as either high risk (HR) or low risk (LR) depending on their ability to cause cancer. HPV can also be classified as cutaneous or mucosal depending on the cellular surfaces infected and specific tissue tropism (Suzich et al., 1995).

High risk HPV can be divided into the most prevalent strains of either species 7 (HPV 18, 39, 45, 59, 68, 70 and 85) or those of species 9 (HPV 16, 31, 33, 35, 52, 58, and 67) (D'Souza et al., 2007a).

Studies have identified the following high risk strains in normal oral mucosa: HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59, in addition to LR-HPV strains. (Kreimer et al., 2010). Oral HPV 16 is the most frequently identified HPV genotype (Kreimer et al., 2010; Kreimer et al., 2011) in the healthy population.

1.6 Oral HPV detection techniques:

Detection of HPV in oral samples is very important for accurate diagnosis and risk stratification. A number of detection techniques currently exist but standardised protocols for HPV testing are required to facilitate clinical usefulness (Suwalska et al., 2014). The following HPV detection methods have been used: p16 immunohistochemistry, in-situ hybridisation (ISH), polymerase chain reaction (PCR), reverse transcriptase PCR, and serology (Schiffman et al. 1991; Blitzer et al., 2014; Singhi and Westra 2010).

Detection of cyclin-dependent kinase-inhibitor 2A (CDKN2A or p16) by immunohistochemistry is widely used in the initial evaluation of HPV status. It is relatively cheap and quick to perform but p16 positivity is not necessarily indicative of HPV infection (Shah et al., 2009; Reimers et al., 2007; Weinberger et al., 2006; Harris et al., 2011).

Southern blots are useful for identifying specific HPV subtypes (Schiffman et al., 1991). The role of southern blot technique in routine clinical practice is limited as it requires large amounts of DNA and takes up time to perform. In-situ hybridisation is another HPV detection technique which has a lower specificity compared to Southern blots (Singhi and Westra, 2010) but it has increased utility as it can be used on fixed or paraffin embedded samples. ISH utilises oligonucleotide probes that can detect high-risk HPV. Highly sensitive detection techniques include polymerase chain reaction (PCR) and reverse transcriptase PCR. PCR detects viral DNA while reverse transcriptase PCR detects RNA. The presence of viral DNA alone is not predictive of cancer development while the overexpression of E6 and E7 predicts carcinogenesis (Blitzer et al., 2014). The Roche Linear Array was used by (Kreimer et al., 2011) for HPV identification while PCR and type-specific hybridisation was employed by (Gillison et al., 2012). Hybrid capture 2 assay has also been proposed as a reliable detection technique for HPV in head and neck squamous cell carcinomas (Smith et al., 2014). This however lacks genotype-specificity.

1.7 Oral HPV prevalence in healthy populations:

The currently available data on the prevalence of HPV infection in the mouth in the general population and the age and gender distribution of this infection is limited to small studies. These are usually retrospective and most of the study cohorts are over ten years old – therefore most of the prevalence data available is at least ten years old. Furthermore, the studies used a variety of sample types and different HPV detection techniques which are polymerase chain reaction (PCR), p16 immunohistochemistry, in-situ hybridisation (ISH). Therefore the results vary widely with reported HPV prevalence rates in tonsil specimens using Polymerase Chain Reaction (PCR) to identify viral DNA reported between 0% and 16% (Kim et al., 2007, Klingenberg et al., 2010, Ernster et al., 2009, Sisk et al., 2006, Ribeiro et

al., 2006, Mammas et al., 2006) and in oropharyngeal exfoliated specimens between 0% and 31.6% (D'Souza et al., 2007a, Smith et al., 2004, Summersgill et al., 2001, Rintala et al., 2006, Jenison et al., 1990, Puranen et al., 1996). These data were compared with the findings from the meta-analysis in this study in Table 2.3.

In children, oral HPV transmission is documented as this can be vertically transmitted during delivery from mother to child (Syrjanen and Puranen, 2000; Pakarian et al., 1994; Chatterjee et al., 1998; Rombaldi et al., 2009; Syrjanen et al., 2010; Pinheiro et al. 2011; Koskimaa et al., 2012). A recent study by Martinelli et al (2012) has identified HPV in the oropharyngeal mucosa in 25 out of 177 (14.1%) newborns in Italy which indicates further studies are required to fully understand the methods of HPV acquisition. The mode of delivery (either via caesarean section or normal delivery) has not been shown to impact on the oral HPV prevalence in 190 children assessed using PCR (Mammas et al., 2012).

Reported prevalence rates of oral HPV infection in adults varies considerably as described below. This is partly due to the lack of standardised HPV testing regimes and the variability in the interpretation of what constitutes a positive result.

In the most recent systematic review on oral HPV infection, Kreimer et al (2010) documented pooled prevalence estimates from 4,070 healthy individuals and found 4.5% (95% CI: 3.9-5.1) for any oral HPV, 3.5% (95% CI: 3.0-4.1) for HR-HPV and 1.3% (95% CI: 1.0-1.7%) for HPV 16. The pooled prevalence obtained from the meta-analysis in this thesis will be compared directly to those published by Kreimer et al. (2010).

In a large cross-sectional study of 5579 adults in America, (Gillison et al., 2012) described a 6.9% (95% C.I: 5.7% - 8.3%) prevalence of oral HPV infection and 1.0% (95% CI 0.7% – 1.3%) prevalence for HPV 16 detected with PCR and ISH using oral rinse/gargle. This study

also shows a double peak for oral HPV infection in the age groups 30 – 34years and 60 – 64years. In men, 10.1% (95% CI: 8.3 – 12.3%) prevalence was found in this population compared with a prevalence of 3.6% (95% CI: 2.6% - 5.0%) in women. Men had a significantly higher prevalence than women, with oral HPV infection more common among those with a history of any type of sexual contact, multiple sexual partners and cigarette smoking (Gillison et al., 2012).

Similarly, Kreimer et al. (2011), sampled 1680 healthy men in the US population for oral HPV infection and found a prevalence of 4.0% (95% C.I: 3.1% - 5.0%). Prevalence of oral HPV infection in the US population was reported by (Sanders et al., 2012) as 7.3% (95% CI: 6.0 – 8.9) from oral rinse with HR-HPV accounting for 3.1% (95%CI: 2.4, 3.9).

The literature on the natural history of oral HPV infection in adults comes largely from the Finnish family study where 324 pregnant women and their partners were tested for oral HPV infection at intervals over a 7year period prospectively (Kero et al., 2012). In this prospective study, point prevalence of oral HPV in oropharyngeal scrapings obtained from 131 males in Finland was determined with the Multimetrix® test (Progen, Heidelberg, Germany). Detection of oropharyngeal HPV DNA in men was found to be common, with HPV16 being the most prevalent genotype. Point prevalence of oropharyngeal HPV infection fluctuated from 15.1% to 31.1% over the study period (Kero et al., 2012). It was suggested that the oropharyngeal mucosa may play a significant role in HPV transmission.

A recent Swedish study found oropharyngeal HPV prevalence of 9.3% in 483 patients in a youth clinic (Du et al., 2012). Detection of oral HPV was found to be significantly higher among females with cervical HPV infection compared to those without cervical HPV infection. Most oropharyngeal HPV genotypes identified in this study matched the cervical

genotypes in women with cervical HPV infection (Du et al., 2012). This suggests some population groups may be at-risk of having higher proportions of oral HPV infection. In Pakistan, oropharyngeal cavity HPV was found in 47 of 200 normal subjects (23.5%) using real-time PCR and a correlation with smoking was documented (Gichki et al., 2012). This high rate of oral HPV infection could be indicative of underlying cultural differences in Pakistan or a reflection of detection technique sensitivity. In contrast, in a sample of healthy Australian University students, a prevalence of 2.3% was found for oral HPV infection with a male preponderance. This study was conducted on 307 students. (Antonsson et al., 2014). This lower oral HPV prevalence rate (when compared with the studies in America), and the high oral HPV prevalence in the Gichki et al. (2012) study in Pakistan, could suggest geographic variation in oral HPV epidemiology. This thesis will explore any such geographic variation in pooled prevalence of worldwide studies.

Prevalence rates of 2.3 – 25% have been reported from saliva testing using lavage for sample collection (Herrero et al., 2003; Goncalves et al., 2006; Seaman et al., 2010; Zhao et al., 2005; D'Souza et al., 2007a; Tachezy et al., 2009; Fakhry et al., 2006; Turner et al. 2011). In contrast to oral rinse, biopsy specimens taken from healthy adults have reported prevalence rates between 0 – 15% (Ostwald et al., 1994; Mao, 1995; Bouda et al., 2000; do Sacramento et al., 2006). This thesis investigate any differences in sample type used for the detection of oral HPV as compared with these literature (Table 2.3).

Data on oropharyngeal HPV prevalence is essential for the assessment of the need for and cost effectiveness of HPV related public health measures, such as including boys in the HPV vaccination programmes. It is also important for the determination of the age of viral acquisition in the general population and hence the age of start of vaccination programmes (Kim et al., 2008; Smith et al., 2007). This data is also essential for the evaluation of the

potential feasibility cost-effectiveness of screening programmes, and preventative educational programmes (Smith et al., 2007).

1.8 Risk factors and transmission routes for oral HPV infection:

Current smoking, male gender, and number of sexual partners have all been associated with oral HPV infection (Gillison et al., 2012). In particular, current tobacco use seemed to increase the odds of oral HPV infection 2.4 times (Kreimer et al., 2011). Oral HPV infection has been shown to have a high male preponderance (Antonsson et al., 2014). It is unclear why males are more predisposed but it could be linked to at-risk sexual behaviours.

Maternal transmission has been suggested as a plausible route of oral HPV transmission (Syrjanen and Puranen, 2000; Pakarian et al., 1994; Rombaldi et al., 2009). Up to 80% of children born to mothers who are positive for genital HPV will have acquired HPV infection in their nasopharynx or oral cavity (Martinez and Troconis, 2014). HR-HPV infections in children may persist for a considerable length of time (Syrjanen et al., 2010; Pakarian et al., 1994; Kravchenko 2011). Rombaldi et al., (2009) reported a 24.5% perinatal transmission rate of HPV DNA. A recent meta-analysis has also shown that vertical transmission could occur in up to 20% of cases (Syrjanen, 2010).

Oral HPV infection appears to be sexually transmitted in adults, with strong associations with multiple sexual partners, an early age of sexual debut, oral sex and French kissing (Satterwhite et al., 2013; Pickard et al., 2012; Gillison et al., 2012; D'Souza et al., 2009; Sanders et al., 2012). People with Human Immunodeficiency Virus (HIV) infection are at an increased risk for simultaneous oral HPV infection too (Beachler et al., 2012). There appear to have been changing trends in sexual behaviours that has coincided with the increasing trend of oral HPV positivity in head and neck cancers (Smith et al, 2004; Johnson et al., 2001).

Some authorities have suggested sexual behaviours such as fellatio and cunnilingus may be high risk behaviour in the acquisition of oral HPV and that past sexual behaviours have an impact on the risk of OPC (Schnelle et al., 2017) but there is a lack of conclusive evidence to support this view (Rosenquist 2012). There has definitely been a shift in recent times in sexual attitudes and practices observed between the younger generation and the older cohort of participants in the NATSAL studies in the UK (Mercer et al., 2013). The younger participants had higher proportions participating in orogenital and anal sex.

D'Souza et al. (2014a) similarly documented recently in a study on a US population the differential rates of oral HPV infection by increasing age based on the sexual behaviour reports provided in the National Health and Nutrition Examination Survey (NHANES). The older generation appeared to have lower oral HPV incidence rates compared to the younger cohort which was associated with less high risk sexual behaviours. This landmark study was on 2,116 men and 2,140 women and showed that although high proportions of both men and women (85.4% vs 83.2%) had ever performed oral sex, men were more likely to have higher numbers of lifetime oral and vaginal sexual partners with higher oral HPV 16 rates. There were statistically significant decreasing proportions of participants that had ever performed oral sex with advancing age as follows in these age groups: 30 – 44 years (90.3%), 45 – 59 years (84.8%) and 60 – 69 years (72.7%), $p<0.001$ (D'Souza et al., 2009). For oral HPV 16 in particular, multivariate analysis showed oral sexual behaviour as the primary predictor. There appeared to be no statistically significant difference reported in the prevalence of oral HPV in gay / bisexual men compared with heterosexual men (13.3% vs 11.0% $p=0.65$) or between lesbian / bisexual women compared with heterosexual women (6.1% vs 3.2%, $p=0.45$). (D'Souza et al., 2014a). It is unclear if digital-oral transmission is plausible. Further studies are interrogating all the possible routes of HPV transmission to the oral cavity.

Recent work investigating if partners of individuals with HPV positive oropharyngeal cancer were at any increased risk, also showed no difference in oral HPV prevalence rates in these partners compared to the rest of the normal population (D'Souza et al., 2014b).

1.9 Oral HPV persistence:

Persistence of high risk HPV especially HPV 16 has been shown to be a critical step in ano-genital carcinogenesis and associated with progression of disease in dysplastic cervical lesions (Remmink et al., 1995) and the development of cervical neoplasia (Cuschieri et al., 2005; Schettino et al., 2014). Syrjanen et al., (2012) suggested that a similar natural history in the oral cavity could mean that persistent infection with HPV 16 especially over 18 months leads to early oropharyngeal cancer. The mechanisms that help oncogenic HPV escape detection and clearance in the oropharynx remain unclear. The majority of oral HPV infections will be cleared by the immune system within 12 – 18 months (Kero et al., 2012) but sometimes oral HPV infections persist (Brendle 2014). New infections can occur in the natural history of oral HPV infection with previously HPV negative individuals becoming positive in consecutive interval testing. Multiple infections can also occur either with several high-risk HPV or a combination of both high- and low-risk strains (Brendle 2014).

Work done in the prospective Finnish HPV family study on 324 pregnant women who had repeated oral HPV testing with nested PCR and Multimetrix® test (Progen, Heidelberg, Germany) over 6 years, showed HPV 16 and 6 were the genotypes that tended to persist most often (Rautava et al., 2012). The definition of persistence used in this study was having at least two positive oral samples with the same HPV strain. 74 of the 324 women (22.8%) had persistent HPV infection in the oropharynx although point prevalence varied between 15 – 24% (Rautava et al., 2012).

1.10 Oral HPV carcinogenesis:

High risk HPVs are the oncogenic genotypes that have been shown to be associated with cancers and include genotypes 16, 18, 31, 33, 35, 51, 56, 66 (Munoz et al., 2006). HPV DNA has been found in cervical cancer (96%), anal cancer (90%), vaginal / vulval cancers (40%), oral cancers (26%) (Suwalska et al., 2014).

Exposure to tobacco has been shown to be associated with oral premalignant lesions. HPV 16 in particular has also been found to have high expression in oral leukoplakia (Sikka and Sikka 2014). Benign oral lesions associated with HPV infection include condyloma acuminatum, squamous cell papilloma, verruca vulgaris and focal epithelial hyperplasia (Syrjanen et al., 2003; Gonzalez et al, 2007)).

High-risk HPV has been associated with oral epithelial dysplasia (McCord et al., 2013) and oral intraepithelial neoplasia (Woo et al., 2013). A recent meta-analysis by (Syrjanen et al., 2011) has shown conclusively that HPV is associated with premalignant lesions of oral cancer. This is a confirmatory work since the group first documented the association between HPV and oral squamous cell carcinomas in 1983. This systematic review looked at 1885 cases of oral squamous cell carcinoma and 2248 controls as well as 956 cases of oral potentially malignant disorders with 675 controls. HPV DNA was found to have significant associations with oral squamous cell carcinoma (OR=3.98; 95% CI: 2.62 – 6.02) and oral potentially malignant disorders (OR=3.87; 95% CI: 2.16 – 6.86) (Syrjanen et al., 2011).

Oncogenic HPV especially HPV 16 has been implicated in the vast majority of HPV-associated head and neck squamous cell carcinoma (HNSCC). HPV 16 is the most prevalent genotype in HPV-associated cancers of the oropharynx, particularly in tonsils and tongue base (Chaturvedi et al., 2013; Combes et al, 2014). HPV is now associated with 30 – 50% of all

head and neck cancers with this proportion rising in oropharyngeal cancer to 50 – 80% (Ang et al., 2010; Chaturvedi et al., 2008). A systematic review by Hobbs et al (2016) showed the oropharyngeal cancer (tonsil and tongue base) had the strongest association with HPV. In a historic paper on risk factors for HPV-OPC, D'Souza et al (2007b) showed oro-genital transmission was important in the pathogenesis of HPV-OPC as well as multiple sexual partners, multiple lifetime sexual partners >6, and cigarette smoking.

The anatomic origin of HPV-OPC is unknown. Some studies have suggested the tonsillar crypts as the site where carcinogenesis commences (Begum et al. 2005). Kim et al (2007) also published work that showed HPV integration begins in the tonsillar crypts and leads to the alteration of p16, EGFR and c-myc during tumourigenesis. Most tonsil cancers associated with HPV arose from the tonsillar crypts, and tended to be inverted and poorly differentiated. P16 expression localizes to HPV-positive cancers, and has been found to be intrinsic to the specialized epithelium of the tonsillar crypts (Begum et al., 2005). HPV has been isolated in the saliva of patients presenting with oral squamous cell carcinoma (Sahebjamee et al., 2009)

A landmark research (Ang et al., 2010) in the RTOG 0129 randomised controlled trial has indicated OPC can be further stratified into high, intermediate and low-risk based on the HPV status, the primary and nodal stage of the disease. This prognostic classifier has been validated in an Italian population (Granata et al., 2012). This has implications for survival as demonstrated by (Fakhry et al., 2008) with a 95% 2 year overall survival in the HPV-OPC group compared with 62% in the HPV negative controls. The low risk group consist of patients who are HPV positive and do not smoke cigarettes; the intermediate group are HPV positive and smoke whilst the high risk group have HPV negative disease and smoke cigarettes. The best survival estimates have been demonstrated in the HPV positive, non-smokers. 3year survival in the low risk, intermediate risk and high risk groups were 93.0%,

70.8% and 46.2% respectively. It is clear HPV status has a strong prognostic effect coupled with these risk factor and disease stage parameters. HPV 16 positive cervical lymph node metastases strongly predict the primary tumour site as oropharyngeal in cancers of unknown primary (Weiss et al., 2011).

1.11 Oropharyngeal cancer risk factors and clinical staging:

Cancers affecting the oropharynx can involve any or a combination of the anatomic sub-sites that include the tonsils, base of tongue, soft palate, uvula and pharyngeal walls. The tonsils are the most commonly involved oropharyngeal sub-site in cancers of the oropharynx followed by the base of tongue (Syrjanen et al., 2003). The vast majority of oropharyngeal cancer are squamous cell carcinomas with lymphomas less frequently presenting as asymmetrical tonsil enlargement. Risk factors for oropharyngeal cancer include tobacco smoking, excessive alcohol consumption (working synergistically) and infection with high-risk human papillomavirus in a subset of cases. Chewing of betel quid or areca nut is also a recognised risk factor in oropharyngeal cancer development especially in individuals from the Asian sub-continent where this practice is rife. There is also a male predisposition as is similarly observed for head and neck cancers generally.

Patients present with a sore throat, swallowing difficulty, and a neck lump (which may be the only presenting symptom). There can be weight loss, tonsillar hypertrophy or ulceration, trismus and on occasion, voice change and breathing problems. The tumour can metastasize to distant organs especially the lungs and bone.

1.11.1 Oropharyngeal cancer clinical staging:

Staging of oropharyngeal cancer is undertaken using the TNM Classification based on the Tumour size, Nodal disease and Metastasis. The most recent AJCC TNM 8th edition (Amin et al., 2017) includes p16 positive (HPV) oropharyngeal cancer for clinical classification:

T0: No primary identified

T1: Tumour measures ≤ 2cm in greatest dimension

T2: Tumour measures >2cm but ≤ 4cm in diameter

T3: Tumour measures >4cm in diameter or extension to lingual surface of epiglottis

T4: Moderately advanced local disease. Tumour extension involving muscles of tongue base, hard palate, constrictor muscles, medial and lateral pterygoid plates, pre-vertebral fascia, or carotid encasement

Nx: Regional lymph nodes cannot be assessed

N0: Absence of nodal involvement in neck

N1: One or more ipsilateral lymph nodes measuring ≤ 6cm in diameter

N2: Contralateral or bilateral metastasis in lymph nodes, all ≤ 6cm in dimension

N3: Nodal involvement >6cm in diameter

Mx: Distant metastasis not assessed

M0: Absence of metastasis to distant organs

M1: Presence of metastasis to distant organs

Depending on the stage of disease, current treatments for oropharyngeal cancer include chemo-radiation and surgical ablation using trans-oral approaches and selective neck dissections. Patients who have had chemo-radiotherapy for oropharyngeal cancer can now be monitored for recurrent disease in the primary site or neck using PET CT scans done 3 months post-treatment (Mehanna et al., 2016). This paradigm shift has ensured patients avoid neck dissections (with the attendant complications) if their PET CT scans do not indicate avid uptake of 5 Fluoro-deoxy-glucose after primary treatment.

1.12 HPV vaccination:

The significant burden of cervical cancer propelled the development of the HPV vaccine. It is hoped that the vaccination of pre-pubertal girls will reduce the incidence of cervical cancer and the attendant mortality. There are currently two HPV vaccines available called Gardasil (Merck and Co. Whitehouse Station, NJ) and Cervarix (GlaxoSmithKline, Philadelphia PA). They both cover the high-risk genotypes HPV 16 and 18 while Gardasil in addition covers for the low-risk HPV 6 and 11. Neither is constituted with HPV DNA but contains virus-like particles made up of the L1 capsid protein of the targeted genotypes. Vaccination stimulates the recipient's immune system to subsequently recognise the L1 capsid protein which constitutes 90% of the outer layer of HPV. Vaccination prevents development of HPV-associated lesions (Suzich et al., 1995).

Population coverage varies from country to country but at present HPV vaccination across Europe is primarily given to girls. There is now a ground-swell of opinion amongst experts in epidemiology, virology, public health and ENT surgery in favour of gender-neutral HPV vaccination, given the increasing incidence of HPV-associated oropharyngeal cancer and its predilection towards young males (Syrjanen et al., 2012). In the UK, HPV vaccination commenced in 2008 and has been delivered to girls aged 12 – 13 years, which is thought to be the age-group that precedes the age of sexual acquisition. The United States policy on HPV vaccination has been modified recently and the programme now includes the vaccination of males. The impact of vaccination both boys and girls will be evaluated in the coming years. There are however protagonists on either side of the discourse around the cost effectiveness of instituting gender-neutral vaccination in the UK. It must be borne in mind the cost of treating these young males who are in the prime of their productive years and the resultant dependence on rehabilitation resources long-term (Mehanna et al., 2010; Kubba 2008).

Since the background on this thesis was written, there has been a significant development with both boys and girls now included on the HPV vaccination programme commencing 2018. This will be further reviewed in the discussion to follow on current trends.

Certainly vaccinating boys will also provide protection for penile cancers, anal cancers, anogenital intraepithelial neoplasia, condylomata acuminate (genital warts) and recurrent respiratory papillomatosis which affect the general population and in addition may be commoner in the immunocompromised (Palefsky, 2010). This systematic review and meta-analysis will provide required data on the baseline prevalence of oral HPV infection amongst healthy individuals stratified by their geographic locations and will help to inform vaccination strategies.

1.13 HPV 16 genome

HPV 16 is a double-stranded DNA virus composed of approximately 8,000 base pairs. The viral life-cycle is linked to the host cell's differentiation and control mechanisms (Graham 2013) as described in section 1.4.

The HPV genome encodes for early and late genes (Figure 1.5). The early genes include E1, E2, E4, E5, E6, and E7. The late genes are L1 (major capsid protein) and L2 (minor capsid protein) and are required for viral assembly and release (Schiffman et al., 2007). The HPV E1 gene is involved in viral replication, the E2 gene in replication and transcription, the E4 gene in viral release, the E5 gene in immune evasion mechanisms, E6 oncogene binds p53, and the E7 oncogene binds the tumour suppressor gene retinoblastoma, RB (McLaughlin-Drubin et al., 2012; Ghittoni et al., 2010). This ultimately leads to a loss of cell-cycle regulation. HPV-associated carcinogenesis has the hallmark of transcriptionally active viral DNA presence and

a loss of regulation for and over-expression of viral oncogenes E6 and E7 (Doorbar et al., 2006).

The expression of E6 and E7 genes critical in carcinogenesis as described, is controlled by the upstream regulatory region (URR) of the HPV 16 genome. The HPV 16 URR consists of p97 promoter, enhancer elements which contain E2 binding sites and the viral origin of replication (ORI) (Figure 1.5). The role of E2 binding in E6 / E7 transcription regulation has been well described. In addition to E2 binding sites, the HPV 16 URR contains TAT and E2 binding site 2. Regulation of HPV gene expression also occurs at the level of RNA splicing and polyadenylation with dysregulation leading to persistent infection and ultimately cancer (Johansson and Schwartz 2013). The early (A_E) and late (A_L) polyadenylation sites in the HPV 16 genome are also shown in Figure 1.5.

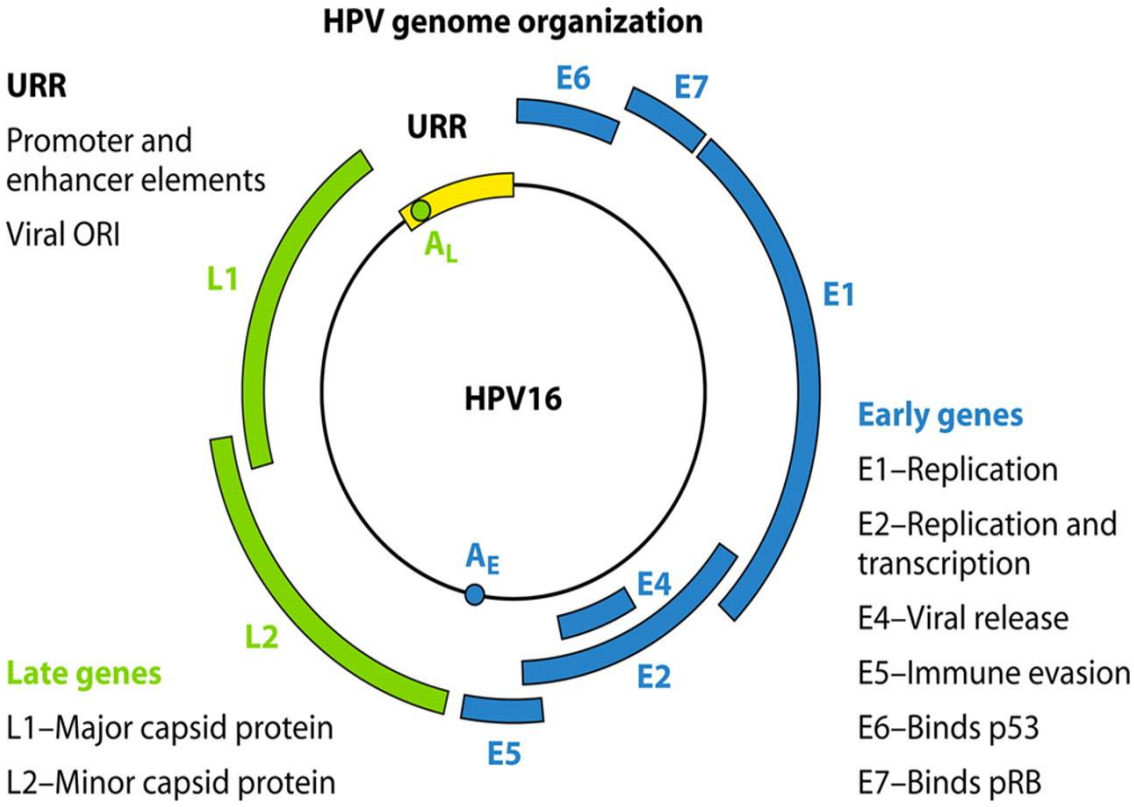


Figure 1.5: Human Papillomavirus 16 genome (Stanley et al., 2012)
HPV 16 genome encodes for early genes E1, E2, E4, E5, E6 and E7 as well as late genes L1 and L2. The viral upstream regulatory region (URR) is also shown containing the origin of replication (ORI), promoter and enhancer elements.

1.14 HPV 16 E2 gene:

The HPV 16 E2 gene encodes for a transcription factor that plays an important role in viral regulation at multiple levels and via different mechanisms. These include initiation of viral DNA replication, maintenance and segregation of viral genomic DNA (Graham 2013). It is also required for replication stability long-term with regulation of the expression of viral helicase E1. Importantly, the E2 protein binds to E2 binding sites located in the Upstream Regulatory Regions (URR) of the HPV genome in order to modulate the p97 promoter in HPV 16 and subsequently the expression of E6 and E7 (zur Hausen 2009; Bernard 2002). Loss of E2 regulation, as a consequence of viral integration and disruption of the E2 open reading frame (described in detail in 1.18) is considered a key step in carcinogenesis especially in anogenital cancers (Kalantari et al 1998, Arias-Pulido et al 2006).

1.15 HPV 16 E2 binding sites:

The HPV 16 E2 recognizes and binds to palindromic DNA sequences *(*5^1*-ACCG-(N)*$_4$*-CGGT-*3^1*)* (Phelps and Howley 1987; Romanczuk et al 1990). There are four recognized E2 binding sites in the HPV 16 URR, with three of them being transcriptionally relevant: the *high-affinity* E2 binding site 1 (E2BS1) which involves CpG positions 7452 and 7458 and the *low-affinity* E2 binding sites 3 (E2BS3) located at CpG positions 37, 43 and the E2 binding site 4 (E2BS4) located at CpG position 52, 58. The fourth E2 binding site – E2BS2 located at CpG position 7460 has not been shown to be involved in transcription regulation (Thierry 2009).

1.16 HPV E6 and E7 mechanisms of action in carcinogenesis:

Early oncoproteins produced by HPV include E6 and E7, with their expression retained in all HPV-induced cancers. These two oncoproteins lead to deregulation of tumour suppressor genes and ultimately genomic instability (Roberts 2015). Several cellular pathways responsible for cell cycle control are targeted by both E6 and E7, in addition to regulation of apoptosis and cell polarity (Vjekoslav et al., 2016). The dysregulation of cell cycle pathways and apoptosis caused by oncoproteins E6 and E7 allows HPV proliferation to be maintained in suprabasal epithelium and eventual viral genomic amplification in the HPV lifecycle (Ganti et al., 2015; Doorbar et al., 2012).

When viral integration occurs into host genome during carcinogenesis, most of the early proteins are lost but E6 and E7 expression remains uncontrollable leading to cellular immortalization and transformation (Smotkin and Wettstein 1986). The actions of E6 and E7 have been shown to be complementary with E6 promoting survival pathways in the cell while E7 promotes cellular proliferation (Barbosa et al., 1989; Hawley-Nelson et al., 1989; Mantovani and Banks 2001). Overexpression of E6 and E7 leads to dysregulation of p53 and RB proteins respectively (Villavicencio-Torres et al., 2013) (Figure 1.6). However, in head and neck squamous cell cancers, oncoprotein E7 appears to play a more dominant role in tandem with E6 which acts in later cancer stages (Strati and Lambert 2007).

The E7 oncogene is involved in DNA synthesis initiation and also stimulates continuous cell growth. The retinoblastoma gene pRB, is an important negative regulator of the cell cycle S phase entry. When RB is unphosphorylated (in the early G1 phase), it interacts with E2F transcriptional factors and functions as a transcriptional repressor of promoters that contain sites with E2F (Dyson et al., 1989). E2F family are involved in the regulation of many genes

associated with cell cycle progression (Slansky and Farnham 1996). G1 cyclin-dependent kinases mediate phosphorylation of p RB which causes E2F release accompanied by cell cycle progression to the S phase (Figure 1.6). In HPV 16 infected cells, E7 targets and degrades unphophorylated RB through the ubiquitin-proteasome pathway (Boyer et al., 1996) (Figure 1.6). This interaction can cause a disruption of the p RB – E2F complexes and induce premature entry of cells into the S phase of the cell cycle (zur Hausen 2000, 2002; Boyer et al., 1996; Jones et al., 1997; Berezutskaya et al., 1997; Nguyen et al., 2002). Specifically in HPV 16, the association of E7 with the cullin 2 ubiquitin ligase complex is also contributory (Huh et al., 2007). The HPV 16 oncoprotein E6 is involved in a number of interactions in HPV infected cells. Broadly, these important interactions include a proteasome-mediated degradation of p53 tumour suppressor gene, degradation of PDZ and expression of hTERT (Senba and Mori 2012) (Figure 1.6).

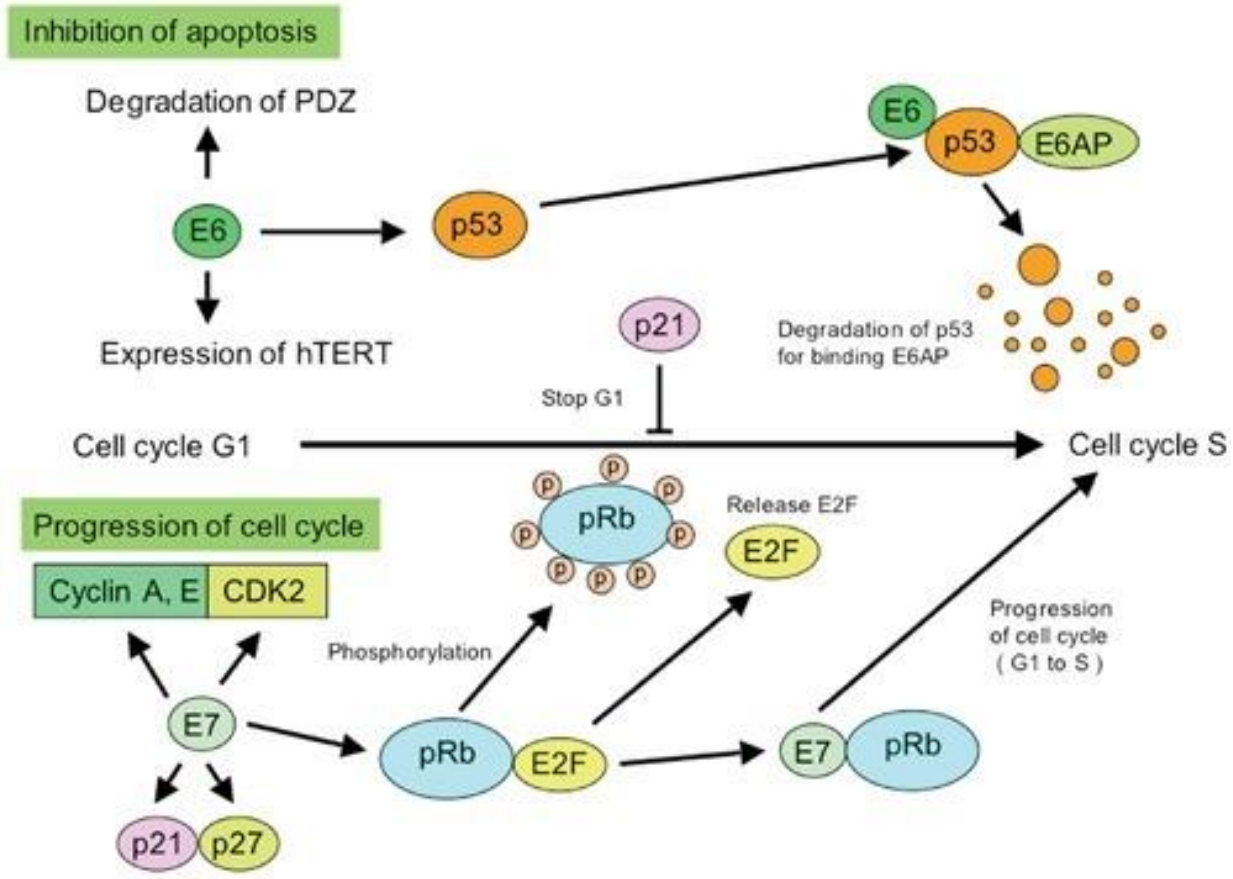


Figure 1.6: HPV E6 and E7 critical interactions in cell cycle deregulation and carcinogenesis. *High-risk human papilloma virus E6 and E7 oncoproteins react with the tumor suppressor gene products p53 and pRb in host cell proteins, respectively. E7 inhibits the retinoblastoma family and constitutive activation of E2F responsive genes. E6 inhibits p53-dependent growth arrest and apoptosis. Degradation of p53 bypasses the normal growth arrest of the cell cycle from G1 to S phases (Senba and Mori 2012).*

Following DNA damage in untransformed cells, the tumour suppressor gene p53 is involved in apoptosis. p53 is also involved in the regulation of cell cycle phases G1/S and G2/M (Thomas et al., 1999). HPV infected cells cause a loss of inhibition of apoptosis via E6 inactivation of p53. The HPV E6 oncoprotein labels the p53 tumour suppressor gene for degradation by transporting ubiquitin peptides from the E6 associated protein (E6AP) to p53 (Figure 1.6). This results in activation of human telomerase reverse transcriptase (hTERT) with an attendant increase in cellular telomerase activity in tumour cells and subsequently in

tumour cells not being subject to cellular aging (Senba and Mori 2012). E6 also causes a dysfunction of some proteins that induce apoptosis namely, BAK, FADD and procaspase 8 (Thomas et al., 1999; Garnett et al., 2006). BAK proteins are responsible for apoptosis in differentiated epithelium. High-risk HPV E6 oncoprotein binds to BAK proteins with high affinity, resulting in suppression of cellular apoptosis and immortalization (Modrow et al., 2013; Haedicke and Iftner 2013).

The complex viral and host cell interactions initiated by HPV oncoproteins E6 and E7 are aptly summarised by Chung and Gillison (2009) in Figure 1.7. It shows ubiquitination by E7 and the cullin 2 ubiquitin ligase complex leads to degradation of p RB (Dyson et al., 1989; Huh et al., 2007; Harbour et al., 1999; Rubin et al., 2005) (Figure 1.9A). Inhibition of cell cycle arrest occurs when E7 interacts with p27Kip1 (Zerfass-Thome et al., 1996) and p21Cip1 (Jones et al., 1997b) respectively (Figure 1.9B and C). Ubiquitination by E6 and ubiquitin ligase E6AP leads to p53 degradation (Scheffner et al., 1990 and 1993; Huibregtse et al., 1991) (Figure 1.9 D) with apoptosis resistance and genomic instability. This complex degradation of p53 also leads to deregulation of DNA damage repair and cellular senescence. There is increased expression of p16INK4A as a consequence of feedback loops resulting from pRB dysfunction (Khleif et al., 1996) (Figure 1.9 E). The transcriptional repressor of hTERT called NFX-1 is also degraded by the E6-E6AP complex with resultant cellular immortalization (Katzenellenbogen et al., 2009) (Figure 1.9F). The combination of these cell cycle dysregulations by HPV oncoproteins E6 and E7 lead to genomic instability and malignant progression.

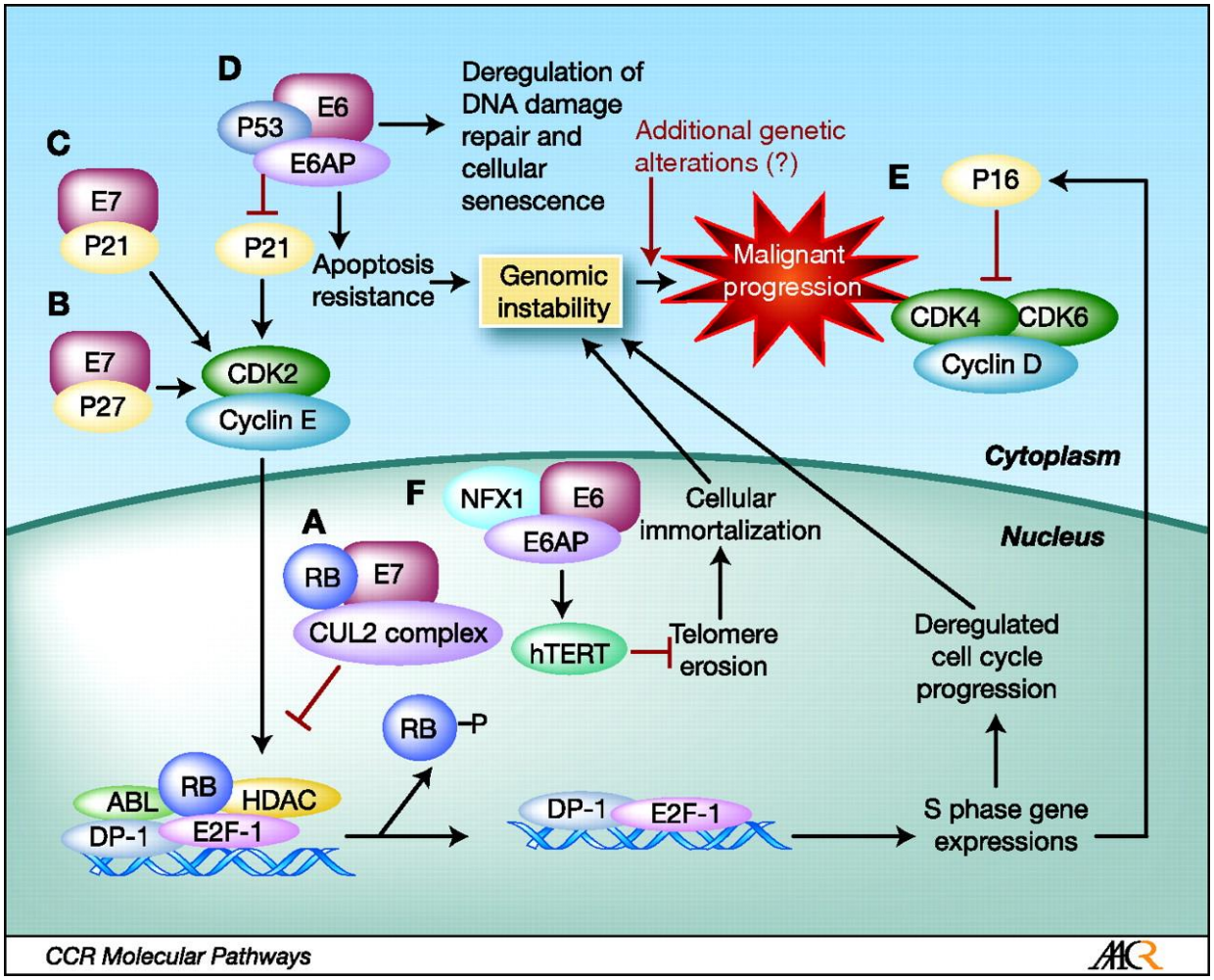


Figure 1.7: Malignant transformation in keratinocytes mediated by HPV E6 and E7 oncoproteins (Chung and Gillison 2009; Used with permission – RightsLink licence number 4447620877537)

Clockwise from A, ubiquitination by E7 and the cullin 2 ubiquitin ligase complex leading to pRb degradation; B, interaction between E7 and p27Kip1 resulting in inhibition of cell cycle arrest contributing to carcinogenesis; C, interaction between E7 and p21Cip1 resulting in inhibition of cell cycle arrest contributing to carcinogenesis; D, ubiquitination by E6 and ubiquitin ligase E6AP leading to p53 degradation; E, increased expression of p16INK4A by a consequent of feedback loops from the absence of pRb function; and F, degradation of NFX1, a transcriptional repressor of hTERT, by association with E6/E6AP resulting in hTERT activation and cellular immortalization (Chung and Gillison 2009 – Used with permission).

1.17 HPV 16 viral auto-regulation

Auto-regulation of the HPV 16 viral genome ensures low levels of E6 and E7 transcripts when the E2 gene is intact (Reuschenbach et al., 2015). This auto-regulation is shown in the upper panel of Figure 1.8 (A). The E2 protein will preferentially bind to the high-affinity E2BS1 resulting in activation of the p97 promoter which leads to increased expression of E6 and E7 but also E2 itself. The increased levels of E2 would in turn result in enhanced binding to the low-affinity E2BS3 and E2BS4 sites, responsible for promoter inhibition. Therefore, binding of E2 protein to these 2 binding sites results in a counteracting inhibitory effect to the promoter stimulatory effect produced by E2 binding to E2BS1. Loss of E2 protein and hence this auto-regulation results in over-expression of oncogenes E6 and E7 which inhibit the actions of tumour suppressor proteins p53 and RB respectively as shown in the lower panel, Figure 1.8 (B). There are two proposed mechanisms as to the loss of viral autoregulation: disruption of the E2 gene and epigenetic mechanisms involving methylation at E2 binding sites.

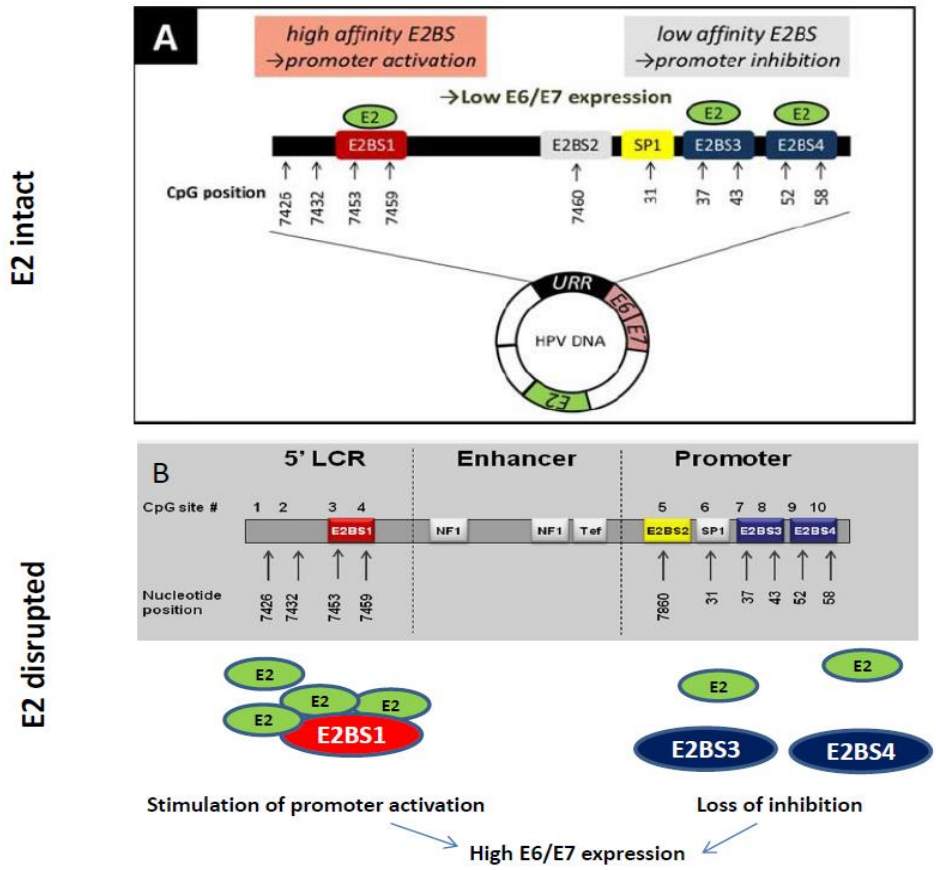


Figure 1.8: Auto-regulation of E6/E7 oncogene expression by E2 protein binding in the HPV 16 URR and the methylation-mediated over-expression of E6/E7. (Adapted from Reuschenbach et al 2015). *When the E2 gene is intact as shown in upper panel, the E2 protein binds avidly to the high-affinity E2BS1(resulting in activation of p97 promoter) as well as the low affinity E2BS3 and 4(with promoter inhibition). The balance achieved results in low levels of E6/E7 expression. With E2 gene disruption as shown in lower panel, the low concentrations of E2 protein preferentially bind to the high-affinity E2BS1 resulting in promoter hyperactivation. The reduced binding to E2BS3 and E2BS4 leads to a loss of promoter inhibition. The combined effect is evidenced by high expression levels of E6 and E7. E2BS1 location overlaps with CpG sites 7453, 7459; E2BS3 with CpG sites 37, 43; and E2BS4 with CpG sites 52, 58.*

1.18 HPV 16 integration in oropharyngeal cancer and E2 gene disruption

HPV integration results in disruption of the viral regulatory E2 gene. This step in carcinogenesis has been shown to be an associated but not necessary step for all cervical cancers (Graham 2013, Raybould et al., 2014). The disruption of the E2 gene occurs in approximately 70% of cervical cancer cases in published literature. However, in tonsillar cancers, the viral genome integration has been reported at much lower frequency with approximately 16 – 43% E2 gene disruption assessed by PCR, reported in the available literature as shown in Table 1.1. This range of integration in tonsil cancer is pooled only from studies using PCR-based methods and that include tumour samples for ten or more patients (Table 1.1).

When the E2 gene is disrupted in a mixed (intact and episomal) infection, there is less E2 protein concentration available to bind to the E2 binding sites in the HPV 16 URR. In this scenario, the available E2 protein binds preferentially to the high-affinity E2BS1 with unopposed activation of the p97 promoter (Figure 1.8, lower panel). In addition to disruption of E2 sequence itself, it has become clear from recent studies using next-generation sequencing techniques, that HPV integration might occur at multiple sites outside of E2 gene, including E1, L1, and L5 (Wilson et al., 2013). The HPV 16 genome transcription mechanisms are poly-cistronic and a dysfunction along the chain upstream of the E2 gene could affect its function even if the E2 gene remains undisrupted and intact when assessed by PCR. Therefore, the conclusion that HPV integration is absent based on E2 disruption solely should be drawn with caution. This is a crucial point that exposes the limitation in the technique used in determining the physical status of the viral E2 gene in this study i.e. PCR with overlapping primers to the HPV 16 E2 gene. Different techniques might be more sensitive and this could also

explain the differences in HPV integration observed across published studies as shown in Table 1.1. For example, in-situ hybridisation identified 100% HPV integration when used in combination with p16 immunohistochemistry on 8 OPC samples (Begum 2005) but on its own, in-situ hybridisation demonstrated HPV integration in 67% of 12 OPC samples (Hafkamp et al., 2003) and 42% integration in the largest OPC cohort of 77 samples (Mooren et al., 2013) (see Table 1.1). APOT-PCR on the other hand, as shown in Table 1.1, showed HPV integration in 37% - 67% of the OPCs analysed in various studies (Reuschenbach et al., 2015; Olthof et al., 2015; Lace et al., 2011; Wiest et al., 2002).

In addition, integration and/or E2 disruption is rarely accompanied by a loss of E2 expression in OPC due to the presence of both intact (episomal) and disrupted viral genomes (Reuschenbach et al., 2015). Therefore, additional mechanisms of E2 dysregulation are considered.

Table 1.1: Literature on HPV 16 integration and E2 disruption in OPC. *This table lists the studies reporting HPV integration in OPC and the different techniques used including detection of integrated papillomavirus sequences by ligation-mediated polymerase chain reaction (DIPS-PCR), amplification of papillomavirus oncogene transcripts (APOT)-PCR, fluorescence in-situ hybridisation (FISH), p16 immunohistochemistry (IHC), real-time q PCR, genome-wide next generation sequencing, restriction enzyme cleavage, ligation and PCR (rliPCR), E2 mRNA level mapping, E2: E6 ratio <0.5, and bioinformatics analysis of RNA Seq data.*

First Author	Year	OPC sample size	HPV integration technique	Percentage HPV integration
Olthof	2015	75	DIPS and APOT-PCR	39
Reuschenbach	2015	49	APOT and DIPS-PCR	37
Khoury	2013	36	Bioinformatics analysis of RNA Seq data	67
Vojtechova	2016	14	Mapping E2 mRNA level, APOT and Southern blot	43
Gao	2014	13	Genome-wide mate pair next generation sequencing	15
Hafkamp	2003	12	FISH	67
Mellin	2002	12	Restriction enzyme cleavage, ligation and PCR (rliPCR)	0
Wiest	2002	9	APOT-PCR	67
Begum	2005	8	HPV 16 in situ hybridization and P16 IHC	100
Lace	2011	4	APOT-PCR	50
Wilson	2013	3	HPV whole genome sequencing	
Mooren	2013	77	FISH	42
Barbieri	2014	51	real-time q PCR, E2: E6 ratio <0.5	16
Park	2011	20	real-time q PCR, E2: E6 ratio <0.5	35
Koskinen	2003	23	real-time q PCR, E2: E6 ratio <0.5	48
Deng	2013	38	real-time q PCR, E2: E6 ratio <0.5	4.4

1.19 HPV epigenetics and the role of DNA methylation:

DNA methylation is an addition of a methyl group to cytosine to form 5-methylcytosine (Figure 1.9). Enzymes that add a methyl group are called DNA methyltransferases. Cytosines are methylated within CpG dinucleotides. The CpG sites or CG sites are dinucleotides where a cytosine nucleotide occurs next to a guanine nucleotide in the linear sequence of bases along its length. High densities of CpGs are often present at gene promoters and transcriptional start sites and such regions are referred to as CpG islands. In mammals, CpG island methylation can silence the gene.

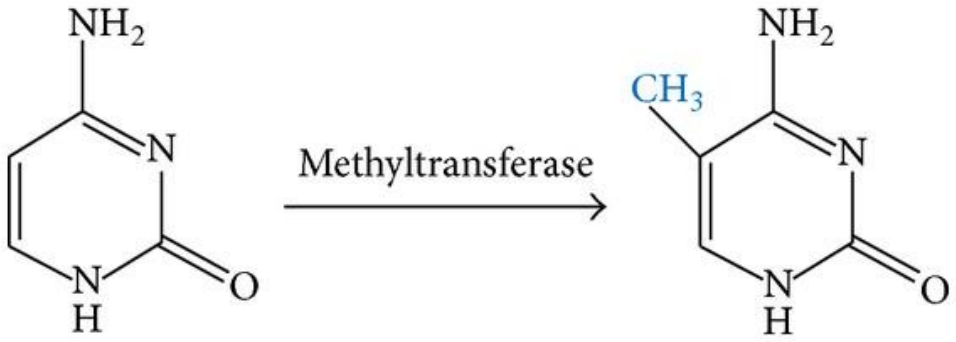


Figure 1.9: Conversion of cytosine to 5-Methylcytosine (Fakhr et al., 2013)

In mammals, DNA methylation is essential for normal development and is associated with a number of key processes including genomic imprinting, inactivation of the X-chromosome, suppression of repetitive elements, and carcinogenesis. In HPV 16 genome, methylation has been involved in regulation of the HPV E2 protein binding affinity to the URR, leading to abnormal expression of E6 and E7 oncoproteins (Doeberitz and Vinokurova 2009)

1.20 HPV 16 E2 binding sites methylation and epigenetic controls

In cervical cancer research, it has been shown that increasing levels of HPV methylation correlate with increasing severity of cervical cancer grade i.e. progressing from intraepithelial neoplasia to frank cervical cancer (Chaiwongkot et al., 2013). In oropharyngeal cancer, work done by Doeberitz and Vinokurova (2009) and subsequently by Reuschenbach et al., (2015), revealed the impact of methylation on the auto-regulation of the HPV 16. They hypothesized that hypermethylation of the high-affinity E2 binding site 1 (E2BS1) located in the long control region (LCR) of the HPV 16 Upstream Regulatory Region (URR) at the 5' end results in further promoter activation (Figure 1.10). In contrast, methylation of the low-affinity inhibitory E2 binding sites (E2BS3 and E2BS4) located in the promoter region of the HPV 16 URR results in loss of promoter inhibition. Similar conclusions have been drawn from the *in vitro* studies (Thain et al., 1996). The combination of these two methylation events at the E2 binding sites culminates in over-expression of oncogenes E6 and E7.

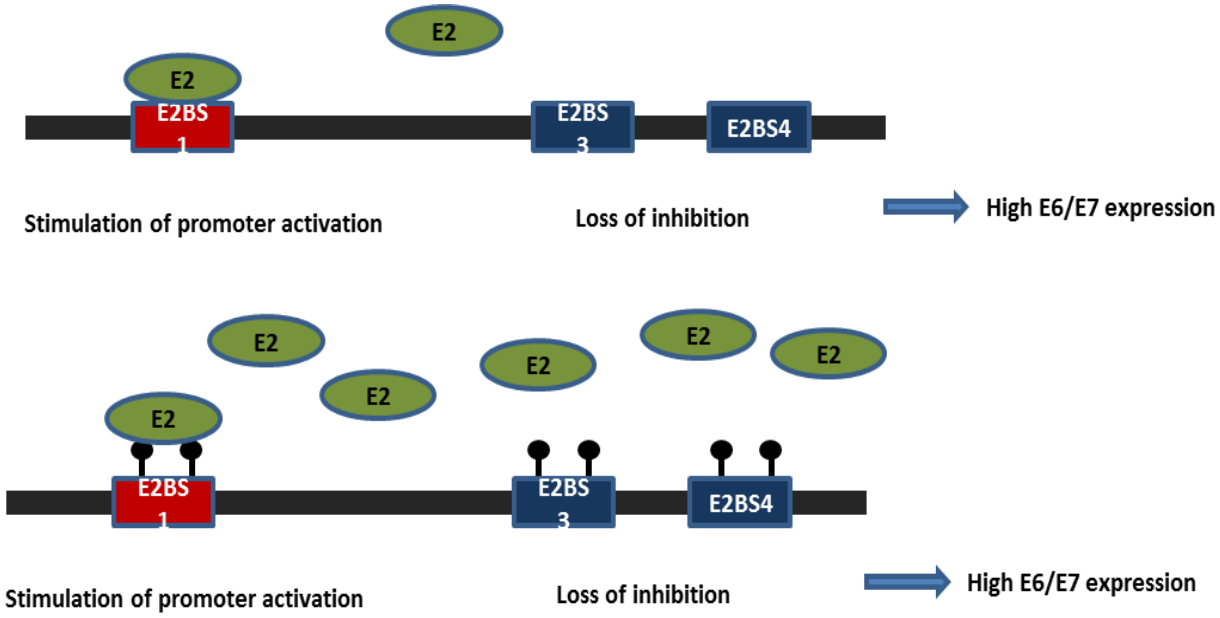


Figure 1.10: Impact of E2 binding sites methylation on E6 and E7 expression. *The upper panel shows preferential binding of the E2 protein to E2BS1 and reduced binding to inhibitory E2BS3 and E2BS4 in the presence of E2 gene disruption (in mixed infection), with over-expression of E6 and E7 oncoproteins. The lower panel depicts the potential impact of methylation on the E2BSs with methylation at E2BS1 resulting in promoter activation while methylation at E2BS3 and E2BS4 results in loss of promoter inhibition. The combination of these effects leads to over-expression of E6 and E7 (Reuschenbach et al., 2015)*

So far, studies on methylation in oropharyngeal cancer indicate that the LCR region of HPV 16, location of the E2BS1, is mostly unmethylated (see Table 1.2). However, Barbieri et al. (2014) investigated the methylation profiles at the HPV 16 URR in a cohort of 60 tumours and found that the promoter activating E2BS1 had higher methylation frequency when compared with the promoter inhibiting E2BS3 and E2BS4. This was similar to the observations made in cervical cancer (Jacquin et al., 2013). In another study on HPV OPC, Reuschenbach et al. (2015) found E2 gene disruption by PCR in 37% of their cohort. In this study, there was no significant difference in methylation at the high-affinity E2BS1 regardless of the E2 status (intact or disrupted) while these Oropharyngeal cancer could be stratified into low, intermediate, high methylation at the inhibitory E2 binding sites (E2BS3 and E2BS4).

Significantly, cases of oropharyngeal cancer with hypermethylation at the inhibitory E2 binding sites had a worse 5 year survival when compared to those cases with intermediate methylation (Reuschenbach et al., 2015).

Table 1.2: HPV methylation in oropharyngeal cancer

Author	Year	OPC sample size	Viral methylation findings
Reuschenbach	2015	57	3 subgroups with differential methylation at E2BS 3 and 4; high methylation associated with worse overall 5yr survival
Barbieri	2014	60	E2BS1 has higher methylation frequency compared to E2BS 3 and 4
Park	2011	22	HPV 16 epigenome is hypomethylated in advanced HPV HNC with the LCR preferentially hypomethylated
Wilson	2013	3	453 differentially methylated regions in the HPV host methylome. HPV 16 integrates at multiple sites and predominantly involves E2 viral genome. LCR unmethylated

Some of the reported techniques for determination of viral methylation patterns in head and neck squamous cell carcinomas are listed in Table 1.3 and discussed by van Kempen et al. (2014) in a systematic review that also covers changes in DNA methylation in the host. These include methylation-specific PCR (MSP), combined bisulfite restriction analysis, pyrosequencing analysis, bead array method and quantitative MSP.

Table 1.3: Methylation techniques in head and neck cancers

Methylation determination technique	Literature
Methylation specific PCR (MSP)	Ren 2018
Combined bisulfite restriction analysis (COBRA)	Bennett 2010 Furniss 2008
Pyrosequencing analysis (PMA)	Poage 2011 Richards 2009
Bead Array Method (BAM)	Colacino 2013 Lechner 2013 Sartor 2011
Quantitative MSP (q-MSP)	Kostareli 2013 Gubanova 2012

Pyrosequencing has been used to quantify HPV 16 DNA methylation in transcriptionally-active HPV OPC (Barbieri et al., 2014). The E2 binding site 1 (E2BS1) had higher mean methylation frequency compared with E2BS3/4. In this study, mean methylation frequency >10% in the promoter region of the HPV genome, was associated with a worse clinical prognosis (Barbieri et al., 2014).

Hypomethylation of the long control region (the regulatory region of the viral genome) was also observed in advanced stage III/IV HPV-associated oropharyngeal cancer which correlated with expression levels of the E6 and E7 oncogenes (Park et al., 2011). There appears to be variability in the viral load in head and neck cancers as well as the existence of both episomal and integrated HPV in oropharyngeal cancer (Park et al., 2011). The authors also found that methylation status of the long control region in advanced oropharyngeal

cancer was detectable in saliva and serum using methylation-specific PCR. This could be a significant biomarker for HPV DNA integration if confirmed in a study with a larger sample size.

1.21 PhD project rationale:

The epidemiology and natural history of oral HPV infection in OPC remains incompletely described. This information is crucial to our understanding of how high-risk human papillomavirus (HR-HPV) infection in the oral cavity progresses through persistence, latency, immune system evasion or modification and ultimately cancer causation. The systematic review and meta-analysis in **Chapter 2** of this thesis aims to provide a comprehensive overview of oral HPV epidemiology and natural history in healthy populations with respect to prevalence, persistence, associated risk factors, and any variations in geographic location, detection technique or oral sample type. This has implications for the development of any screening strategies, the identification of any at-risk population groups and the evaluation of the need for and efficacy of HPV vaccination programs.

HPV oropharyngeal cancer is significantly different to HPV negative oropharyngeal cancer in every sphere – epidemiologically, biologically, in disease progression and in response to current treatments. In the main, patients with HPV positive oropharyngeal cancer tend to be males, are younger than patients with HPV negative OPC and do not necessarily have a history of tobacco smoking or excessive alcohol consumption.

There is an on-going discourse around the potential impact of tonsillectomy as an intervention that abrogates the natural history of oral HPV infection and therefore protects against the development of HPV-associated oropharyngeal cancer long-term. A recent Danish study using population-based cancer registry data observed a reduction in oropharyngeal cancer

rates in a cohort of people who had a history of tonsillectomy when compared with individuals who did not have tonsillectomy (Fakhry et al., 2015). The study also showed this protective effect was greatest when tonsillectomy was performed over a year prior to a diagnosis of oropharyngeal cancer. This finding needs validation in other populations as it could mean that tonsillectomy is a cost-effective preventative secondary measure alongside primary prevention with HPV vaccinations. **Chapter 3** of this thesis presents for the first time in the UK, data on the potential impact of tonsillectomy on the prevalence of oral HPV infection in a prospective cohort of healthy individuals who had tonsillectomy for non-cancer reasons.

Oropharyngeal cancer is linked to various demographic and risk factor behaviours which also have a significant impact on recurrence-free and overall survival. **Chapter 4** of this thesis describes a retrospective review of 28,846 oropharyngeal cancer cases registered in the UK Hospital Episodes Statistics (HES) database and identifies the impact of age, gender, socioeconomic status or deprivation scores on survival and cancer recurrence. This chapter also investigates any regional variations in the UK in the incidence of oropharyngeal cancer. This is important for targeted cancer-risk assessments and resource allocation with a view to better understand the underlying reasons for any regional predispositions and the institution of appropriate preventative and diagnostic pathways.

There is very little published on the HPV genome and epigenome in HPV-OPC despite its distinct molecular profile when compared with HPV negative OPC. The physical status of the virus and methylation has been studied in some centres with conflicting accounts on relevance to clinical outcomes. This thesis explores genomic or epigenetic modifications such as HPV E2 gene disruption and aberrant methylation at the E2 bindings sites in the viral upstream regulatory region that could impact on clinical outcomes such as cancer recurrence and

survival in a cohort of HPV positive OPC described in **Chapter 5**. The possible impact of risk factors such as smoking and demographic variables such as gender and age, on the HPV genome and epigenome itself, are also investigated in this thesis.

The findings from this thesis will corroborate current literature on the epidemiology of oral HPV infection in healthy individuals in a global meta-analysis and also reiterate determinants of survival in a cohort of UK patients with oropharyngeal cancer. There will be new information on the potential impact of tonsillectomy on oral HPV burden as well as new observations on viral genomic and epigenetic mechanisms in a cohort of HPV-associated oropharyngeal cancer that adds to the limited data currently available.

PhD CHAPTER 2

PREVALENCE AND PERSISTENCE OF OROPHARYNGEAL HUMAN PAPILLOMAVIRUS INFECTION IN HEALTHY ADULT POPULATIONS BY SAMPLE TYPE, DETECTION TECHNIQUE AND GEOGRAPHY:

A SYSTEMATIC REVIEW AND META-ANALYSIS

2.1 Introduction:

There remain significant gaps in our knowledge of HPV-OPC and HPV infection of the mouth at this time including its prevalence, natural history, risk factors, transmission, viral persistence, the presence of premalignant lesions, and the effect of vaccination, effective treatment modalities and the projected disease burden. This systematic review of the current literature focused on the natural history of oral HPV infection in particular viral prevalence and persistence, with a view to facilitating our understanding of this disease. This information is important for the assessment of the need for and cost effectiveness of HPV related public health measures, such as including boys in the HPV vaccination programmes. It is also important for the determination of the age of viral acquisition in the general population and hence the age to start vaccination programmes (Kim et al., 2008; Smith et al., 2007). These data are also essential for the evaluation of the potential feasibility cost-effectiveness of screening programmes, and preventative educational programmes (Smith et al., 2007). There is currently no UK data available on the subject of persistence of oropharyngeal HPV infection which this study aimed to provide with an elucidation of the possible risk factors that ultimately lead to cancer development. It is also unclear if geographic location plays any role in the prevalence of oral HPV infection and this review interrogated this further.

2.2 Review questions:

1. To describe the prevalence of oral and oropharyngeal HPV infection in healthy adult populations by age, gender, viral type, sample type, HPV detection technique, and geography.
2. To identify the high risk HPV genotype(s) that persists in the oral cavity / oropharynx and the duration of persistence.

2.3 Review protocol:

The protocol for this systematic review and meta-analysis was published online in accordance with international guidelines for the conduct of systematic reviews on PROSPERO (International Prospective Register of Systematic Reviews) in 2014: Oladejo Olaleye, Esmond Carr, Jayne Wilson, Hisham Mehanna. *Prevalence and Persistence of oral and oropharyngeal human papilloma virus infection in the healthy population by sample type, detection technique and geography: a systematic review and meta-analysis.* PROSPERO 2014 CRD42014009205 Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42014009205

2.4 Population, exposure and outcomes:

Population: Normal population.

Exposure: Oral or Oropharyngeal human papilloma virus infection.

Outcomes: Prevalence and persistence of oral HPV infection

- Stratified by sample type, detection technique and geography.

2.5 Inclusion criteria for participants:

1. Patients or healthy volunteers aged 12 years and above, from any geographic location in the world and regardless of whether they were sexually active or not.
2. Participants with or without any documented risk factors such as cigarette smoking, alcohol consumption, oral sex, French kissing or multiple sexual partners.
3. Any human studies with oral HPV status documented as well as duration of persistent infection in any age group and with any HPV detection technique done anywhere in the world.

2.6 Primary outcomes:

1. Prevalence of oral HPV infection expressed as pooled prevalence with 95% confidence intervals
2. Prevalence of oral HPV infection by detection technique, tissue type, gender, age, and geography
3. Persistence of oral and oropharyngeal human papilloma virus infection and the duration of persistence.
4. The analyses were expressed as weighted proportions, with accompanying 95% confidence intervals.

2.7 Secondary outcomes:

1. Investigated time-trends of oral HPV prevalence reported in all included studies.
2. Identified by age-distribution analyses, any particular age-group susceptible to higher oral HPV prevalence
3. Documented any variations in oral HPV persistence by gender, sample type (tonsil, oral rinse, tongue base / pharyngeal swabs), detection technique (p16 immunohistochemistry, HPV DNA by polymerase chain reaction, ELISA) and geography.
4. The analyses were expressed as weighed proportions, with accompanying 95% confidence intervals.

2.8 Types of studies included:

1. Randomised controlled clinical trials (RCTs) were considered eligible for inclusion if:
   - Oral HPV status was included as part of the demographics for healthy adults
   - Oral HPV detection techniques were specified
   - Healthy adults were in the non-intervention arm of the RCT e.g. receiving no HPV vaccination
2. Cross-sectional studies on healthy adult populations that tested for oral HPV infection
3. Prospective or retrospective Cohort studies of healthy adults for oral HPV testing
4. Population screening studies for oral HPV infection.
5. Studies that documented oral HPV prevalence by any detection technique in any geographic location
6. Studies that assessed the diagnostic efficacy of oral HPV screening tests including use of PCR, ELISA, p16 immunohistochemistry, E4 or E6 proteins
7. Studies that assessed differences in HPV detection in various samples such as oral rinse, oral / throat swabs, pharyngeal swabs / brushes, tongue base brushes, tonsil specimens
8. Studies that documented repeated sample testing for persistence of oral HPV infection regardless of duration from index testing

2.9 Exclusions:

1. Case-control studies with oral HPV infection status including both cancer and non-cancer case-control studies
2. Studies in populations that could not be classed as healthy adults or could be at-risk populations for higher oral HPV infection rates such as head and neck cancer patients,

immunocompromised (HIV), sexually transmitted diseases clinic attendees, women with HPV positive cervical tests, genetic abnormalities, Fanconi's anaemia patients, and participants with any clinically detectable oral lesions.

3. All oral HPV studies on children or anyone less than 12 years old were excluded as viral acquisition and transmission routes differ in this group.

4. All studies with sample size less than 50 healthy adults.

5. Exclusions also included letters, opinion papers, editorials, case reports

6. All animal studies

7. All in-vitro studies

2.10 Searches:

A comprehensive search of the published literature was performed as detailed in the published protocol on PROSPERO in 2014 cited in section 2.3. Studies were identified from reviews, database searches, clinical trials, specialized databases and grey literature.

2.10.1 *Reviews* included Cochrane database of systematic reviews, Database of Abstracts of Reviews on Effects (DARE), International Prospective Register of Systematic Reviews (PROSPERO) and Health Technology Assessment (HTA).

2.10.2 *Databases* searched included MEDLINE, EMBASE, NICE and the Cochrane Central Register of Controlled Trials

2.10.3 *Clinical trials* included trials on the current controlled metagister: http://www.controlled-trials.com/mrct/ ClinicalTrials.gov http://clinicaltrials.gov/, the UK International Clinical Research Network Portfolio (NIHR CRN) http://www.crncc.nihr.ac.uk/about_us/processes/portfolio and the WHO

International Clinical Trials Registry Platform (ICTRP) http://www.who.int/ictrp/en/

2.10.4 *Specialised databases* were searched on the websites for the International Agency for Health Technology Assessment (INAHTA), the International Agency for Research into Cancer (IARC), and the University of York National Health Service (NHS) Centre for Reviews and Dissemination (CRD)

2.10.5 *Grey literature* search and other literature searches were conducted by reviewing relevant paper from Conference papers index, Zetoc database of conference abstracts https://zetoc.mimas.ac.uk/ , internet searching (e.g. Google scholar), hand-searching of Journals such as Head Neck, Clinical Otolaryngology, European Archives of Otolaryngology, Head and Neck Surgery, Virology, Oral Oncology. Content experts were also contacted for additional studies, and for further information regarding on-going trials or studies. No language restrictions were used. The reference lists of all potentially relevant papers including reviews, meta-analyses, editorials, consensus statements, letters, recent text books and conference proceedings were reviewed for studies that could be included. In order to update the literature, alerts were established in PubMED/MEDLINE, EMBASE, ZETOC and current additions to the literature were added throughout the conduct of the systematic review.

2.11 Data selection and coding

The initial literature search results were examined independently by two reviewers (Oladejo Olaleye and Esmond Carr). Pre-specified inclusion and exclusion criteria were applied and potentially relevant papers were selected by both reviewers separately. Conference abstracts

or citations were selected if their title, abstract or key words suggested inclusion of relevant information on prevalence and persistence of oral HPV infection.

The two reviewers then independently reviewed each retrieved article to formally appraise it against the pre-specified inclusion and exclusion criteria. All articles selected and excluded by each reviewer were then compared. Any disagreements regarding article inclusion were resolved by consensus or if required, with the assistance of a third reviewer (Jayne Wilson or Hisham Mehanna).

At every screening or selection stage, the reason(s) for exclusion of any study were documented. Reference lists of excluded manuscripts were however searched for any relevant trials or studies.

2.12 Data extraction and management:

For each included study or review, data was collected including year of data collection, year of publication, study design, geographic location of study, patient / participant demographics (mean age, median age, gender, number of males and females in total, number of males and females with oral HPV infection), risk factor exposure such as cigarette smoking, alcohol consumption, oral sex, french kissing, multiple sexual partners. Oral HPV positive data was sub-divided into overall, high-risk HPV, HPV 16, and HPV 18 total numbers and proportions.

Data on oral HPV persistence were extracted from all studies that retested participants after a period of time from the index test using any HPV detection techniques. Oral HPV detection techniques used in each study reporting prevalence were collected as well as the sample type e.g. oral rinse, oral swab / brush, pharyngeal swab / brush, tonsil specimens, exfoliated samples, or biopsies. A pre-determined data extraction form was used to extract data from

each included study. Differences were resolved by consensus or with the help of an independent third expert (Jayne Wilson or Hisham Mehanna). Whenever it was required, further information was sought from the authors of studies.

2.13 Risk of bias (quality) assessment

All included studies were assessed critically for methodological quality, internal and external validity by two independent reviewers. For included systematic reviews, the checklist used was AMSTAR (A Measurement Tool to Assess Systematic Reviews) www.amstar.ca. Observational studies were also assessed for quality.

To explore the quality, internal/external validity and overall strength of the evidence included, the following were considered:

- Quality of randomised controlled trials including the study designs. Strength of the evidence where non-RCTs were included

- Potential sources of bias including publication bias, publication status (i.e. published reports vs grey literature), language bias and sources of funding.

2.14 Strategy for data synthesis

2.14.1 Statistical analysis:

Depending on the quantity, quality and statistical/clinical homogeneity of the included studies, pooled estimates of oral HPV prevalence and persistence in healthy populations, by gender, detection technique or samples were computed in order to produce a quantitative meta-analysis for each domain. The choice of a fixed or random effects model was based on the degree of homogeneity assessed by a statistician (Piers Gaunt). A narrative systematic

review was conducted when the obtained results for any of the domains could not be appropriately combined.

STATA was used to analyze the dichotomous (binary) data obtained from this review. The analyses was expressed as pooled prevalence and persistence, with accompanying 95% confidence intervals. Separate analysis of oral HPV prevalence by detection technique was performed including for HPV DNA by PCR, ELISA, p16 immunohistochemistry. Separate analysis of oral HPV persistence by tissue type was also undertaken for samples grouped as oral rinse, oral swab/brush, pharyngeal brush/swab, or tonsil specimens.

2.14.2 Heterogeneity:

Heterogeneity was tested using a Chi-squared test with (N - 1) degrees of freedom, where N equals the number of trials from which data was included (Petitti et al., 2001). Standard forest plots were also be used to explore the observed heterogeneity graphically.

2.14.3 Publication bias:

Publication bias was explored with the use of the inverted funnel plot technique. Specific tests of publication bias were utilised as appropriate (Egger et al., 1997).

2.14.4 Reporting of results:

The systematic review was conducted and reported based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

2.14.5 Analysis of subgroups or subsets:

Subgroup analyses of oral HPV prevalence and persistence was performed where possible and

appropriate, to investigate the differential prevalence and persistence in age-groups and the associations with risk factors such as cigarette smoking, alcohol consumption, and sexual practices. Subgroup analysis of the various screening tests currently available for oral HPV testing were performed where appropriate.

2.14.6 Dissemination plans:

We aim to publish our findings in a leading peer-reviewed journal with a focus on Ear, Nose and Throat Surgeons, Epidemiologists, Virologists, General Practitioners, Public Health physicians and Pharmaceuticals. We have also disseminated our findings through oral and poster presentations at both national and international conferences including ENT UK and the European Congress on Head and Neck Oncology (ECHNO), as well as on the InHANSE website.

2.15 Meta-analysis results:

The flow chart of this systematic review and meta-analysis is shown in Figure 2.1.

4,389 abstracts were initially identified and after removal of duplicates, screening against inclusion criteria and exclusions, 159 studies were included in the qualitative analysis and 45 studies in the quantitative synthesis (Figure 2.1).

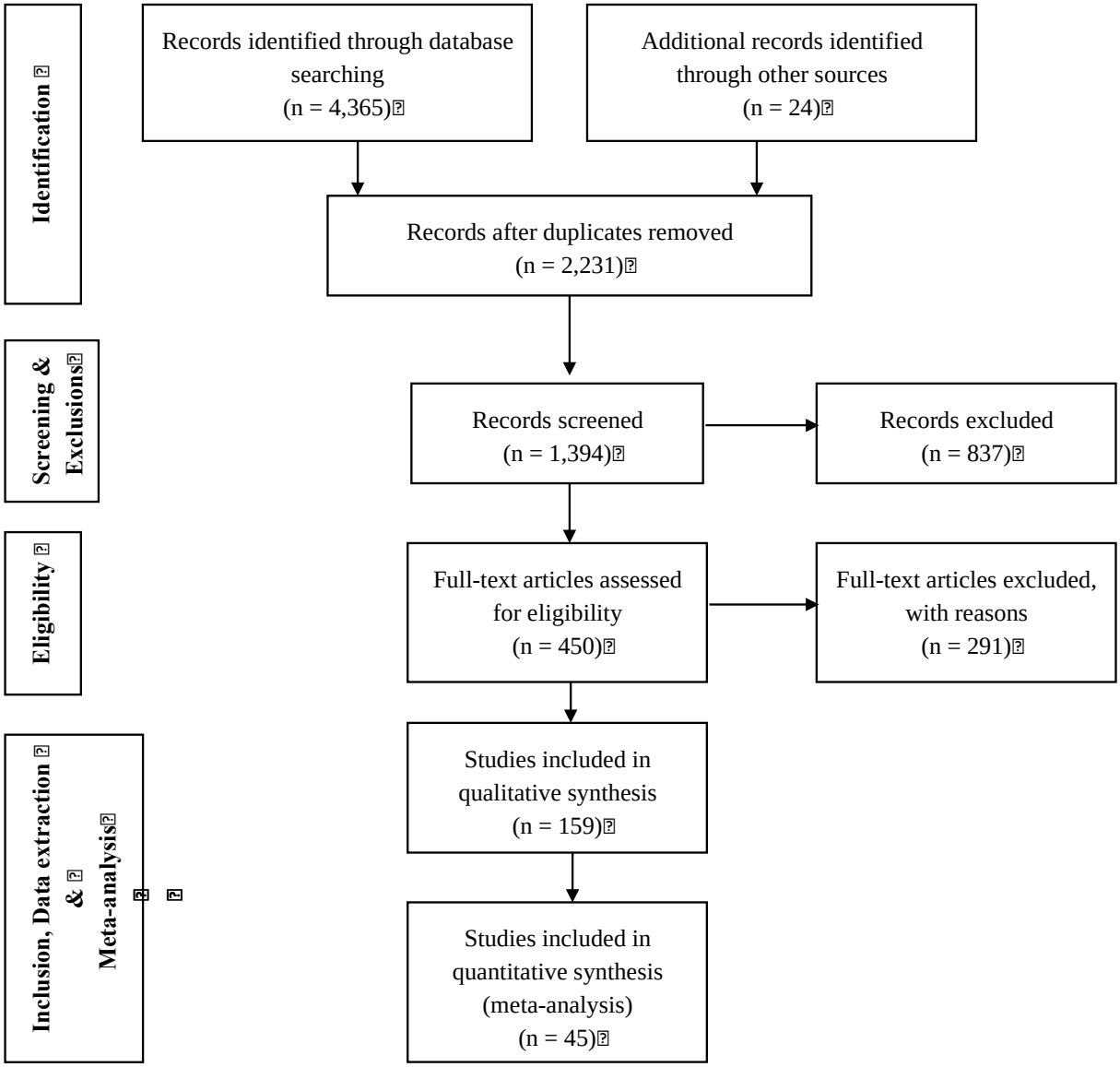


Figure 2.1: Systematic review flow chart. *The flow chart shows the number of abstracts included, number of exclusions and final number of abstracts in the meta-analysis.*

45 studies were included in this systematic review and meta-analysis on oral HPV prevalence in healthy individuals. This analysis included cross-sectional studies (n=33), cohort studies (n=10), randomised controlled trial (n=1) and cluster randomised study (n=1). There were 42 prospective studies and 3 retrospective studies. A total of 29,050 healthy individuals (cancer-free) were included from a population with age range of 12years – 89years, with a worldwide distribution. Population sizes varied from 50 subjects (the least) to 5501 subjects (the maximum size). Of the 45 included studies, 36 studies reported on at least 100 or more healthy individuals, with 11 of these studies reporting on over 500 participants and remarkably, 7 studies with 1000 or more healthy subjects. Two studies (Gillison et al., 2012; Hang et al., 2014) accounted for over 5,000 participants in each study. Table 2.1 shows the included 45 studies, the year of study, total sample size, geographic location, study types and % positivity for any HPV, high-risk HPV and HPV 16. Figure 2.2 shows the distribution of included studies and their respective population sizes.

The meta-analysis of HPV proportions in each study was conducted using a random effects model. For interpretation, each meta-analysis transcript has the included studies, the effect size (ES) and the 95% confidence intervals. There is also the % weighting assigned to each study dependent on the sample size and the statistical corrections inherent in the model. The overall HPV proportion output in each figure is the diamond right at the bottom.

Table 2.1: Studies included in oral HPV prevalence review in healthy individuals.

1st Author	Year of study	Total Sample size	Country	Study Type	Any HPV Positive (%)	High risk HPV (%)	HPV 16 positive (%)
Gillison	2012	5501	USA	Cross-sectional	6.9	3.7	0.9
Hang	2014	5351	China	Cross-sectional	6.2	0.5	0.43
Kuhs	2013	2926	Costa Rica	Randomised controlled trial	1.9	1.3	0.4
Durzynska	2011	2837	Poland	Cross-sectional	1.3	0	0
Kreimer	2011	1680	USA	Cross-sectional	4	1.3	0.6
Cook	2014	1010	USA	Cross-sectional	1.9	0.6	0.2
Pickard	2012	1000	USA	Cohort	2.4	1.2	0.2
Rosen	2016	980	Peru	Cross-sectional	7.35	2.24	1.22
Kurose	2004	662	Japan	Cohort	0.6	0.15	0.15
Smith	2007	536	USA	Cross-sectional	2.6	0.93	0.75
Machado	2014	514	Brazil	Cross-sectional	1.36	0.39	0
Pedersen	2012	442	Denmark	Cross-sectional	0.7	0.7	0.45
Fu	2015	409	USA	Cohort	2.4	2.2	0.7
Gonzalez-Losa Mdel	2015	383	Mexico	Cross-sectional	14.8	2.35	0.78
Nordfors	2013	335	Sweden	Cross-sectional	1.8	1.8	1.2
Rautava	2012	324	Finland	Cohort	17	14.5	10.5
Forslund	2013	311	Sweden	Cross-sectional	5.7	0.3	0.3
Antonsson	2014	307	Australia	Cross-sectional	2.3	1.3	0.33
Ernster	2009	226	USA	Cohort	0	0	0
Edelstein	2012	212	USA	Cohort	7.5	7.5	2.8
D'Souza	2009	210	USA	Cross-sectional	2.9	2.4	0.48
Bekker	2012	200	USA	Cohort	1	Not available	Not available
Wimardhani	2015	200	Indonesia & Thailand	Cross-sectional	6	6	4
Klingenburg	2010	195	Netherlands	Cross-sectional	1.03	1.03	1.03
Marais	2006	194	South Africa	Cross-sectional	4.1	1.03	0.52
Gichki	2012	192	Pakistan	Cross-sectional	24.5	7.8	2.1
Araujo	2014	166	Brazil	Cross-sectional	24.1	3.6	0
Turner	2011	151	USA	Cluster randomized	2.6	2.6	2.6
Kero	2012	131	Finland	Cohort	18.3	9.9	6.1
Davidson	2014	125	South Africa	Cross-sectional	5.6	1.6	0.8
Tristao	2012	125	Brazil	Cross-sectional	23.2	Not available	Not available
Cavenaghi	2013	124	Brazil	Cross-sectional	2.4	0.8	0
Knight	2016	124	United Kingdom	Cross-sectional	4	Not available	Not available
Seifi	2013	114	Iran	Cross-sectional	6.1	5.3	0
Esquenazi	2010	100	Brazil	Cross-sectional	0	0	0
Arbabi-Kalati	2014	100	Iran	Cross-sectional	0	0	0
Summersgill	2001	97	USA	Cohort	5.2	3.1	3.1
Schlecht	2013	93	USA	Cohort	12	4.3	2
Ruiz	2009	85	Venezuela	Cross-sectional	8.2	Not available	Not available
Migaldi	2012	81	Italy	Cross-sectional	1.2	0	0
Gonzalez-Losa Mdel	2008	77	Mexico	Cross-sectional	1.3	1.3	Not available
Eike	1995	61	Denmark	Cross-sectional	0	0	0
Shindoh	1994	56	Japan	Cross-sectional	3.6	1.8	1.8
do Sacramento	2006	50	Brazil	Cross-sectional	14	10	4

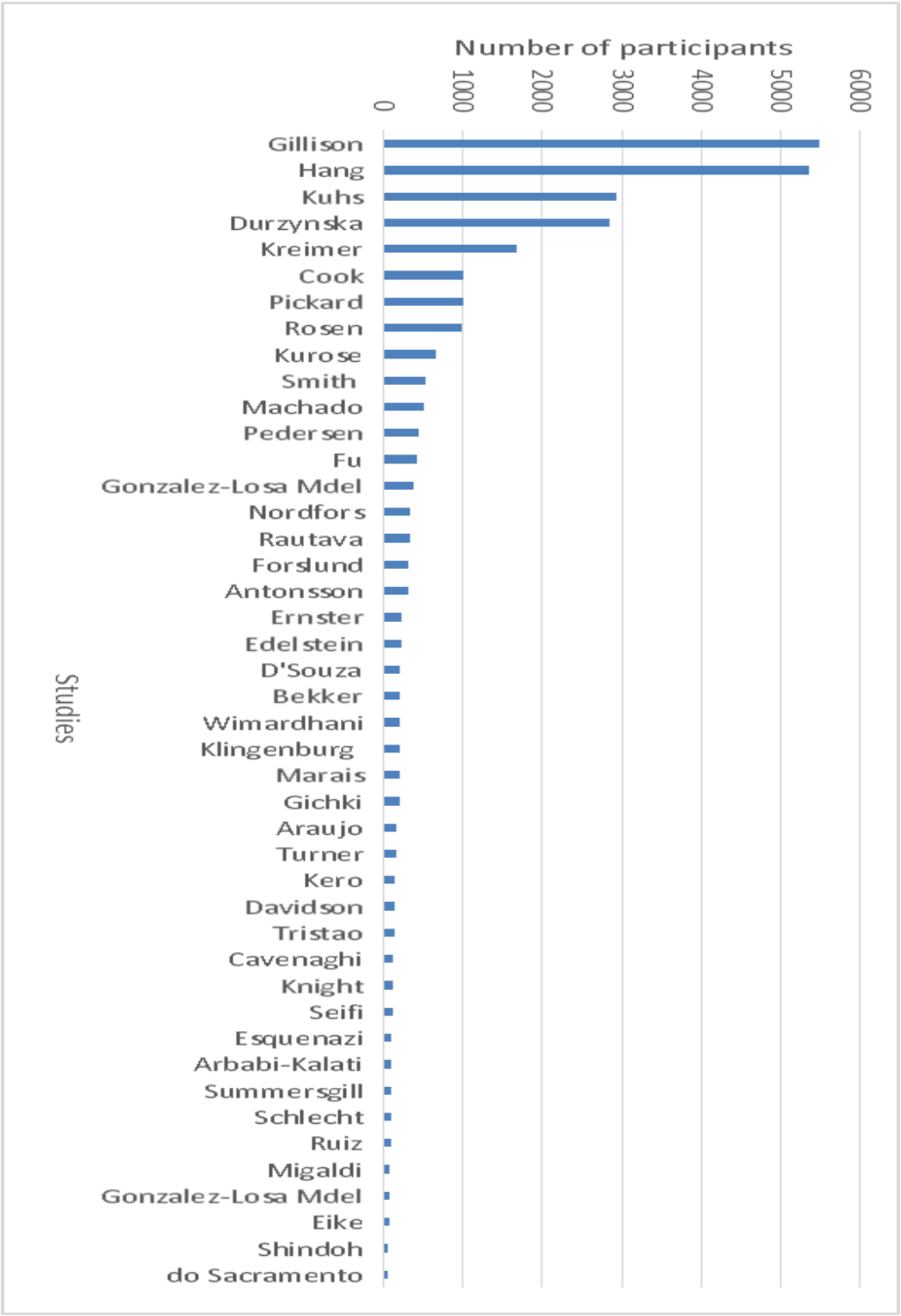


Figure 2.2: Studies reporting oral HPV prevalence in healthy populations. *Studies with at least 50 healthy participants included with the highest number of participants in 2 studies with over 5,000 participants (Gillison et al., 2012 and Hang et al., 2014).*

There were 1,442 healthy individuals positive for any HPV type (high-risk or low-risk) out of the 29,050 healthy individuals included. This is a pooled prevalence of any HPV type of 4.9% (95% CI: 3.7% - 6.3%) in healthy subjects (Figure 2.3). This pooled prevalence is further stratified by gender, geography, and oral sample type and detection technique in our subsequent analyses.

2.15.1 Oral HPV detection techniques:

43 out of the 45 studies included in this review used PCR based detection methods for HPV detection. This was used either alone or in combination with other techniques such as p 16 immunohistochemistry, in-situ hybridisation and type-specific PCR. Hybrid capture was used in 2 studies with the limitation that they could not report on the specific HPV genotypes identified.

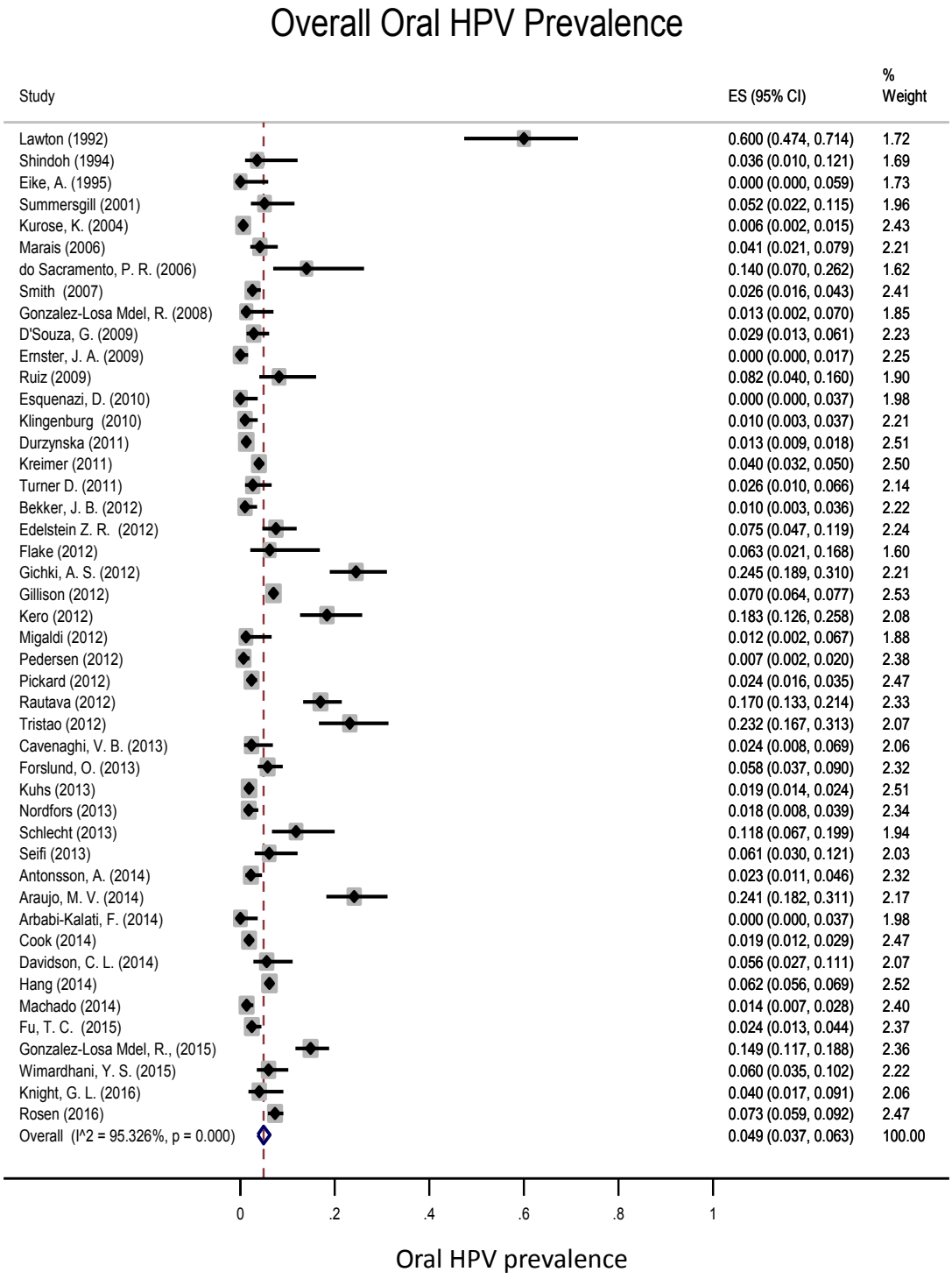


Figure 2.3: Pooled overall oral HPV prevalence. *The pooled overall prevalence for any oral HPV infection was calculated using STATA and presented as a proportion with effect size (ES) and 95% confidence intervals depending on the weighting assigned based on sample size. The x-axis measures the effect size (oral HPV prevalence) in proportions from 0 – 1.*

There were 536 healthy individuals with oral high-risk HPV from a total sample size of 28,516 individuals available from all studies. High-risk HPV included genotypes 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 66. The pooled prevalence of high-risk oral HPV (carcinogenic HPV) was therefore 1.8% (95% CI: 1.2% - 2.6%) as shown in Figure 2.4. (Rautava et al., 2012) reported the highest prevalence of 14.5% (95% CI: 11.1% - 18.8%) in Finland. This study was a prospective cohort study on 324 pregnant women on the Finnish Family Study who were followed up for 6 years. 47 of these pregnant women tested positive on nested PCR for high-risk oral HPV at baseline.

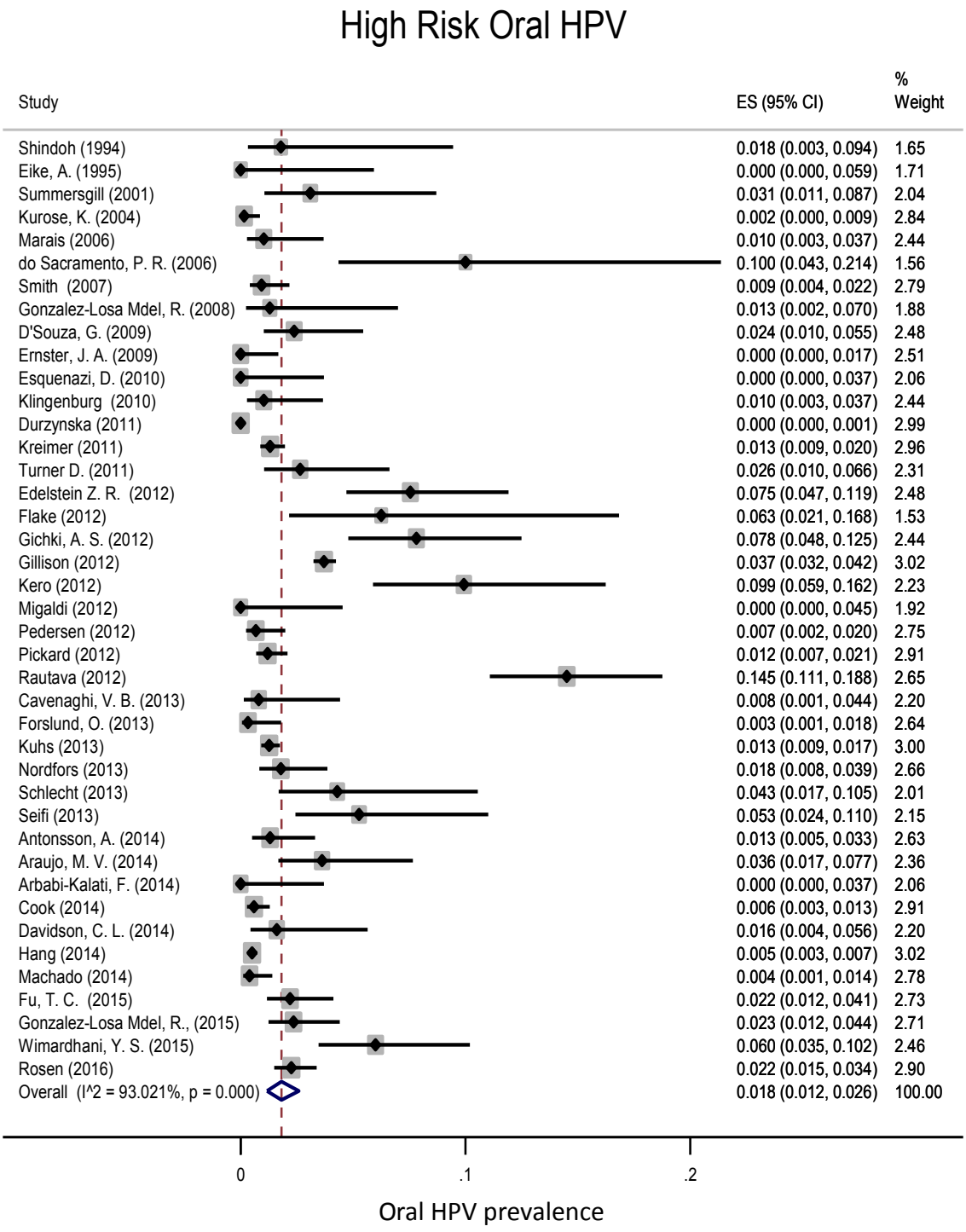


Figure 2.4: High risk oral HPV prevalence in healthy populations. *The pooled prevalence for high-risk oral HPV infection was calculated using STATA and presented as a proportion with 95% confidence intervals depending on the weighting assigned based on sample and effect size. The pooled prevalence of high-risk oral HPV was 1.8% (95% CI: 1.2% - 2.6%).*

Oral HPV 16 accounts for over 90% of HPV positive head and neck squamous cell carcinomas. It is therefore important to establish its prevalence in healthy individuals. In this meta-analysis, Oral HPV 16 pooled prevalence in healthy individuals was 0.7% (95% CI: 0.4% - 1.0%) (Figure 2.5). Studies that reported oral HPV 16 prevalence were included and showed 221 individuals positive out of a total sample size of 28,439 individuals. Rautava et al., (2012) similarly reported the highest oral HPV 16 baseline prevalence in their Finnish cohort of 10.5% (95% CI: 7.6% - 14.3%). The majority of studies however reported oral HPV 16 prevalence of less than 2.0%.

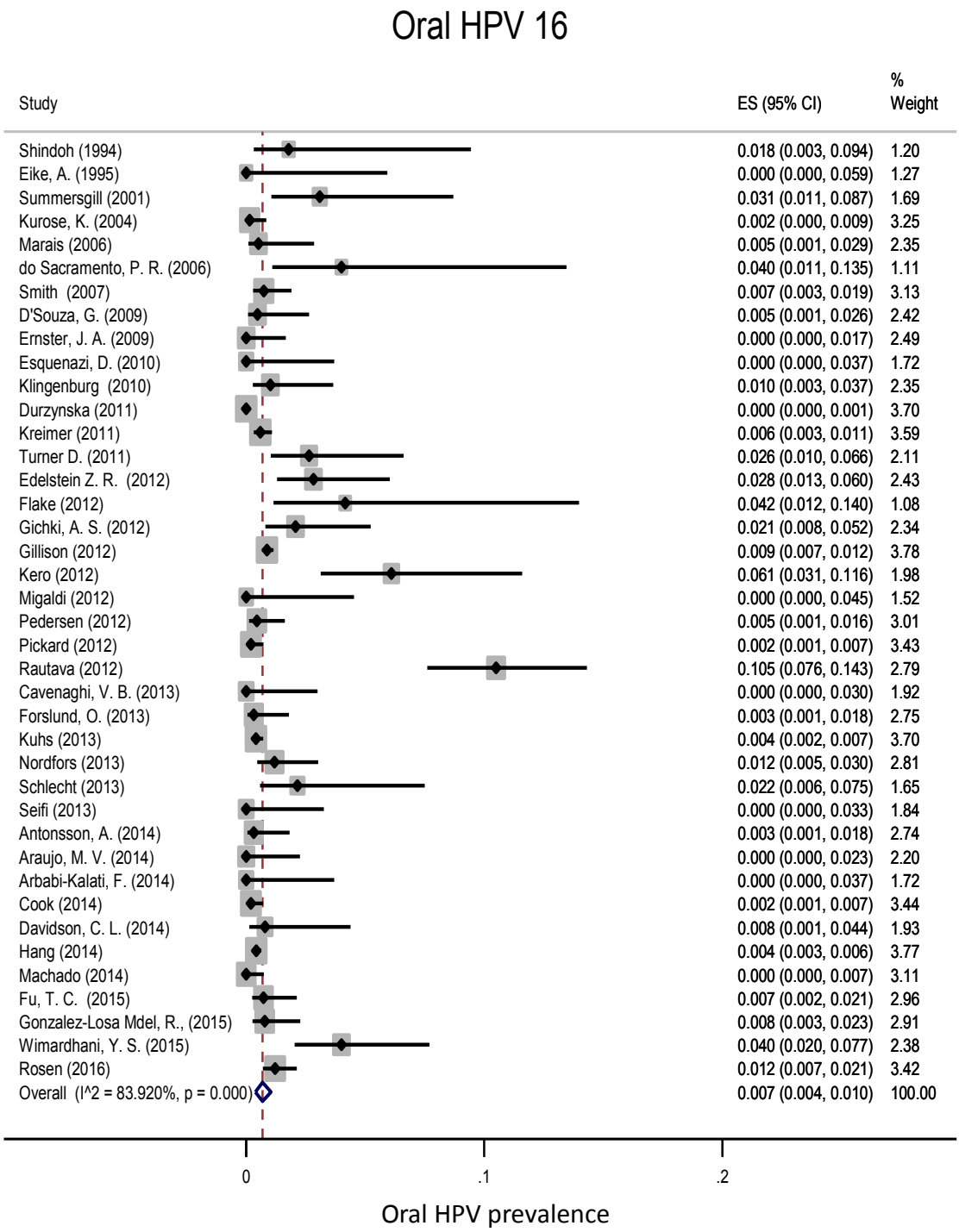


Figure 2.5: Pooled oral HPV 16 prevalence in healthy populations. *The pooled prevalence for oral HPV 16 infection was calculated using STATA and presented as a proportion with 95% confidence intervals depending on the weighting assigned based on sample and effect size. The pooled prevalence of oral HPV 16 was 0.7% (95% CI: 0.4% - 1.0%).*

Further estimates of oral HPV 18 were extracted from studies that provided the information. There were 66 out of 28,439 healthy individuals positive for oral HPV 18. The pooled prevalence of oral HPV 18 was thus 0.1% (95% CI: 0.0% - 0.3%) (Figure 2.6). The majority of the included studies reported very low oral HPV 18 prevalence of <1.0%, with the highest prevalence of 5.7% (95% CI: 3.2% - 10.0%) reported in Pakistan (Gichki et al., 2012). The second highest prevalence for oral HPV 18 was 4.4% (95% CI: 1.9% - 9.9%) reported by (Seifi et al., 2013) in Iran.

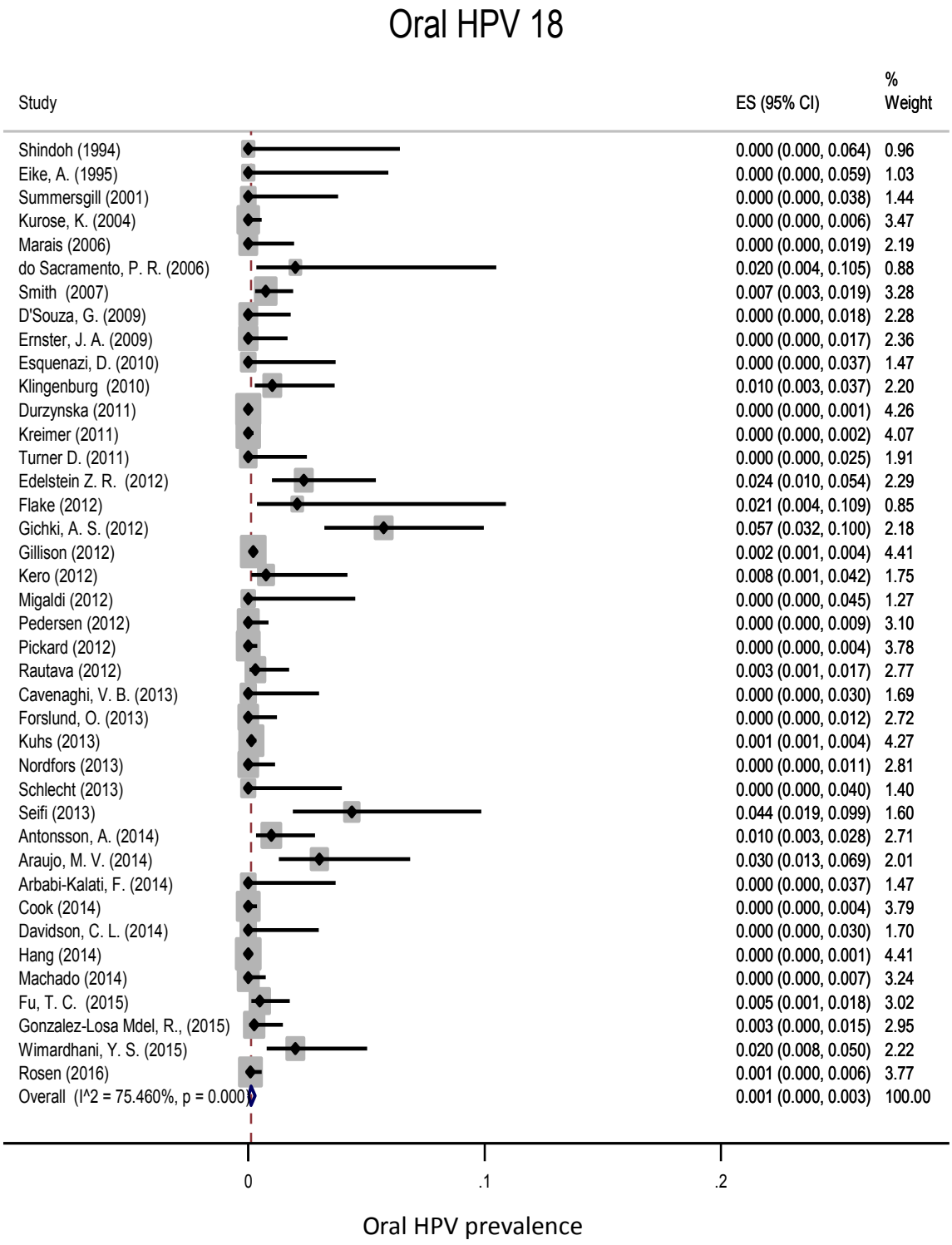


Figure 2.6: Pooled oral HPV 18 prevalence in healthy populations. *The pooled prevalence for oral HPV 18 infection was calculated using STATA and presented as a proportion with 95% confidence intervals depending on the weighting assigned based on sample and effect size. The pooled prevalence of oral HPV 18 was 0.1% (95% CI: 0.0% - 0.3%).*

2.16 Regional variation in oral HPV prevalence by geographic location

The impact of geographic location on oral HPV prevalence (high-risk or low-risk) was investigated by grouping studies from the same WHO geographic region together and calculating regional pooled prevalence. Any variability in pooled prevalence between geographic regions was further assessed. The 45 included studies were divided into the following world regions as shown in Figure 2.7: North America (n=13), South America (n=8), Central America (n=3); Europe / Scandinavia (n=6), Western Europe (n=3), Eastern Europe (n=1); Asia (n=2), Far East (n=3), Middle East (n=2); Africa (n=2); Australasia (n=2).

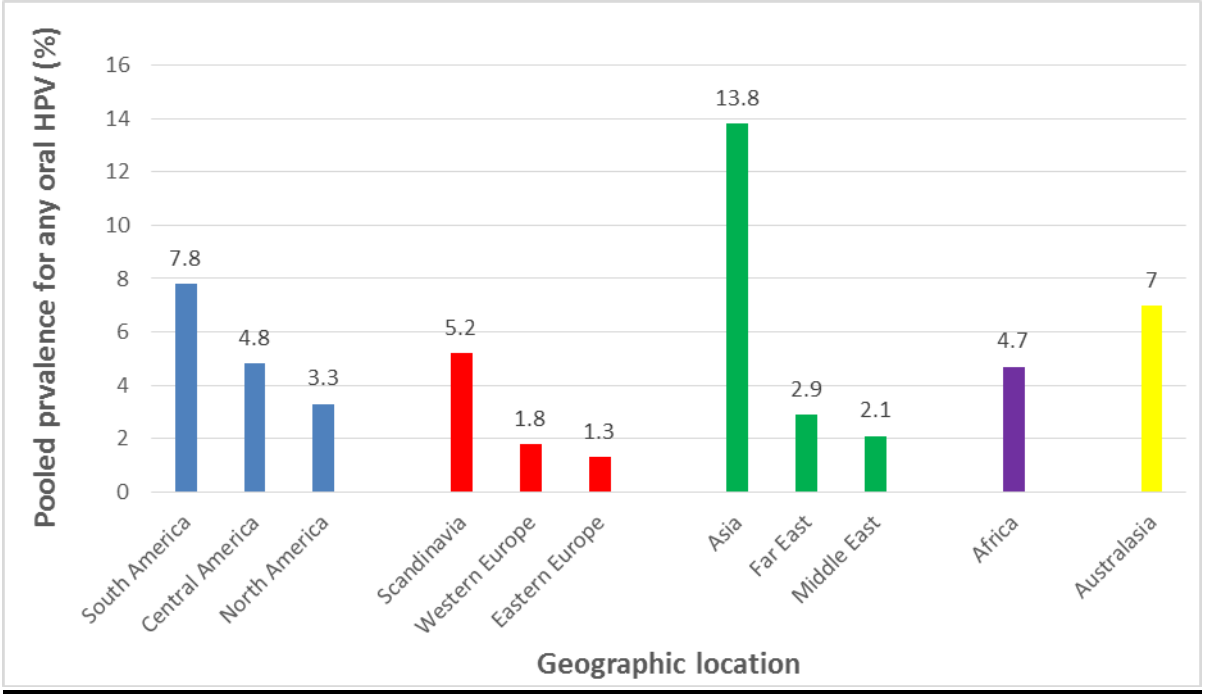


Figure 2.7: Pooled prevalence for any oral HPV infection in healthy populations by geographic region. *The pooled prevalence by geographic region for any oral HPV infection was calculated by amalgamating all the studies from each region and presented as proportions. The studies from the American continent coloured in blue, included 13 studies from North America, 8 from South America, and 3 from Central America. In Europe, studies are coloured in red with 6 from Scandinavia, 3 from Western Europe and 1 from Eastern Europe. Studies from Asia are coloured green with 2 studies from Asia, 3 from the Far East and 2 from the Middle East. There were 2 studies from Africa coloured purple and 2 studies from Australasia coloured in yellow.*

In ***North America*** (n=13 studies), the pooled prevalence of any oral HPV is 3.3% (95% CI: 2.0% - 4.9%) (Figure 2.7 and Table 2.2). The included studies (in time order) were by (Summersgill et al., 2001), (Smith et al., 2007), (D'Souza et al., 2009), (Ernster et al., 2009), (Kreimer et al., 2011), (Turner et al., 2011), (Bekker et al., 2012), (Edelstein et al., 2012), (Flake et al., 2012), (Gillison et al., 2012), (Pickard et al., 2012), (Schlecht et al., 2013), (Cook et al., 2014), and (Fu et al., 2015). The largest study in North America was the Gillison et al (2012) cross-sectional study of 5,501 healthy American men and women aged 14years – 69 years, who had oral rinse / gargle testing for any oral HPV by multiplex PCR and type-specific hybridisation, with 385 participants positive. They reported a prevalence of 6.9% for any oral HPV in a population representative for the background American population using the National Health and Nutrition Exercise Survey (NHANES). Heterogeneity in the North American series was I^2=91.98%, p=0.000.

In ***South America***, 8 studies were available with 6 prospective cross-sectional studies from Brazil (do Sacramento et al., 2006; Esquenazi et al., 2010; Tristao et al., 2012; Cavenaghi et al., 2013; Araujo et al., 2014; Machado et al., 2014), one large prospective cross-sectional study with 980 individuals from Peru (Rosen et al., 2016) and one prospective cross-sectional study of 85 people from Venezuela (Ruiz et al., 2009) (Figure 2.7 and Table 2.2).

In the South American studies, the pooled prevalence for any oral HPV was 7.8% (95% CI: 2.9% - 14.6%). Two studies that accounted for the largest effect sizes were (Araujo et al., 2014) that reported a prevalence of 24.1% (95% CI: 18.2% - 31.1%) and (Tristao et al., 2012) that reported a prevalence for any oral HPV of 23.2% (95% CI: 16.7% - 31.3%). Both studies (Tristao et al., 2012; Araujo et al., 2014) were conducted in Brazil and were partly responsible for the relatively higher oral HPV prevalence in the region. In contrast, there were three other

studies conducted in Brazil with oral HPV prevalence lower than the 4.9% pooled prevalence in the meta-analysis (Esquenazi et al., 2010; Cavenaghi et al., 2013; Machado et al., 2014). Esquenazi et al., (2010) analysed oral scrapings from 100 healthy individuals (40 males, 60 females) by PCR and found 0% prevalence (95% CI: 0% - 0.4%). Machado (2014) undertook a cross-sectional study of 514 asymptomatic men and found 7 oral brushings positive on PCR with a prevalence of 1.4% (95% CI: 0.7% - 2.8%). Cavenaghi et al. (2013) also reported oral HPV prevalence of 2.4% (95% CI: 0.8% - 6.9%) in a sample size of 124 individuals in Sau Paulo. Overall heterogeneity for the South American cohort was I^2=95.01%, p=0.000.

The pooled prevalence of oral HPV in ***Central America*** was 4.8% (95% C.I: 0.0 – 16.3%). (Figure 2.7 and Table 2.2). The three studies included were Gonzalez-Losa Mdel et al. (2008), Kuhs et al. (2013) and Gonzalez-Losa et al. (2015). Gonzalez-Losa Mdel et al. (2008) used hybrid capture 2 technique to identify oral HPV in oral brushings of 77 participants in Mexico and found a prevalence of 1.3% (95% C.I: 0.2 – 7.0%). There was no type-specific prevalence possible with this hybrid capture 2 technique. Gonzalez-Losa et al. (2015) study however used nested PCR with MY09/11, GP5+/6+ primers and sequencing to characterise oral HPV in a cross-sectional prospective study of 383 women from Yucatan, Mexico and found a prevalence of 14.9% (95% C.I: 11.7 – 18.8%). Kuhs et al. (2013) was a randomized controlled trial on 2,926 women in Costa Rica allocated to the control arm of a HPV vaccine trial. 55 women aged 18 – 25years old, were found to have oral HPV at the end of the trial on oral rinse / gargle using HPV SPF10 PCR-DEIA (DNA enzyme immunoassay)-LiPA25. This indicated an oral HPV prevalence of 1.9% (95% C.I: 1.4 – 2.4%). The results published in this RCT (Kuhs et al., 2013), are similar to the data published by Herrero et al. (2013) and Katki et al. (2013) from the same trial and population. These latter two publications focused

on HPV vaccine efficacy rather than prevalence and were excluded to avoid duplication. Overall heterogeneity in this *Central America* cohort was I^2=97.81%, p=0.000.

In ***Europe / Scandinavia,*** the pooled prevalence for any oral HPV infection was 5.2% (95% C.I: 0.9 – 12.5%) (Figure 2.7 and Table 2.2). The studies included were Eike et al. (1995), Kero et al. (2012a), Pedersen and Hariri (2012), Rautava et al. (2012), Forslund et al. (2013), and Nordfors et al. (2013). There were two studies that accounted for a higher oral HPV prevalence in this Europe / Scandinavia group and both were published from the same group working on the Finnish Family Study (Kero et al., 2012b; Rautava et al., 2012). Kero et al. (2012a) reported oral HPV prevalence of 18.3% (95% C.I: 12.6 – 25.8%) in 131 male partners of pregnant women on a prospective cohort study over a 7 year period. Oral HPV prevalence fluctuated between 15.1% and 31.1% in oral scrapings with cytobrush using nested PCR with MY09/11, GP5+/6+ primers and Multimetrix kit (Progen Biotechnik, Heidelberg, Germany). Similarly Rautava et al. (2012) reported oral HPV prevalence of 17.0% (95% C.I: 13.3 – 21.4%) in the same prospective cohort but only on the 324 pregnant women in the Finnish family Study over a 6 year period. Data from these two studies are therefore correlated and accounted for higher oral HPV prevalence rates from both the pregnant women and their male partners when compared to the other studies in the region. The other studies from Scandinavia in this series were Eike et al. (1995) from Denmark reporting any oral HPV prevalence 0% in 61individuals; Pedersen and Hariri (2012) from Denmark reporting 0.7% prevalence in 442 individuals; Forslund et al. (2013) from Sweden reporting 5.7% prevalence in 311 individuals; and Nordfors et al. (2013) from Sweden reporting oral HPV prevalence of 1.8% in 335 high school students. Overall heterogeneity for this Europe / Scandinavia cohort was I^2=96.08%, p=0.000.

In ***Western Europe,*** pooled oral HPV prevalence in healthy individuals was 1.8% (95 C.I: 0.4 – 4.0%). The three studies included were Klingenburg et al. (2010) from the Netherlands, Migaldi et al. (2012) from Italy, and Knight et al. (2016) from the United Kingdom. Klingenberg et al. (2010) reported a prevalence of 1.0% (95% C.I: 0.3 – 3.7%) in 195 tonsils using PCR, p16 immunohistochemistry, and in-situ hybridization techniques. Migaldi et al. (2012) reported oral HPV prevalence in oral smears of 81 participants from a Northern Italian population of 1.2% (95% C.I: 0.2 – 6.7%). In the United Kingdom, Knight et al. (2016) in a prospective cross-sectional study of 124 young adults, found a prevalence of 4.0% (95% C.I: 1.7 – 9.1%). Overall heterogeneity in this *Western Europe* cohort was I^2=33.98%, p=0.220.

In ***Eastern Europe,*** the only study included was (Durzynska et al., 2011) which was a prospective cross-sectional study of 2,837 adolescents in Poland using PCR to identify HPV in oral rinse. The reported prevalence for any HPV was 1.3% (95% C.I: 0.9 – 1.8%) (Figure 2.7 and Table 2.2)

In the ***Far East,*** the pooled prevalence of oral HPV was 2.9% (95% C.I: 0.0 – 9.3%) (Figure 2.7 and Table 2.2). The 3 studies included were Shindoh et al. (1994), Kurose et al. (2004) and Hang et al. (2014). Shindoh et al. (1994) was a cross-sectional prospective study done in Japan on oral scrapes from 56 individuals using nested PCR and found oral HPV prevalence of 3.6% (95% C.I: 1.0 – 12.1%). Kurose et al. (2004) was a prospective cohort of 662 residents of Miyako Island, Japan. Buccal mucosa scrapes were analysed for oral HPV using consensus PCR with MY09/11 primers and found a low prevalence of 0.6% (95% C.I: 0.2 – 1.5%). Hang et al. (2014) is a recent cross-sectional prospective study done on a large cohort of 5410 healthy adults in China. 332 adults were positive for any oral HPV in oral cavity

swabs and exfoliated cells using general primer-mediated (SPF1/G6+) PCR and sequencing, resulting in a prevalence of 6.2% (95% C.I: 5.6 – 6.9%). Overall heterogeneity in this *Far East* cohort was I^2=97.06%, p=0.000.

In ***Asia,*** the two studies included were Gichki et al. (2012) and Wimardhani et al. (2015) with a pooled prevalence of 13.8% (95% C.I: 10.5 – 17.4%) (Figure 2.7 and Table 2.2). This prevalence was the highest when compared with the other world regions. A closer review of the studies indicates that (Gichki et al., 2012) reported the highest prevalence of all included studies in this review at 24.5% (95% C.I: 18.9 – 31.0%). This was a cross-sectional prospective study on 192 adults from Pakistan using real-time PCR. High-risk HPV was 7.8% while HPV 16 was 2.1% in this cohort. Wimardhani et al. (2015) conducted a cross-sectional prospective study on 200 female Indonesian and Thai Dental students, using PCR to analyse saliva samples for presence of HPV. The oral HPV prevalence was 6.0% (95% C.I: 3.5% - 10.2%). Overall heterogeneity in this *Asia* cohort was I^2=66.99%, p=0.082. The high oral HPV prevalence observed in the Asia cohort must be interpreted cautiously as the driver for the high rate comes from a particular study in Pakistan on 192 adults using real-time PCR (Gichki et al., 2012). This technique can be considered highly sensitive for HPV detection.

In the ***Middle East,*** the pooled prevalence for oral HPV was 2.1% (95% C.I: 0.5 – 4.7%). (Figure 2.7 and Table 2.2). The two studies in this group were (Seifi et al., 2013) and (Arbabi-Kalati et al., 2014). (Seifi et al., 2013) reported a prevalence of 6.1% (95% C.I: 3.0 – 12.1%) while (Arbabi-Kalati et al., 2014) reported oral HPV prevalence of 0% (95% C.I: 0 – 3.7%).

In ***Africa,*** pooled oral HPV prevalence was 4.7% (95% C.I: 2.5% - 7.3%) from 2 included studies – (Marais et al., 2006) and (Davidson et al., 2014). In ***Australasia,*** the pooled oral HPV prevalence was 7.0% (95% C.I: 4.5 – 9.9%) accrued from 2 studies – Lawton et al., 1992) and (Antonsson et al., 2014) (Figure 2.7 and Table 2.2).

Table 2.2: Oral HPV prevalence worldwide by geographic location. *Pooled prevalence for any oral HPV infection by geographic location is presented as % with 95% confidence intervals. The pooled prevalence was calculated from studies published in each world region.*

Geographic location	Oral HPV prevalence (%)	95% Confidence Intervals (%)
South America	7.8	2.9 – 14.6
Central America	4.8	0.0 – 16.3
North America	3.3	2.0 – 4.9
Scandinavia	5.2	0.9 – 12.5
Western Europe	1.8	0.4 – 4.0
Eastern Europe	1.3	0.9 – 1.8
Asia	13.8	10.5 – 17.4
Far East	2.9	0.0 – 9.3
Middle East	2.1	0.5 – 4.7
Africa	4.7	2.5 – 7.3
Australasia	7.0	4.5 – 9.9

2.17 Oral HPV prevalence by sample type:

The data included from the 45 studies included in this review was further stratified into the sample type in which HPV was detected. Sample type was broadly divided into the following:

a) Oral rinse / gargle / saliva
b) Scrapings / brushings / swabs
c) FFPE (Formalin Fixed Paraffin Embedded) tonsil tissue
d) Fresh frozen tonsils

Pooled prevalence for any oral HPV in the studies that collected oral rinse / gargle / saliva from participants was 3.5% (95% C.I: 2.4 – 4.9%) (Table 2.3). In the studies that collected scrapings / brushings / swabs from the oral cavity / oropharynx, the pooled prevalence was found to be higher at 5.9% (95% C.I: 3.4 – 8.9%) (Table 2.3). Ernster et al. (2009) undertook a retrospective analysis of 226 FFPE from tonsils using type-specific PCR to HPV 16 and 18. There was no single positive FFPE sample to oral HPV with a prevalence of 0.0% (95% C.I: 0.0 – 1.7%) (Table 2.3). There were two studies that reported HPV prevalence in tonsils (Klingenberg et al., 2010; Bekker et al., 2012). Klingenberg et al., (2010) performed a retrospective cross-sectional study in the Netherlands of 195 tonsils using PCR, p16 immunohistochemistry and in-situ hybridisation. They found only 2 tonsils positive for HPV with a prevalence of 1.0% (95% C.I: 0.3 – 3.7%). Similarly, Bekker et al. (2012) in America, undertook a retrospective analysis of 200 clinically benign tonsils using p16 immunohistochemistry, PCR and in-situ hybridisation techniques. They reported 2 tonsils positive for HPV with a prevalence of 1.0% (95% C.I: 0.3 – 3.6%) (Table 2.3).

Oral HPV positivity was highest in those studies that used scrapings / brushings / swabs at 5.9%; followed by those studies that analysed HPV presence in oral rinse / gargle / saliva at

3.5%; and low oral HPV positivity in actual tonsils at 1.0% or FFPE at 0.0% (Table 2.3).

Overall heterogeneity between these groups was 95.33%, p = 0.000.

Table 2.3: Pooled prevalence for any oral HPV infection by sample type. *The sample types in all studies were grouped as (oral rinse/ gargles/ saliva), (scrapings/ brushings/ swabs), Fresh frozen tonsils, and FFPE (Formalin Fixed Paraffin Embedded). The pooled prevalence for each sample type is presented as % with the corresponding 95% confidence intervals. The comparative literature for each sample type cited in section 1.7 is also shown but with a wide variation in sensitivity of different techniques.*

Sample Type	Pooled prevalence for any oral HPV (%)	95% confidence intervals	Comparative literature for any oral HPV (%)
Oral rinse / gargles / saliva	3.5	2.4 – 4.9	2.8 – 25
Scrapings / Brushings / Swabs	5.9	3.4 – 8.9	0 – 32
Fresh frozen tonsils	1.0	0.2 – 2.3	0 – 16
FFPE (Formalin Fixed Paraffin Embedded)	0.0	0.0 – 1.7	0 – 2

2.18 Oral HPV 16 prevalence:

Pooled prevalence of oral HPV 16 by geographic location of the included studies showed some variation in prevalence across different world regions. The overall pooled prevalence of oral HPV 16 worldwide was low at 0.7% (95% C.I: 0.4 – 1.0%). In ***North America,*** the pooled prevalence of oral HPV 16 from the 13 included studies was 0.6% (95% C.I: 0.3 – 1.1%). In the ***Far East***, this pooled prevalence from 3 studies (Shindoh et al., 1994; Kurose et al., 2004; Hang et al., 2014), was lower at 0.2% (95% C.I: 0.0 – 0.5%). Similarly, 2 studies from ***Central America*** (Kuhs et al., 2013, Gonzalez-Losa et al., 2015) indicated a pooled prevalence for oral HPV 16 of 0.4% (95% C.I: 0.2 – 0.7%). There was only one study from ***Eastern Europe*** (Durzynska et al., 2011) that reported a prevalence of oral HPV 16 of 0% (95% C.I: 0.0 – 0.1%) in oral rinse in a cross-sectional study of 2,837 teenagers aged 16 – 18 year old using PCR. There were 6 studies from ***South America*** (do Sacramento et al., 2006; Esquenazi et al., 2010; Cavenaghi et al., 2013; Araujo et al., 2014; Machado et al., 2014; Rosen et al., 2016) from which oral HPV 16 pooled prevalence of 0.2% (95% C.I: 0.0 – 1.1%) was calculated. The prevalence of oral HPV 16 was generally low in the South American series with only 2 studies reported prevalence above the global average of 0.7%. (do Sacramento et al., 2006) reported the highest oral HPV prevalence of 4.0% (95% C.I: 1.1 – 13.5%) while Rosen et al. (2016) reported oral HPV 16 prevalence of 1.2% (95% C.I: 0.7 – 2.1%).

In ***Europe / Scandinavia,*** the pooled prevalence of oral HPV 16 from the included 6 studies (Eike et al., 1995; Kero et al., 2012a; Pedersen and Hariri, 2012; Rautava et al., 2012; Nordfors et al., 2013; Forslund et al., 2013) was 2.1% (95% C.I: 0.1 – 5.8%). This prevalence was higher than the average reported in other world regions. The two studies that accounted for the high oral HPV 16 rates were published by Kero et al. (2012a) that reported 6.1% (95%

C.I: 3.1 – 11.6%) and Rautava et al. (2012) reporting 10.5% prevalence (95% C.I: 7.6 – 14.3%). It has been previously noted that these two studies are correlated as they represent data from the Finnish Family Study. Rautava et al. (2012) reported prevalence in 324 pregnant women over a six-year period whilst Kero et al. (2012a) reported prevalence in 131 of their male partners over a 7 year period during which point prevalence for any oral HPV fluctuated between 15.1 – 31.1%. Both studies utilised nested PCR with MY09/11, GP5+/6+ primers and Multimetrix kit (Progen Biotechnik).

Antonsson et al. (2014) in ***Australasia*** reported an oral HPV 16 prevalence of 0.3% (95% C.I: 0.1 – 1.8%). In ***Western Europe,*** 2 studies reported oral HPV 16 prevalence in healthy individuals (Klingenberg et al., 2010; Migaldi et al., 2012) with a pooled prevalence of 0.5% (95% C.I: 0.0 – 2.0%).

In ***Asia,*** the pooled prevalence for oral HPV 16 was 3.0% (95% C.I: 1.5 – 5.0%) obtained from two studies (Gichki et al., 2012; Wimardhani et al., 2015). In ***Africa,*** there were only 2 studies (Marais et al., 2006; Davidson et al., 2014), with a pooled prevalence of 0.6% (95% C.I: 0.0 – 1.9%). There was no reported detection of oral HPV 16 in the 2 studies in the Middle East (Seifi et al., 2013; Arbabi-Kalati et al., 2014), giving a pooled prevalence of 0% (95% C.I: 0.0 – 0.9%). Overall heterogeneity between the various world regions, $I^2 = 83.92\%$, p=0.000.

2.19 Oral HPV prevalence in males:

A meta-analysis of the data in this systematic review showed an overall oral HPV prevalence in males for any HPV genotype was 4.1% (95% C.I: 2.4 – 6.1%) (Figure 2.8). The vast majority of the studies that included a gender distribution in their publication reported oral HPV prevalence in males close to this pooled prevalence of 4.1%. There were outlier studies that reported higher oral HPV prevalence though, as shown in Figure 2.8 including Flake et al. (2012), Gichki et al. (2012), Kero et al. (2012a), and Tristao et al. (2012). These studies may have inherent flaws in their design, variations in the HPV detection technique utilised, or differences in their baseline population characteristics.

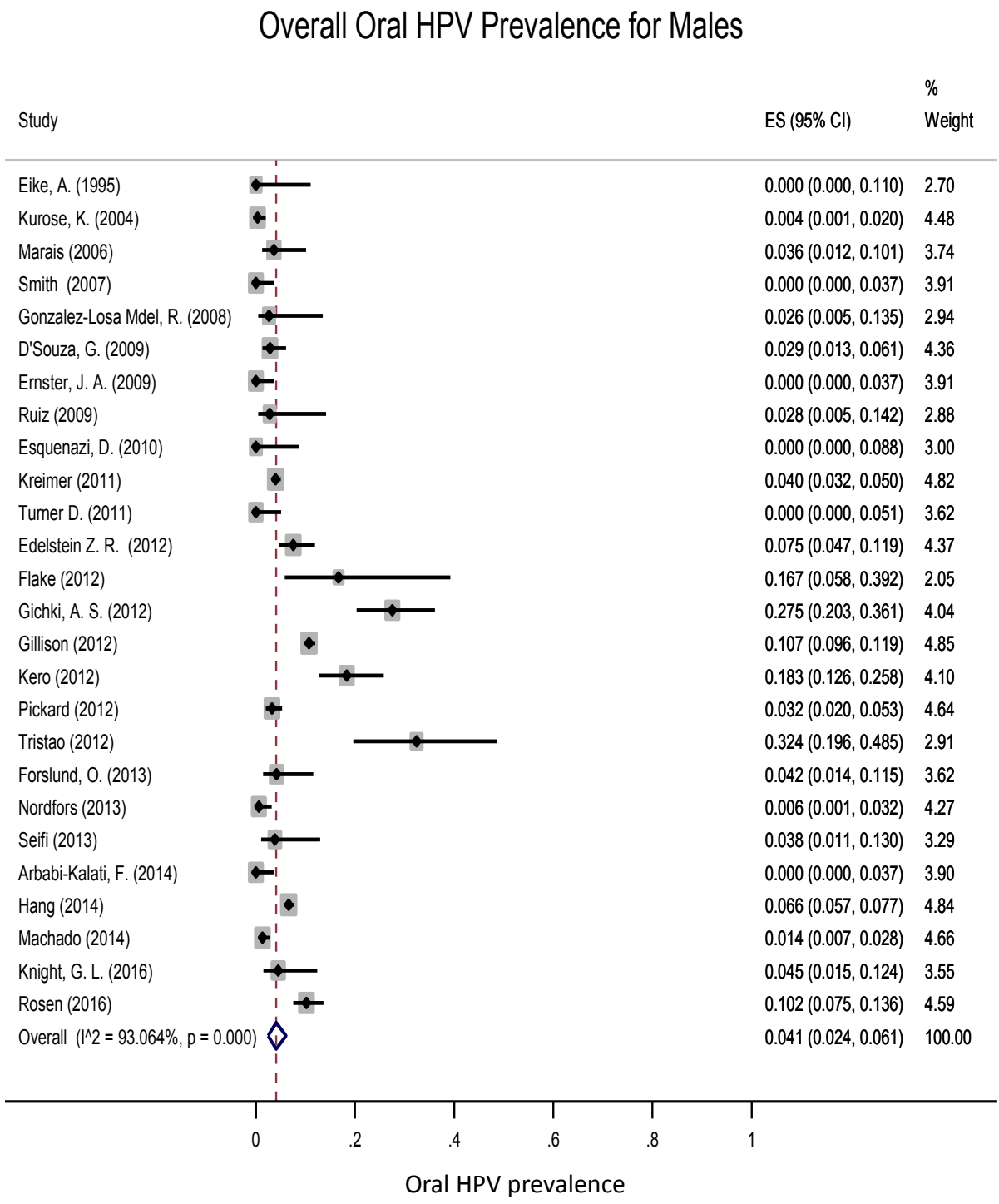


Figure 2.8: Pooled prevalence for any oral HPV infection in healthy males. *The pooled prevalence for any oral HPV infection in males was calculated using STATA and presented as a proportion with 95% confidence intervals depending on the weighting assigned based on sample and effect size. The pooled prevalence for any oral HPV in males was 4.1% (95% C.I: 2.4 – 6.1%).*

2.20 Oral HPV prevalence in females:

In females, the pooled oral HPV prevalence for any HPV genotype worldwide in those studies reporting a gender distribution was 4.5% (95% C.I: 3.1 – 6.1%) (Figure 2.9). Some studies had higher prevalence reported (Gichki et al., 2012; Rautava et al., 2012; Ruiz et al., 2009; Tristao et al., 2012; Schlecht et al., 2013; Gonzalez-Losa et al., 2015). This is comparable to the pooled prevalence of 4.1% shown in males. There was no statistically significant difference in the pooled prevalence of any oral HPV between males and females in this meta-analysis on Chi-square (X^2) testing for independent variables.

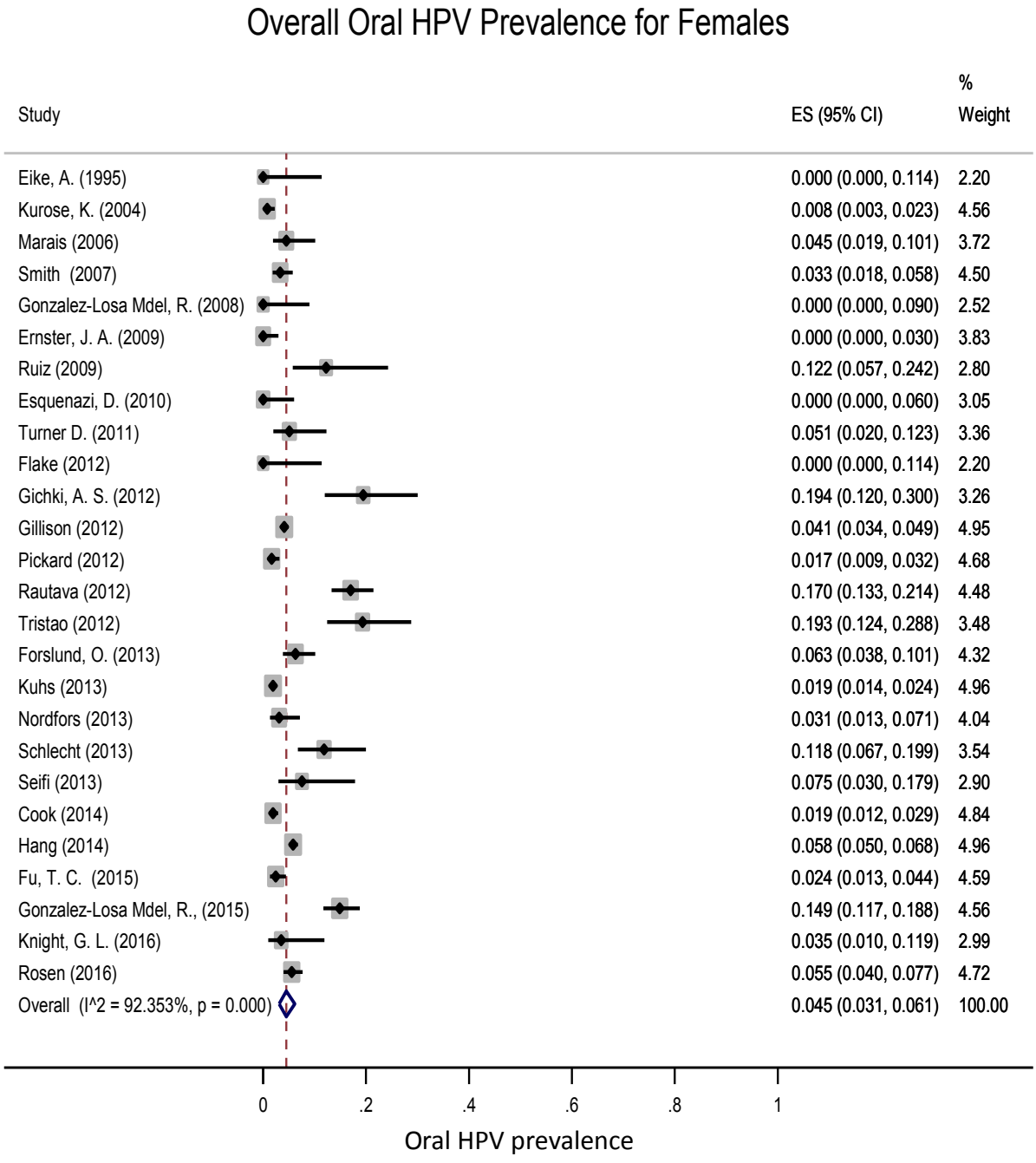


Figure 2.9: Pooled prevalence for any oral HPV infection in healthy females. *The pooled prevalence for any oral HPV infection in females was calculated using STATA and presented as a proportion with 95% confidence intervals depending on the weighting assigned based on sample and effect size. The pooled prevalence for any oral HPV in females was 4.5% (95% C.I: 3.1 – 6.1%).*

Meta-analysis summary:

Our meta-analysis using 45 studies in current literature has provided the following *pooled prevalence for oral HPV in healthy individuals:*

- Any HPV: 4.9% (95% CI: 3.7% - 6.3%)
- High-risk HPV: 1.8% (95% CI: 1.2% - 2.6%)
- HPV 16: 0.7% (95% CI: 0.4% - 1.0%)
- HPV 18: 0.1% (95% CI: 0.0% - 0.3%)

Oral HPV prevalence by gender:

- Males: 4.1% (95% CI: 2.4% - 6.1%)
- Females: 4.5% (95% CI: 3.1% - 6.1%)

There was no statistically significant difference in oral HPV pooled prevalence by gender.

2.21 Sub-populations at higher risk of oral HPV infection:

The meta-analysis presented so far has focused on oral HPV prevalence amongst healthy individuals from 45 published studies that meet the inclusion criteria. It is important to try and identify any sub-sets of populations that may be at increased risk of acquiring oral HPV infection. The systematic review conducted in achieving the final 45 studies included in the prevalence in healthy individuals, also provided concurrent literature on sub-populations potentially at-risk of higher oral HPV prevalence rates. The studies excluded from the healthy cohort were reviewed and stratified into populations with higher oral HPV prevalence. There were 4,365 papers identified on initial search of the relevant databases. After duplicates were removed, there were 2, 231 papers on oral HPV prevalence from which the final 45 studies on healthy populations were extracted. The remaining studies not selected for the 'healthy population' meta-analysis contained the required data to assess high-risk groups.

In 4,628 individuals worldwide with positivity to Human Immunodeficiency Virus (HIV) in this meta-analysis, 31.0% had oral HPV positivity for high-risk HPV and 11.5% prevalence for high-risk HPV (Table 2.4). In comparison to the healthy population, this indicates a relative risk of 11.5 for HIV positive individuals in terms of acquisition of oral high-risk HPV (Table 2.4).

Oral high-risk HPV prevalence was found to be 28.9% in 373 sex workers (relative risk – 10.7), 14% in 920 women with simultaneous HPV positivity in the cervix (relative risk – 5.2). T-test for each group compared with the healthy meta-analysis cohort was significant ($p<0.05$) (Table 2.4).

Table 2.4: Populations at-risk for higher oral HPV prevalence. *This table summarises the pooled prevalence for oral high-risk HPV infection obtained in sub-populations considered to be at higher risk compared to healthy populations. Papers reporting oral high-risk HPV prevalence in women with cervical HPV positivity, sex workers and individuals with HIV were pooled using STATA and the relative risk calculated in relation to the prevalence identified in healthy individuals from this study.*

Population Type	Population Aggregate Total (N)	Oral High-Risk HPV (%)	Relative risk for Oral High-Risk HPV
Healthy	**28,516**	**1.8**	**1.0**
HPV positivity in Cervix	**920**	**14.0**	**5.2**
Sex Workers	**373**	**28.9**	**10.7**
HIV positive	**4,628**	**31.0**	**11.5**

2.22 Oral HPV persistence in healthy individuals (N – 11 studies):

Persistence of oral HPV infection in healthy individuals was investigated in this systematic review. 'Persistence' was defined as the detection of any oral HPV genotype in oral cavity samples in the included studies on healthy individuals using any HPV detection techniques on two consecutive HPV testing time points. Genotype-specific persistence indicates the same oral HPV genotype was detected on two consecutive tests. General persistence is when different genotypes of HPV are identified at index testing and the follow up testing. The limitation with these definitions is that 'true oral HPV persistence' cannot be determined as it is impossible to show that the oral HPV genotype identified on follow up testing is the exact same virus detected initially at the index testing. The concept of oral HPV clearance is also based on the premise that tested individuals are shown to be negative for oral HPV infection on two consecutive tests after initially testing positive for any HPV genotype.

2.22.1 Systematic review papers on oral HPV persistence:

There were 11 studies identified from the 45 included studies on oral HPV prevalence in healthy individuals that had tested for the presence of oral HPV infection at least on two occasions (Table 2.5). Repeat detection of oral HPV genotypes in healthy individuals in this review ranged between 0% - 14.0%. The time lapse on follow up testing after the initial test for oral HPV infection varied between the included studies from 3 months (Pickard et al., 2012) to 7 years (Kero et al., 2014). Some observational studies employed multiple repeat tests over a long study period while some studies only reported two testing time points. The majority of included studies utilised PCR technique for oral HPV detection.

Table 2.5: Studies on oral HPV persistence in healthy individuals. *Oral HPV persistence in healthy individuals was extracted from the prevalence studies in the systematic review. Any studies that carried out repeat testing after a period of time were included. The sample size, duration of follow up and HPV detection technique are described in 11 studies.*

Study ID	1st Author	Year	Country	Total Sample size	Follow up duration	HPV detection	Oral HPV prevalence at index testing - overall	Prevalence overall at index testing (%)	Overall persistence N (%)
1	Kero	2014	Finland	129 fathers to be	7 years	PCR	24/131	18.30%	18/129 (14.0%)
2	Louvanto	2013	Finland	329	6 years	PCR	115/329	34.95% Incidence	36/329 (10.9%)
3	Kurose	2004	Japan	662	2.5 years	PCR	4/662	0.60%	2/662 (0.3%)
4	Garolla	2014	Italy	98 (49 couples)	2 years	PCR INNO-LiPA	5/49 couples (10.20%)	10.20%	0/25 couples (0%) Counsel grp
5	Edelstein	2012	USA	212 adults	1 year	PCR	16/212	7.5%	5/212 (2.4%)
6	Kreimer	2013	USA	1626 men	4 years	Linear Array	115/1626	4.4% Incidence	8/1626 (0.49%)
7	Pickard	2012	USA	1000	3 months	PCR	24/1000	2.4%	9/1000 (0.9%)
8	D,Souza	2007	USA	63 HIV negative women	6 months	PCR	10/63	15.9%	6/63 (9.5%)
9	Cook	2014	USA	7 University women	12 months	Linear Array	19/1010	1.9	0/1010 (0.0%)
10	Herrero	2013	Costa Rica	7466	4 years	PCR SPF/DEIA	127 / 7466	1.7%	15/2924 (0.5%)
11	Fu	2015	USA	409	6 months	PCR Linear Array	10/409	2.4%	3/381 (0.8%)

Data on the oral HPV genotypes identified on repeat testing was also extracted in this systematic review. HPV 11 (which causes warts) had the longest mean detection time in repeat sampling of 55.2 months, followed by HPV 66 (33.4 months), HPV 51 (30.7 months), HPV 12 (28 months), HPV 78 (28 months), and HPV 16 (20.15 months) (Figure 2.10). Significantly, the main oral HPV genotypes associated with head and neck cancer – namely HPV 16 and HPV 18 – had duration of persistence in healthy individuals of 20.15 months and 6 months respectively.

2.22.2 Oral HPV genotypes and persistence:

High-risk HPV genotypes found on repeat oral HPV testing in these included studies (from the lowest to highest persistence in mean months) were HPV 18, 39, 33, 58, 16, 51, and 66 (Figure 2.10). Similarly, low risk oral HPV identified and their relative persistence in mean months from least to longest included: HPV 6, 78, 12 and 11 (Figure 2.10).

Oral infections with multiple HPV genotypes had a mean duration for persistence on repeat testing of 11.65 months. Overall persistence for any oral HPV genotype ranged between 6 months – 55.2 months while oral HPV 16 specifically, had a persistence range of 18.6 – 21.7 months.

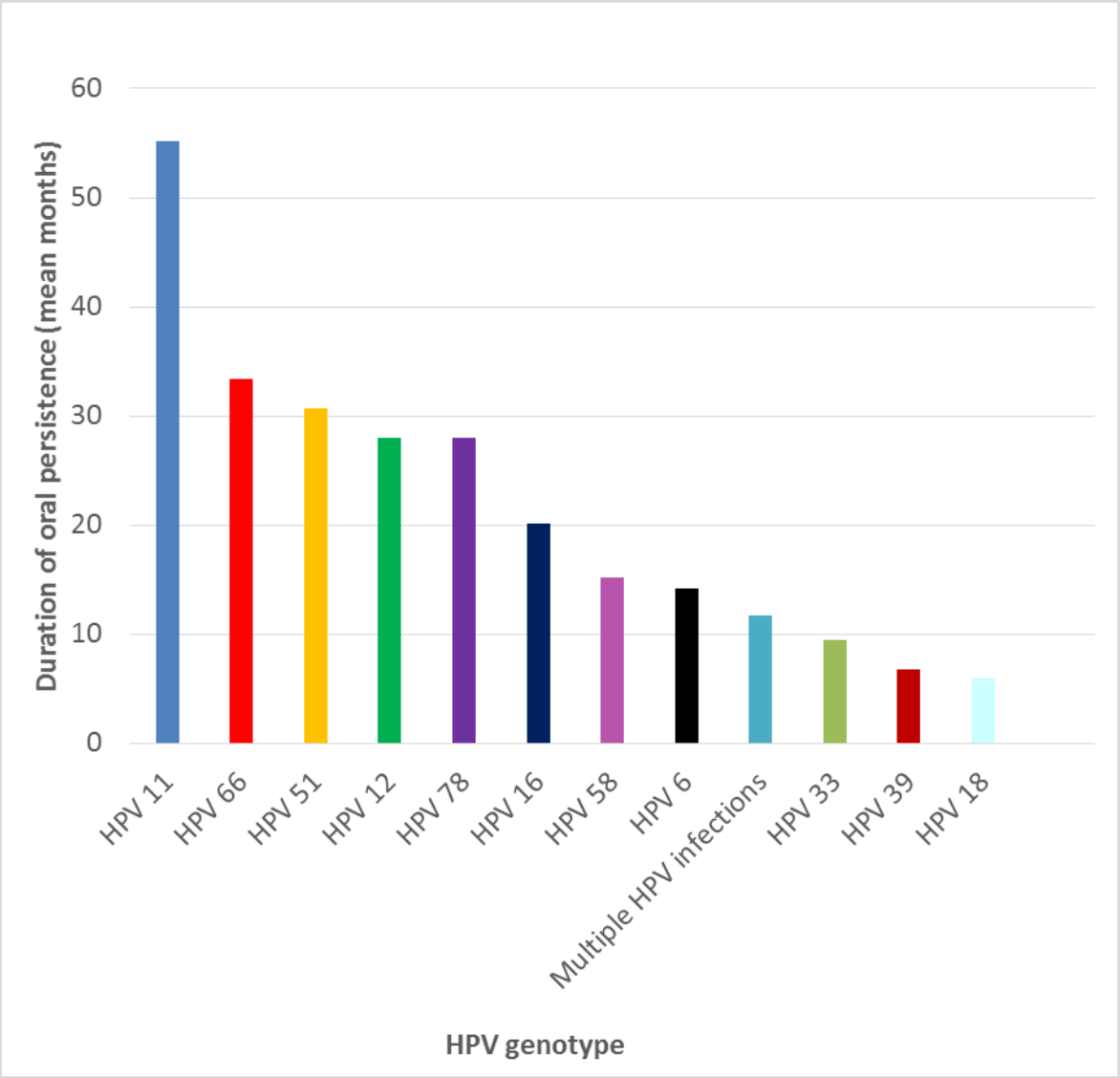


Figure 2.10: Duration of persistence of oral HPV genotypes (mean months). *Persistence of oral HPV infection in included studies showed a range of 6 – 55 mean months. Among low-risk HPV, genotype 11 had the longest duration of persistence while in the high-risk group, HPV 16 persisted for 20 mean months.*

2.23 Discussion:

2.23.1 Oral HPV prevalence in healthy individuals:

Our current understanding of HPV-mediated carcinogenesis comes from robust studies and clinical trials conducted into how persistent infection with high-risk HPV leads to carcinoma in-situ in the cervix (CIN) and ultimately cervical cancer (Cuschieri et al., 2005; Remmink et al., 1995; McCredie et al., 2008). High-risk HPV genotypes 16 and 18 cause the vast majority of cervical cancer worldwide. In addition, high-risk HPV genotypes cause a significant proportion of other genital cancers including vulval cancers, vaginal cancers, rectal cancers, penile cancers, and anal cancers. In recent times, high-risk HPV infection has been confirmed as causal for a subset of Oropharyngeal cancer (cancers of the tonsils, tongue base, soft palate and pharyngeal walls). Causality was established by strong epidemiologic and laboratory studies worldwide, with ratification at the International Agency for Research into Cancer (IARC) Conference in Lyon, France in 2007 (WHO 2007).

There is limited data available on the natural history of oral human papillomavirus infection, its associated risk factors, the population groups at increased risk and any variations in oral HPV prevalence by gender, tissue type, and world geographic location. This meta-analysis addresses these fundamental questions and provides further insight into the acquisition and natural history of oral HPV infection. This is important information as it adds to the on-going discourse on the progression of oral HPV infection to the development of HPV-associated oropharyngeal cancer.

The prevalence of oral HPV infection in healthy individuals was described in the most recent systematic review by Kreimer et al. (2011). In 4,070 participants, pooled prevalence for any oral HPV was 4.5% (95% C.I: 3.9% – 5.1%); in 4441 participants, high-risk oral HPV was

found in 3.5% (95% C.I: 3.0% - 4.1%); while in 3,977 participants, oral HPV 16 was found in 1.3% (95% C.I: 1.0% - 1.7%) (Kreimer et al., 2011). Direct comparison with this systematic review and meta-analysis shows a significantly higher proportion of healthy individuals (28,516) from the 45 studies included. This demonstrates the increasing number of publications in this field in the last 6 years. The pooled prevalence for any oral HPV, oral high-risk HPV and oral HPV 16 reported in this meta-analysis is comparable to the figures published by (Kreimer et al., 2011) and therefore corroborates their findings. This meta-analysis reports pooled prevalence in healthy individuals worldwide for any oral HPV of 4.9% (95% CI: 3.7% - 6.3%); oral high-risk HPV of 1.8% (95% CI: 1.2% - 2.6%); oral HPV 16 of 0.7% (95% CI: 0.4% - 1.0%); and oral HPV 18 of 0.1% (95% CI: 0.0% - 0.3%). This meta-analysis reports a comparatively lower pooled prevalence for oral high-risk HPV of 1.8% (95% CI: 1.2% - 2.6%) with distinct 95% confidence intervals when compared with the pooled prevalence for oral high-risk HPV of 3.5% (95% C.I: 3.0% - 4.1%) reported in the systematic review by Kreimer et al. (2011). The difference in reported rates in oral high-risk HPV between the two systematic reviews is likely due to the inclusion of some studies with participants that have been classified as 'at-risk' for their systematic review. In particular, there were studies with sex workers, and participants recruited from sexually transmitted infections clinics in the Kreimer et al. (2011) review that have been excluded from our meta-analysis. The lower rate of high-risk HPV thus found in this review could reflect less heterogeneity in the inclusion criteria for 'healthy individuals' and as such, a better representation of the true prevalence.

2.23.2 Oral HPV prevalence and gender:

This systematic review did not show a statistically significant difference in pooled prevalence for any oral HPV by gender in healthy individuals. In males, pooled prevalence for any oral

HPV infection was 4.1% (95% CI: 2.4% - 6.1%) while the corresponding prevalence in females was 4.5% (95% CI: 3.1% - 6.1%). This finding is consistent with the (Kreimer et al., 2011) systematic review that also did not show a difference in oral HPV prevalence by gender with reported rates of 4.6% in males and 4.4% in females. Significantly, HPV-associated Oropharyngeal cancer disproportionately affects males more than females (Javadi et al., 2017; D'Souza et al., 2007a; Mehanna et al., 2013). It is therefore of interest that gender does not appear to have an effect on acquisition of oral HPV infection in younger healthy individuals but becomes a major factor with male preponderance when Oropharyngeal cancer occurs in the cohort aged over 50years. It is unclear whether hormonal factors play a protective role in females with regards to HPV carcinogenesis in the Oropharynx, but what is clear is the fact that males (as is the case with head and neck cancers generally) represent the vast majority of HPV-associated Oropharyngeal cancer. The impact of sexual habits and engagement in high-risk sexual behaviours may also be a factor to consider in the observed gender disparity in HPV-OPC. Oral sex has been shown to be an independent risk factor in the development of HPV-associated Oropharyngeal cancer (D'Souza et al., 2009). The concept that the female genital tract acts as a reservoir for HPV infection in sexually active women positive with HPV in the cervix, with the attendant risk of transmission to males during oro-genital contact, has been explored as a possible explanation for the higher oropharyngeal HPV load in males that could facilitate oropharyngeal cancer development in males.

2.23.3 Oral HPV prevalence and risk factors:

Risk factors identified to be associated with an increased prevalence of oral HPV infection in healthy populations include male gender, unmarried women, cigarette smoking, number of lifetime sex partners, ever having had sex, number of lifetime oral sex partners (Gillison et al., 2012; Chung et al., 2014).

Oral sexual behaviours have been shown to elevate both the odds for oral HPV infection as well as the odds for developing HPV-associated Oropharyngeal cancer (D'Souza et al., 2009). There exists the possibility that sexual behaviours in Western populations have changed in recent times, with more people engaging in high-risk sexual activity. The current literature in the UK that supports this narrative comes from the three *National Survey of Sexual Attitudes and Lifestyles* 'NATSAL' studies. Johnson et al. (2001) reported on the changing patterns of sexual behaviour in Britain in the NATSAL 2000 survey as compared to NATSAL 1990. The survey concluded there was an increase in risky sexual behaviours over the decade mirroring the observed changes in cohabitation trends and the underlying increase in sexually transmitted diseases. Mercer et al. (2013) went further in NATSAL 3 to confirm that there have been significant changes in sexual behaviours and attitudes in the United Kingdom in the last 60years. In particular, women had experienced more fluctuations in sexual behaviours and in general there had been a distinct rise in oral sex and anal sex over time. The increase in the uptake of these high-risk behaviours has coincided with the rise of HPV-associated Oropharyngeal cancer. There is some evidence to indicate deep kissing as a transmission route for oral HPV infection. Kissing is well documented to be a transmission route for other viruses such as Epstein Barr Virus (EBV) which causes glandular fever or *'infectious mononucleosis'*. EBV also affects the tonsils (as does HPV) and significantly, is causative for nasopharyngeal cancer. EBV has been found to co-exist in some forms of nasopharyngeal carcinoma with HPV (WHO Type III classification). The viral milieu in the oral cavity and oropharynx is therefore transmissible by sexual activity that involves contact with a sexual partner's secretions – be it orogenital contact or oro-oral contact. The number of lifetime sexual partners with whom patients with HPV-OPC engaged in sexual intercourse, kissing or oral sex, was shown to be a significant risk factor in Oropharyngeal cancer development when

compared with other head and neck cancer patients (Schnelle et al., 2017). Patients with HPV-OPC had significantly elevated odds of having participated in oral sex with four or more partners when compared with patients with other head and neck cancers that were not HPV positive (Schnelle et al., 2017).

This systematic review and meta-analysis on oral HPV infection in healthy people did not extract data on sexual behaviours in participants included in the 45 studies as this data was often absent or not uniformly collected. However oral sex has been shown to increase the risk of acquiring HPV infection in the oral cavity even in healthy populations (D'Souza et al., 2009).

2.23.4 Oral HPV prevalence and geography:

This meta-analysis sought to identify any variations in the pooled prevalence of oral HPV infection based on the geographic location of the background population in each of the included 45 studies. The heterogeneity in the populations worldwide as well as variations in the HPV detection techniques utilised by each study means the meta-analysis results should be interpreted with caution. The pooled oral HPV prevalence for any HPV genotype across all world regions was between the range 1.3% (95% C.I: 0.9 – 1.8%) in Eastern Europe and 13.8% (95% C.I: 10.5 – 17.4%) in Asia. The highest prevalence was reported in Asia however this figure was driven by a single study in Pakistan that utilised a highly sensitive HPV detection technique called real-time PCR (Gichki et al., 2012). This may therefore be an over-representation of the true prevalence of oral HPV infection in that country and comparisons with other world regions that used mainly PCR-based methods is sub-optimal. True comparisons in oral HPV prevalence can only be made when HPV detection techniques are

universally standardised and type-specific techniques compared against each other rather than using any HPV detection technique from all studies.

2.23.5 Oral HPV prevalence and sample type:

This meta-analysis investigated any differences in HPV prevalence depending on the oral samples collected in the studies. Using all techniques, HPV detection was highest at 5.9% in oral scrapings / brushings / swabs. Studies that collected oral rinse / gargle / saliva had a lower HPV detection rate of 3.5%. Surprisingly, studies that identified HPV in tonsils reported the lowest rates at 1.0% and FFPE from tonsils at 0.0%. A large UK study corroborates these findings with 0% oral HPV in 3,377 FFPE specimens and 0% oral HPV in 511 non-malignant tonsils (Palmer et al., 2014). Given that HPV has been demonstrated to localise in tonsillar crypts (Begum 2005), a higher prevalence inside tonsil specimens was expected. This fact is corroborated by a recent study investigating oral HPV infection in healthy individuals undergoing tonsillectomy for non-cancer reasons in the UK (OROMOUTH). HPV detection from 937 tonsil specimens using p16 IHC and PCR, yielded only 1.4% prevalence (Mehanna 2017 – unpublished work). The HPV lifecycle could explain the increase in detection in oral cavity specimens and fluids in comparison with the actual tonsil reservoir itself. There is more shedding of viral particles at the epithelial layer when virus assembly is complete with millions of genomic HPV released, in contrast to the significantly smaller viral load present in the basal layer or reticulated epithelium of the tonsillar crypts. This has implications in terms of the development of screening techniques and approaches for the oral cavity and oropharynx. It will be useful for screening programmes to incorporate into their repertoire for oral HPV detection, tissue sampling that will likely provide the greatest oral HPV yield.

2.23.6 Oral HPV prevalence and HPV detection technique:

There are wide variations in HPV detection techniques with a lack of consensus worldwide on which technique or combination of techniques constitute true HPV positivity. In the UK, consensus PCR and positivity on p16 Immunohistochemistry has become the benchmark for defining HPV positivity in recent clinical trials. In clinical practice however, the majority of hospitals and treating centres utilise positivity only on p16 IHC to stratify patient groups into HPV positive or negative. This approach is supported in part by current evidence that shows HPV positivity on p16 IHC is a reliable surrogate marker of HPV positivity for clinical applications. In-situ hybridization (ISH) and real-time PCR are also sensitive HPV detection technique used in some of the included studies. There is however the risk of over-estimating oral HPV prevalence using these techniques. 43 of the 45 studies in this review used PCR alone or in combination with other detection techniques such as ISH and p16 IHC. Other less frequently used HPV detection techniques include hybrid capture which was used in 2 of the included studies in this systematic review. Hybrid capture has been used in the identification of HPV in cervical specimens however, the main limitation with this technique is its inability to identify the exact HPV genotype present in the sample. It simply provides a binary outcome: HPV present or absent. This is therefore less useful in stratifying results into those associated with high-risk or low-risk HPV or even type-specific HPV.

2.23.7 Oral HPV infection and sub-populations at-risk:

An investigation into which groups in the population are at increased risk of acquiring oral HPV infection (with a higher prevalence than the healthy population), is necessary to better understand its aetiopathogenesis and for facilitating resource allocation when considering the feasibility of screening. This meta-analysis has identified sub-populations with significantly

higher pooled prevalence of oral HPV infection when compared to the pooled prevalence in healthy individuals of 4.9% (95% CI: 3.7% - 6.3%) for any HPV, 1.8% (95% CI: 1.2% - 2.6%) for high-risk HPV, and 0.7% (95% CI: 0.4% - 1.0%) for oral HPV 16. The sub-populations at increased risk of oral HPV infection thus include a) individuals with positivity to Human Immunodeficiency Virus (HIV), b) sex workers, c) women with positivity to HPV in the cervix, d) patients with Fanconi's anaemia.

The highest pooled prevalence for oral HPV was identified in individuals positive for HIV. In 4,628 HIV positive individuals, high-risk HPV was found in 31%. This represents a relative risk of 11.5 when compared with healthy populations. The immune system plays a role in ensuring the necessary balance between exposure and infection with the ubiquitous HPV, and ultimately persistent oral HPV infection and progression to disease states. Strong epidemiologic evidence has shown that oral HPV infection is cleared in immune-competent individuals between 18 – 24 months from viral acquisition (Kero et al., 2014, D'Souza et al., 2014a). This may not be the case in individuals with compromise of their immune system such as occurs with HIV infection, renal transplant patients, patients on chemotherapy, or other immune-deficiency states. There are on-going studies investigating the role of the immune system in the natural history of oral HPV infection and progression to oropharyngeal cancer.

Expectedly, among 373 individuals who identified as sex workers in some of the included studies, 28.9% were found to have oral high-risk HPV. This was a relative risk of 10.7 when compared to healthy populations with high-risk HPV of 1.8% as indicated previously. Exposure to multiple sexual partners is clearly the risk factor in this sub-population.

In 920 women positive for HPV in the cervix and simultaneously tested for oral HPV, 14% were found to have high-risk HPV in oral samples – a relative risk of 5.2.

2.23.8 Oral HPV persistence:

This systematic review identified papers that investigated the natural history of oral HPV infection in terms of persistent infection. Persistence was defined as the detection of oral HPV on two consecutive tests done at two separate time points. There was wide variability in the reported rates of oral HPV persistence with a range between 0.0% - 14.0% in the 11 included studies. There were 2 prospective longitudinal studies from the FINNISH family study that took serial oral samples from 325 pregnant women and their partners over a 7year period that have provided the clearest observation of how oral HPV infection progresses in healthy populations. In immune-competent individuals, oral HPV infection was cleared within 18 – 24 months on repeat testing (D'Souza et al., 2014a). Risk factors identified to be associated with persistent oral HPV infection include current smoking and use of oral contraceptives (Kero et al., 2014). There are a few caveats to bear in mind when interpreting results reporting oral HPV 'persistence'. True persistence is difficult to ascertain because studies usually do not test whether the exact HPV detected in the first oral sample is the same virus detected again in the second sample after a time interval (it is however possible to know by genotyping). It is obvious that the oral cavity hosts a dynamic and changing viral community and therefore, HPV infection can be acquired, cleared and re-acquired over a period of time. It follows therefore, that the HPV infection detected on the second test could as well be a new HPV infection rather than the same HPV infection identified ab initio. It is pertinent however to note that persistent HPV infection in the cervix is a recognised precursor to the initiation of cervical intraepithelial neoplasia and ultimately cervical carcinoma. This model of carcinogenesis is proven by robust data over the years and therefore the hypothesis that

persistent oral HPV infection leads to Oropharyngeal cancer, is credible and worthy of scientific consideration. Gillison et al. (2012) described a bi-modal peak for oral HPV infection with the first peak in teenagers and adults in their twenties and the second peak observed in adults aged over 50 years. Our understanding of how oral HPV infection translates into oropharyngeal cancer through the lens of this described bi-modal peak, indicates either that these two peaks are related and as such persistent oral HPV infection leads to OPC development, or that these two peaks are unrelated and represent two separate 'new oral HPV infections'. If the latter hypothesis is true, then the first peak represents the age of initiation of sexual activity while the second peak may be indicative of changing sexual patterns that may occur in middle age such as divorce and sexual activity with new partners.

2.23.9 Oral HPV infection and vaccination:

HPV vaccination is available either as a bi-valent vaccine (Cervarix), a quadrivalent vaccine (Gardasil), and even a vaccine covering 9 HPV genotypes. Pre-pubertal girls aged 12 – 13 years in the UK, USA, Canada, Australia and some countries in Europe receive HPV vaccination on national programmes. The HPV vaccine was primarily aimed at prevention of HPV associated diseases ranging from benign conditions like anogenital warts (HPV 6, 11) to cervical and anogenital cancers (HPV 16, 18 and other high-risk HPV). There are increasing calls for gender-neutral vaccination that protects boys as well as girls. There is an expectation that vaccination will offer a protective effect against anogenital cancers and benign HPV conditions in boys too. Countries such as the United States of America, Canada, Australia and some others across Europe, have in recent times begun vaccinating boys as well as girls against HPV-related diseases. In summer 2018, the UK followed in approving vaccination for both boys and girls. The impact of HPV vaccination on oral infection and HPV-associated

oropharyngeal cancer prevalence is yet to be fully determined however Herrero et al. (2013) showed a reduction in oral HPV prevalence from 1.7% to 0% 4 years after HPV vaccination.

2.23.10 Oral HPV infection and screening:

HPV screening currently exists for cervical intraepithelial neoplasia and cervical cancer based on the observed changes on cervical cytology from scrapings from the transitional epithelium at the squamo-columnar junction of the cervix. Recently, cervical cancer screening programmes have begun the transition from a cytology-based process to a HPV-based screening focused on identifying high-risk HPV. Screening in the cervix for HPV-driven disease and cancer is possible in part due to the presence of a pre-malignant state (CIN) identifiable in the transitional zone of epithelium and changes that can be observed from pre-malignancy to cervical cancer. In contrast, these conditions are not replicated for HPV-Oropharyngeal cancer. In the first instance, there is a lack of a known pre-malignant state in HPV-OPC. Dysplasia is rarely identified in the oropharynx in healthy individuals while HPV positive dysplasia is even rarer. HPV has been shown to localise in the reticulated epithelium of the tonsillar crypts and therefore it is challenging to develop screening tools that can effectively access these sites like a ‘pap-smear’ equivalent of the cervical cytology screening. Further research is required to establish which oral sampling techniques will be most productive in the development of any screening programme for oropharyngeal HPV infection or cancer. It is also important to thoroughly investigate which oral samples – oral rinse, oral scrapings, swabs, biopsies or tonsil tissue – will be adopted into any screening programme on oral HPV infection. There will also need to be universal consensus on the HPV detection techniques and what parameters constitute oral HPV positivity e.g. p16 positivity and the presence of HPV DNA on PCR.

PhD CHAPTER 3

THE IMPACT OF TONSILLECTOMY ON THE NATURAL HISTORY OF ORAL HUMAN PAPILLOMA VIRUS INFECTION IN HEALTHY INDIVIDUALS:

DATA FROM THE UK OROMOUTH COHORT

3.1 Introduction:

High-risk human papillomaviruses (HR-HPV), are causal for a subset of oropharyngeal cancer (OPC) (Gillison et al., 2000; Chaturvedi et al., 2013; D'Souza et al., 2007a). In particular, HPV 16 is associated with over 90% of HPV oropharyngeal cancer. There is currently limited data on the natural history of oral HPV infection itself. Available studies indicate oral HPV acquisition coincides with the onset of sexual activity and in immune-competent individuals, is cleared within 18 – 24 months (Chung et al., 2014; Kreimer et al., 2013; Kero et al., 2012a; Edelstein et al., 2012; D'Souza et al., 2007b). Persistence of high-risk HPV demonstrably increases the risk of progression to high grade squamous intraepithelial cervical neoplasia (Cuschieri et al., 2005; Vega-Pena et al., 2013; Remmink et al., 1995; Schettino et al., 2014). In contrast, persistence of oral high-risk HPV infection has yet to be shown in any longitudinal studies to be a direct prelude to OPC development. The obvious limitation of such a study design (amongst others), is the length of follow up required due to a latency of over 20 years between oral infection and overt oropharyngeal cancer. It is known that p16 expression localises to the specialised epithelium of tonsillar crypts (Begum et al., 2005) but it remains undetermined if tonsillectomy affects the natural history of oral HPV infection. This is the first study to investigate the natural history of oral HPV infection after tonsillectomy.

3.2 Patients and methods:

The OROMOUTH study was a multicentre cross-sectional study on oral HPV prevalence in healthy individuals in the UK undergoing tonsillectomy for non-cancer reasons http://public.ukcrn.org.uk/Search/StudyDetail.aspx?StudyID=1234. The OROMOUTH study (N-937), has now completed recruitment. The study was conducted in multiple centres in the

UK with recruitment of patients undergoing tonsillectomy. Patients signed consent forms and had oral rinse and urine samples taken before their surgery as well as swabs of the posterior pharyngeal wall and tongue base in theatre. All samples including the tonsils were collected and sent for HPV DNA analysis at DDL Belgium. Patients aged 16years and above on OROMOUTH also completed sexual behaviour questionnaire with the data anonymised, returned in a sealed envelope and analysed by the research team in a blinded fashion. The results of the larger OROMOUTH study is currently being analysed and will be published in a peer-reviewed journal soon.

The oral HPV persistence study presented in this thesis was a longitudinal study (N – 126) on a subset of the OROMOUTH cohort in which the student collected oral rinse samples from patients following tonsillectomy. The patients were identified from a random sample of patients who had tonsillectomy on the OROMOUTH study and were willing to provide second oral rinse samples post-operatively. They were approached at the time of surgery and consented to second sample collection. They were reminded by telephone by the student when their appointments were due following surgery. The oral rinse samples were collected by the student in clinic at the hospital.

3.2.1 Setting: Hospital-based, multi-centre study in the West Midlands United Kingdom.

3.2.2 Participants: All participants on the OROMOUTH study who were positive for the presence of HPV DNA (all ages) and a control group of patients negative for HPV DNA (stratified into participants >14 years old and children ≤ 14 years old) with pre- and post-tonsillectomy oral rinse samples available for analysis.

3.2.3 Samples: Oral rinse (gargles with normal saline for 30seconds) was taken on the day of tonsillectomy. Second samples were taken 3 – 12 months following tonsillectomy and both

pre- and post-operative oral rinse samples tested for HPV DNA by PCR. 2nd visit samples were collected in the Ear, Nose and Throat clinic by the student and pipetted into preservative-containing bottles, stored in – 80^0C ready for laboratory analysis by a diagnostic laboratory in Belgium.

3.2.4 Laboratory tests: These tests were not performed by the student but undertaken by scientists at the diagnostic laboratory in Belgium. HPV DNA amplification was performed using a broad-spectrum short-PCR-fragment assay [SPF_{10}] PCR-DNA Enzyme Immunoassay (DEIA) followed by a primer SPF10-based line probe assay (SPF_{10}LiPA) and MPTS 123 (a novel E6-based multiplex type-specific system that uses the Luminex xMAP technology). For this analysis, DEIA positivity was regarded as HPV positivity and the HPV genotype was any strain identified from either the LiPA or MPTS 123 assays.

3.2.5 Statistical analysis: All demographic and laboratory test results were anonymised, sequentially entered on to a central database and extracted on to Excel spreadsheets by the student. The required variables were analysed by the student and correlated with oral HPV positivity using a statistical package (STATA SE).

3.2.6 Ethical approval: Approval was secured for this study with OROMOUTH MREC No.: 11/WM/0283, UKCRN ID: 12344.

The role of the student included arranging second visits for a subset of the OROMOUTH cohort following tonsillectomy, collection of oral rinse samples post-tonsillectomy, statistical analysis of the HPV data on samples processed by DDL Belgium and presentation of the results.

3.3 Results:

All participants included in this analysis (children and adults) have all had pre- and post-tonsillectomy oral rinse samples taken for HPV DNA identification. **126** participants (in total) were included in this analysis. **25** participants were aged < *14 years old*, while **101** participants were aged ≥ *14 years old*. Both groups are analysed separately, as oral HPV acquisition and transmission routes are likely to be different between children and adults.

3.4 Tonsillectomy cohort of participants aged 14 years or more:

Demographic variables:

Overall, the age range of participants aged ≥ 14 years old was 14.8 – 55.2 years, with median age 22.1 years (IQR: 18.8 – 30.5years) and mean age 25.3years. There were 19 males (18.8%) and 82 females (81.2%) in this cohort. The age range for males was 15 – 23 years; median age 28.4years (IQR: 20.8 – 31.5 years) and mean age 28.2 years. The age range for females was 14.8 – 55.2 years; median 22 years (IQR: 18.7 – 29.8 years) and mean age 24.7 years. These demographic data are shown in Table 3.1. The difference in mean age by gender was not statistically significant on Fisher's T-test (p=0.111).

Table 3.1: Tonsillectomy population aged 14 years and over. *This table presents the sample size of participants aged 14 years and above in a tonsillectomy population, their age range, median and mean stratified by gender.*

	Overall	Males	Females	Statistical testing
Sample size (N) (%)	101	19 (18.8%)	82 (81.2%)	
Age range (years)	14.8 – 55.2	15.0 – 23.0	14.8 – 55.2	
Median age (IQR)	22.1 (18.8 – 30.5)	28.4 (20.8 – 31.5)	22.0 (18.7 – 29.8)	
Mean age (years)	25.3	28.2	24.7	Fisher's T-test T=1.61; p=0.111

3.5 Paired oral rinse samples:

In this study, there were 101 participants aged ≥ 14 years with paired oral rinse samples (i.e. before and after tonsillectomy) available for analysis. In the pre-operative oral rinse samples, there were 11/101 (10.89%) positive for oral HPV on testing (Table 3.2). Of the 11 oral HPV positive individuals, there was 1 male and 19 females. The gender distribution of oral HPV positivity pre-operatively was therefore 1/19 males (5.26%) and 10/82 females (12.20%).

Following tonsillectomy, oral rinse samples were collected from all 101 participants (aged ≥ 14 years) and tested using the same HPV testing protocol as performed pre-operatively. Of note, only 2 participants remained HPV positive following tonsillectomy among the 11 participants who were positive pre-operatively. There were however 9 participants with new oral HPV infections detected in oral rinse samples following tonsillectomy from the 90 patients previously HPV negative pre-operatively. This indicates a 9/90 (10%) proportion of the cohort developed new oral HPV infections after tonsillectomy.

Oral HPV positivity for any oral HPV genotype remained the same at 11/101 (10.89%) pre- and post-tonsillectomy. However, following tonsillectomy, 82% of those with existing infections cleared their infection. Four of the six cases with high-risk HPV oral infection cleared their infections after tonsillectomy. The only person with HPV 16 oral infection cleared their infection. There was no patient with oral HPV 18 pre-operatively. Tonsillectomy was associated with a reduction in the burden of any oral HPV infection by 80.1% and high-risk oral HPV by 63.5% but these reductions were not statistically significant given there were 9 new infections during the study period.

Table 3.2: Pre-operative oral HPV infection status in the tonsillectomy cohort aged 14years and over. *This table shows the demographic distribution of the tonsillectomy population pre-operatively stratified by oral HPV status*

	Overall	Oral HPV positive	Oral HPV negative	Statistical testing
Sample size (N) (%)	101	11 (10.9%)	90 (89.1%)	
Age range (years)	14.8 – 55.2	14.9 – 43.8	14.8 – 55.2	
Median age (IQR)	22.1 (18.8 – 30.5)	21.9 (18.6 – 23.8)	22.15 (18.9 – 30.9)	
Mean age (years)	25.3	23.6	25.5	Fisher's T-test T=-0.699; p=0.486

3.6 Oral HPV genotypes:

The oral HPV genotypes identified before tonsillectomy was performed in the 11 participants using LiPA and MPTS, are as follows: oral HPV 11, 16, 39, 51, 52, 59, 66 and 68. There were 5 oral rinse samples in which the oral HPV present was not typeable. Most oral HPV infections were multiple infections with more than one HPV genotypes as there were 8/11 multiple infections noted pre-operatively. Oral HPV infection in these healthy participants were either low risk HPV infection 5/101 (4.95%) or high-risk HPV infection 6/101 (5.94%).

3.7 Oral HPV positivity stratified by oropharyngeal sites:

HPV was detected in oral rinse in 11/101 (10.89%) of participants before tonsillectomy. The tongue base brushes yielded only one positive sample for which the HPV genotype was not typeable. In both the posterior pharyngeal wall brushes and the tonsils specimens, there was no oral HPV identified.

In terms of correlation with other oral samples, only one patient who was found to be positive for oral HPV in oral rinse pre-operatively who was also HPV positive in any other oral sample (tongue base brush) at the same time. This was with a concordant 'not typeable' genotype in both oral rinse and tongue base brush.

3.8 All pre-operative oral samples and HPV detection:

The different oral samples yielded varying degrees of oral HPV positivity. HPV positivity was found in oral rinse in 11/101 (10.89%), in tonsils in 4/101 (3.96%), in tongue base brushes in 3/101 (2.97%) and in pharyngeal wall brushes 0/101 (0%). Oral rinse therefore provided the highest yield of oral HPV pre-operatively. The HPV genotypes found in the 3 positive tongue base brushes were HPV 53, and 2 samples with 'non-typeable HPV' (regarded as low-risk HPV). Only 1 tongue base positive sample also showed positivity in the corresponding oral rinse sample. In the 4 positive tonsils, the oral HPV genotypes identified included HPV 16 (in 2 tonsils), HPV 66 and 'non-typeable HPV'. Remarkably, none of the 4 individuals with HPV detected in their tonsils had simultaneous detection of HPV in their oral rinse. There were no patients with oral HPV positivity on pharyngeal wall brushes so sensitivity on oral rinse testing was not possible. The true sensitivity and specificity of the use of oral rinse and the other sampling techniques (tongue base brush, pharyngeal wall brush, and tonsil samples) will be fully assessed in the main OROMOUTH study (N – 937).

3.9 Impact of tonsillectomy on the natural history of oral HPV infection:

Before tonsillectomy, there were 11 subjects positive for oral HPV and 90 subjects negative for oral HPV. After tonsillectomy, only 2/11 HPV positive patients remained positive on oral rinse testing (Table). Importantly, 9/90 (10%) subjects who were HPV negative before tonsillectomy, were subsequently found to have HPV positivity on oral rinse after their tonsillectomy. These were regarded as new oral HPV infections.

Although tonsillectomy reduced the prevalence of oral HPV infection in the infected people with a reduction from 11/101 (10.89%) pre-operatively to 2/92 (2.17%) post-operatively, this was not statistically significant on McNemar's test due to the 9 new infections observed in the cohort. McNemar chi-squared statistic with Yates correction of 1.0 was 0.056. The corresponding p-value (2 tails) was 0.814. The odds ratio equals 1.0 with 95% Confidence interval from 0.397 to 2.519 (Table 3.3). This indicates a non-significant impact of tonsillectomy on the natural history of oral HPV infection in this tonsillectomy population.

Table 3.3: Impact of tonsillectomy on oral HPV infection in cohort aged 14 years and over. *Table showing number of participants positive for oral HPV infection before and after tonsillectomy in individuals aged 14years and over. Although there was a reduction in oral HPV positivity post-operatively, this was not statistically significant on McNemar's testing due to occurrence of new infections in previously oral HPV negative participants.*

	HPV Positive After Tonsillectomy	HPV Negative After Tonsillectomy	Total
HPV Positive Before Tonsillectomy	2	9	11
HPV Negative Before Tonsillectomy	9	81	90
Total	11	90	101

3.10 Children's cohort:

Only 25 children < 14 years old, had paired oral rinse pre- and post-tonsillectomy available for inclusion in this analysis. Understandably, there were difficulties obtaining oral rinse or urine from children waiting for surgery or even after surgery. The age range of this cohort of children was 3.4 – 13.3 years with a median age of 7.6 years (IQR 5.6 – 10.2 years) (Table 3.4).

Table 3.4: Demographic data for the tonsillectomy cohort of children. *Table showing sample size, age range, median and mean age for children having tonsillectomy aged <14 years, stratified by their oral HPV status.*

	Overall	Oral HPV positive	Oral HPV negative	Statistical testing
Sample size (N) (%)	25	5 (20%)	20 (80%)	
Age range (years)	3.4 – 13.3	3.9 – 7.9	3.4 – 13.3	
Median age (IQR)	7.6 (5.6 – 10.2)	7.4 (6.2 – 7.6)	8.2 (5.6 – 11.0)	
Mean age (years)	8.1	6.6	8.5	T-test T: -1.303 P-value: 0.205

3.11 Oral HPV positivity in children:

5/25 (20%) children had oral HPV positivity in oral rinse pre-operatively (Table 3.4). The majority of these infections were due to low risk HPV as 'non-typeable HPV' was found in 4 children and HPV 35 in one child. Low-risk HPV therefore accounted for 4/25 (16%) in children, while high-risk HPV was found in 1/25 (4%) pre-operatively. There were 3/13 males (23.1%) and 2/12 females (16.7%) positive for oral HPV among this cohort of children. The mean age of children overall was 8.1 years with the breakdown along oral HPV positive (6.6 years) and oral HPV negative (8.5 years) groups. There was no difference in mean age between the two groups on student T-test with a p-value=0.205 (Table 3.4).

3.12 Impact of tonsillectomy on oral HPV infection in children:

Of the 5/25 (20%) children with oral HPV positivity before tonsillectomy, 2/25 (8.0%) children had persistence of oral HPV infection post-tonsillectomy (Table 3.5 and Figure 3.1).

Both children with oral HPV on repeat testing following tonsillectomy had low-risk HPV (both had non-typeable HPV). The only child with high-risk HPV pre-operatively (HPV 35) had cleared this on oral rinse testing post-operatively. There was however a single child with a new infection post-tonsillectomy despite being negative pre-operatively. In assessing the impact of tonsillectomy on oral HPV infection, McNemar's test statistic was 0.25, with a 2-tailed p-value = 0.617 and an odds ratio = 3. This indicates the observed impact of tonsillectomy in reducing oral HPV prevalence, was not statistically significant.

Table 3.5: Impact of tonsillectomy on oral HPV infection in children. *Table showing number of children less than 14 years old with oral HPV infection before and after tonsillectomy. There was a reduction in oral HPV positivity post-operatively but this finding was not statistically significant given the small numbers on McNemar's test.*

	HPV Positive After Tonsillectomy	**HPV Negative After Tonsillectomy**	**Total**
HPV Positive Before Tonsillectomy	2	3	5
HPV Negative Before Tonsillectomy	1	19	20
Total	3	22	25

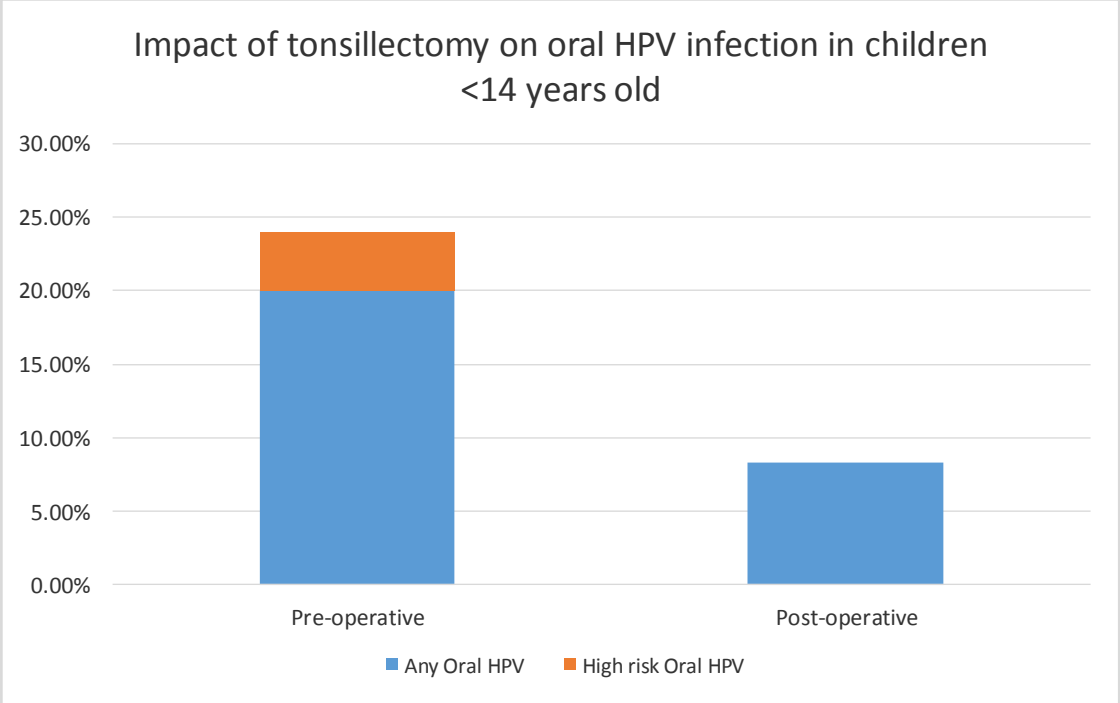


Figure 3.1: Impact of tonsillectomy on oral HPV infection in children < 14 years old. *Chart shows a comparison between oral HPV infection in children before and after tonsillectomy. Any oral HPV is depicted in blue while orange is high-risk oral HPV.*

3.13 Age distribution of oral HPV infection:

The entire dataset obtained from this OROMOUTH cohort was further classified into age-groups as follows (in years): [≤ 14], [14.0 – 19.9], [20.0 – 24.9], [25.0 – 29.9], [30.0 – 34.9], [35.0 – 39.9], [40.0 – 44.9], [45.0 – 49.9], [≥ 50]. The total number of participants within each age-group and the relative numbers of individuals with oral HPV positivity in each group is shown in Table 3.6 below. The age-group with the highest proportion of individuals in this cohort was the 14.0 – 19.9 years group with 34 patients, followed by the 20.0 – 24.9 years age-group with 30 individuals (Table 3.6). With 25 children aged <14 years, it clearly shows that 89/101 participants in this cohort were young and aged <25years. This is a consistent expectation for a tonsillectomy population. The age groups with the highest numbers of

participants with oral HPV infection in this healthy cohort were the <14 years age group with 5/25 (20%) HPV positive, the 20.0 – 24.9 years age-group with 6/30 patients (20%) positive for oral HPV and the the 40.0 – 44.9 years age-group with 1/6 patients (16.7%) positive for any oral HPV infection (Table 3.6 and Figure 3.2).

Table 3.6: Age distribution for any oral HPV infection. *Table showing the age distribution of the entire tonsillectomy cohort with the majority of participants being young and <25years.*

Age group (years)	Total number of individuals	Number Oral HPV positive	% Oral HPV positive
< 14	25	5	20.0
≥14 – 19.9	34	3	8.82
20 – 24.9	30	6	20.0
25 – 29.9	9	0	0.0
30 – 34.9	18	1	5.6
35 – 39.9	1	0	0.0
40 – 44.9	6	1	16.7
45 – 49.9	0	0	0.0
≥ 50	3	0	0.0

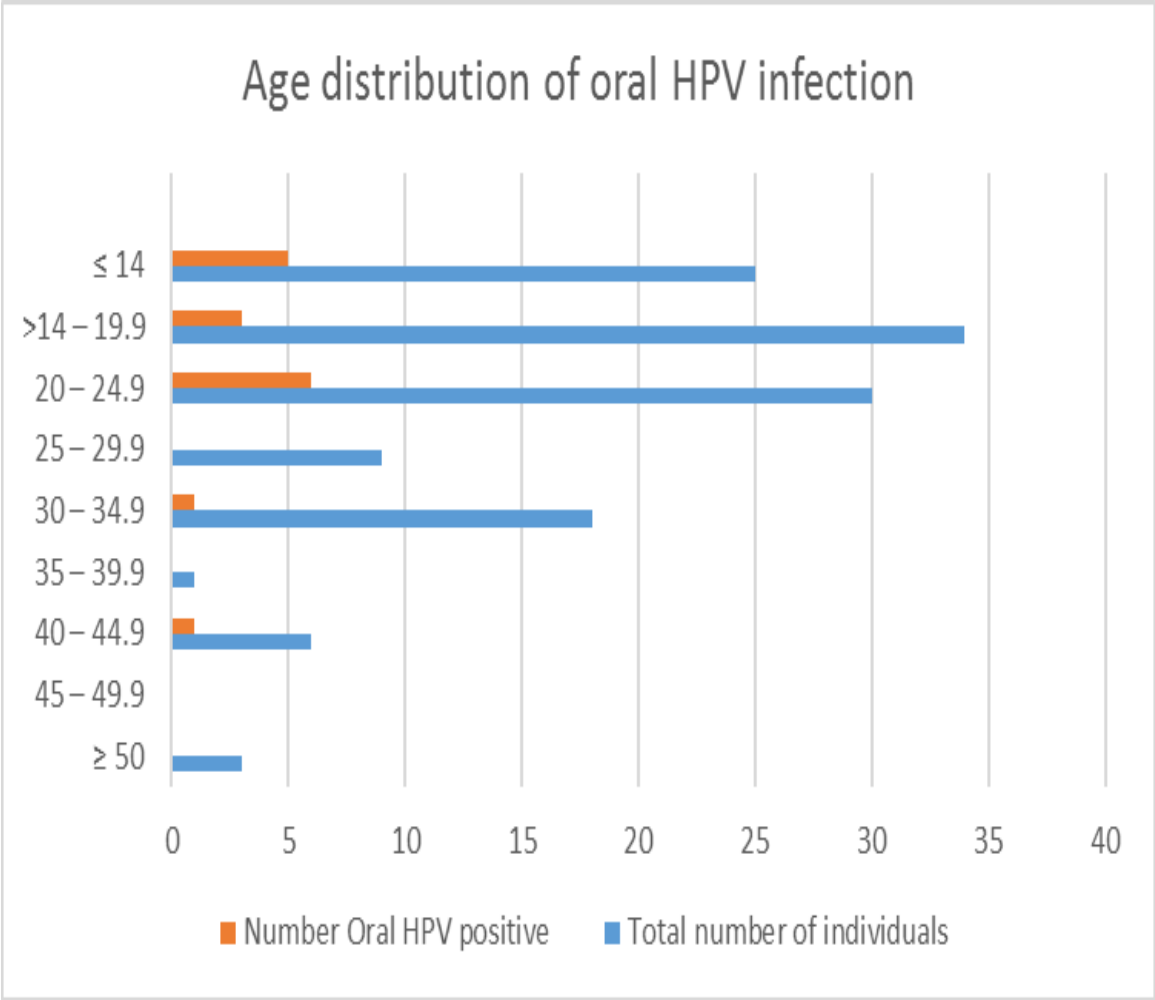


Figure 3.2: Age distribution of oral HPV infection. *Figure showing the age distribution of the tonsillectomy population in this study. The total number of individuals in each age group is depicted in blue columns while the number of individuals with any oral HPV infection is represented in orange.*

3.14 Discussion:

Persistent infection with high-risk HPV has been recognised as a precursor of invasive cervical cancer (Cuschieri et al., 2005; Vega-Pena et al., 2013; Remmink et al., 1995; Schettino et al., 2014). A similar association is yet to be characterised for persistent oral HPV infection and oropharyngeal cancer development. Oral HPV infection is dynamic and varies with the onset of sexual activity, changes in sexual behaviours and exposure to risk factors (Cook et al., 2014). Studies on the natural history of oral HPV infection show that immune-competent individuals are able to clear the infection within 18 – 24 months (Chung et al., 2014; Kreimer et al., 2013; Kero et al., 2012a; Edelstein et al., 2012; D'Souza et al., 2007a). Oral sex is an independent risk factor for acquisition of oral HPV infection when confounders are controlled for. In men, tobacco smoking is associated with persistent oral HPV infection (Kero et al., 2014), as is oral contraception in women. There are higher proportions of oral HPV infection in individuals with HIV infection (Beachler et al., 2012), a history of previous HPV in the cervix, and sex workers (Matsushita et al., 2011), with variations in the factors that will affect persistent infection. No previous studies have found an effect of tonsillectomy on the natural history of oral HPV infection. We sought to demonstrate for the first time, the potential effect of tonsillectomy on oral HPV infection in healthy subjects.

Our study has shown a trend towards an abrogation of oral HPV infection following tonsillectomy in previously positive individuals. However, there was no statistically significant association found in adults or in children because several cases of new infections occurred post-tonsillectomy.

The acquisition of oral HPV infection occurs mainly with the onset of sexual activity in adults. Oral sex is demonstrably a risk factor but the evidence for deep kissing as a

transmission route is less convincing. Vertical transmission has been described in children (Vega-Pena et al., 2013) as well as correlation with maternal cervical HPV status (Chatterjee et al., 1998; Pinheiro Rdos et al., 2011; Koskimaa et al., 2012; Kravchenko, 2011). This study suggests there is a difference in the oral HPV genotype distribution between children and adults as expected, given the different drivers of infection in both groups. Children had more low-risk HPV strains compared with sexually active adults with more high-risk HPV. This is consistent with literature that shows oral HPV peaks in teenagers once they become sexually active (Flake et al., 2012).

Oral HPV persistence: In the children's cohort, only one child out of eleven pre-operatively positive patients (on any sample) remained positive on oral rinse after tonsillectomy. This was a seven year old girl with non-typable HPV (which is low-risk) and no significant risk factors. Amongst adults, two out of fifteen participants had persistence of oral HPV infection following tonsillectomy in any sample, with oral rinse positivity fractions reducing from 7 pre-op to only 1 adult post-op. This was an 18-year old female, regular smoker, with multiple high-risk oral HPV infection (strains 11, 39, 66). She also provided a history of drinking alcohol twice a week, engaging in oral sex and has had 9 lifetime male sexual partners. Smoking does increase the risk of oral HPV persistence (Kero et al., 2014) but it is unclear what other risk factors could explain persistence in this case despite tonsillectomy.

3.15 Study conclusions:

This is the first UK study to provide crucial data on the potential impact of tonsillectomy on the natural course of oral HPV infection. Tonsillectomy appears to be associated with a reduction in oral HPV positivity. However, we do not have a control arm and so we cannot tell whether this was natural clearance, or an effect of tonsillectomy or a combination.

PhD CHAPTER 4

EPIDEMIOLOGIC DETERMINANTS OF OVERALL SURVIVAL IN OROPHARYNGEAL CANCER:

A RETROSPECTIVE POPULATION ANALYSES OF THE UK HOSPITAL EPISODES STATISTICS (HES) DATABASE

4.1 Introduction:

Oropharyngeal cancer (OPC) are a sub-group of head and neck cancers that primarily include cancers of the tonsils, tongue base, soft palate, uvula and pharyngeal walls. Incidence of OPC is mainly in males aged >50 years with recognised risk factors of tobacco smoking and alcohol consumption. High risk strains of human papillomavirus (HPV) are now an established aetiology for a unique subset of oropharyngeal cancer (HPV OPC). HPV OPC is more common among younger males who do not smoke tobacco or consume excessive amounts of alcohol (Gillison et al., 2000). Oral sex and multiple sexual partners over a lifetime are risk factors for HPV OPC (D'Souza et al., 2007b). Determinants of survival include age (elderly patients have worse survival), gender (males worse), socio-economic status (worse in deprived populations), and significant co-morbidities (worse survival with increasing co-morbidity scores).

4.2 Aims: To describe the baseline demographics and the epidemiologic determinants of overall survival in patients with oropharyngeal cancer recorded on the UK Hospital Episodes Statistics (HES) database.

4.3 Methods: Retrospective analyses of the UK HES database for cases of oropharyngeal cancers. The database covers all NHS Trusts in England with admitted patient care information available only since 1989, attendance at Outpatient clinics since 2003 and attendance in Accident and Emergency department since 2007.

4.4 Statistical enquiry: We described demographic data on oropharyngeal cancers recorded on the UK HES database including the age distribution and described the median, interquartile range and range which were then compared with the independent T-test.

The anatomic sub-sites for oropharyngeal cancer included tonsil, base of tongue, and other oropharyngeal sites (namely soft palate, uvula, posterior and lateral pharyngeal walls). These data were all coded using the ICD-10 classification including ICD 10 C09.0, C09.1, C09.8, C09.9 for tonsil cancers and C02.9, C10.9. The HES database was accessed via the licence granted to the Bioinformatics department of the Queen Elizabeth Hospital Birmingham. The data was anonymised and the student only had access to the files on the secure site.

We investigated the gender distribution of oropharyngeal cancers recorded on the HES database as well as the regional proportions, and ethnicity, assessing for any significant differences with the X^2 test. We documented behavioural risk factors such as tobacco smoking and alcohol consumption among patients with OPC on HES.

We assessed for the impact of co-morbidity, deprivation, oropharyngeal cancer anatomic site and behavioural risk factors on overall survival from OPC in the HES database using Kaplan Meier and Cox regression analyses.

We calculated yearly incidence of oropharyngeal cancers in the UK overall and by region. The numbers of oropharyngeal cancers documented on the HES database was utilised per region as the numerator while the population per year in every region was obtained from the UK Office of National Statistics (ONS). There were 21,135 males and 7,711 females with oropharyngeal cancers in the HES database over the study period.

Human papillomavirus (HPV) status on the UK HES database was also reviewed. Only a small proportion of oropharyngeal cancers, specifically 161/28,846 (0.56%), were recorded to be HPV positive. It is therefore difficult to draw any conclusions from this variable in particular because HPV status only recently became part of the documentation records from cancer networks across the UK for oropharyngeal cancer.

4.5 Results:

Using the UK HES (Hospital Episodes Statistics) Database, all data recorded for oropharyngeal cancer in the United Kingdom between the years 2001 and 2016 were analysed. There were 28,846 patients recorded on the HES database with a diagnosis of oropharyngeal cancer that also had demographic data available for analysis. In this oropharyngeal cancer cohort, there were predominantly more males accounting for 21,135 (73.27%) of cases, while females were 7,711 (26.73%). This indicates a male:female ratio of 2.7: 1. (Figure 4.1)

4.5.1 Age distribution:

The mean age for all oropharyngeal cancer in this HES cohort was 60 years. The data was further stratified into age group categories as follows: [‘<40 years’, ‘40 – 49 years’, ‘50 – 59 years’, ‘60 – 69 years’, ‘70 – 79 years’ and ‘≥ 80 years’]. The age group with the highest proportion of oropharyngeal cancer diagnoses in the UK from the HES database was the ‘50 – 59 years’ cohort with 9,642 cases (33.43%). This age group was followed in descending order of cancer frequency by the ’60 – 69 years’ with 8,640 cases (29.95%); the ‘70 – 79 years’ with 4,285 cases (14.85%); the ‘40 – 49 years’ with 4,079 cases (14.14%); the ‘≥ 80 years’ with 1,520 cases (5.27%); and finally the ‘<40 years’ group with only 680 cases (2.36%) [Table 4.1]. Oropharyngeal cancer therefore remains a condition that predominantly affects people aged over 50 years old in the UK with a peak prevalence in people aged 50 – 59 years old.

Table 4.1: Age distribution of Oropharyngeal cancer in the UK by gender. *The total numbers of oropharyngeal cancers in each age group are shown and the proportions by gender.*

Age groups (Years)	Total OPC N (%)	Males N (%)	Females N (%)
<40	680	417 (61.3%)	263 (38.7%)
40 - 49	4079	3080 (75.5%)	999 (24.5%)
50 - 59	9642	7319 (75.9%)	2323 (24.1%)
60 - 69	8640	6445 (74.6%)	2195 (25.4%)
70 - 79	4285	2979 (69.5%)	1306 (30.5%)
≥80	1520	895 (58.9%)	625 (41.1%)
Total	28,846	21,135 (73.3%)	7,711 (26.7%)

4.5.2 Gender distribution:

Males account for a higher proportion of Oropharyngeal cancer in ALL age groups recorded on HES. Males accounted for 417/680 cases (61.3%) in patients aged <40years; 3,080/4079 cases (75.5%) in the 40 – 49 years age group; 7319/9642 cases (75.9%) in the 50 – 59 years age group; 6445/8640 cases (74.6%) in the 60 – 69 years age group; 2979/4285 cases (69.5%) in the 70 – 79 years age group; and 895/1520 cases (58.9%) in the ≥80 years age group. The age group with the highest proportion of OPC was therefore the 50 – 59 years group with 75.9% amongst males (Table 4.1).

The results for both gender, follow a normal distribution with the highest gender fractions for Oropharyngeal cancer seen in the 50 – 59 years age group for both males and females (34.63% and 30.13% respectively) and the lowest proportions in the extremes of age.

In males, there were a total of 21,135 oropharyngeal cancer in HES UK. Of this total, there were 417 (1.97%) in males aged <40years; 3080 (14.57%) in the 40 – 49 years group; 7319 (34.63%) in the 50 – 59 years group; 6445 (30.49%) in the 60 – 69 years group; 2979 (14.10%) in the 70 – 79 years group; and 895 (4.23%) in the patients aged ≥80 years old (Table 4.2 and Figure 4.2).

Similarly in females, there were a total of 7711 oropharyngeal cancer in HES UK. Of this total, there were 263 (3.41%) in females aged <40years; 999 (12.96%) in the 40 – 49 years group; 2323 (30.13%) in the 50 – 59 years group; 2195 (28.47%) in the 60 – 69 years group; 1306 (16.94%) in the 70 – 79 years group; and 625 (8.11%) in the female patients aged ≥80 years old (Table 4.2 and Figure 4.2).

Table 4.2: Age-specific distribution of Oropharyngeal cancer in the UK. *Total number of oropharyngeal cancers in each age group by gender.*

Age groups (Years)	Total OPC N (%)	Males N (%)	Females N (%)
<40	680 (2.4%)	417 (2.0%)	263 (3.4%)
40 - 49	4079 (14.1%)	3080 (14.6%)	999 (13.0%)
50 - 59	9642 (33.4%)	7319 (34.6%)	2323 (30.1%)
60 - 69	8640 (30.0%)	6445 (30.5%)	2195 (28.5%)
70 - 79	4285 (14.9%)	2979 (14.1%)	1306 (16.9%)
≥80	1520 (5.3%)	895 (4.2%)	625 (8.1%)
Total	28,846	21,135	7,711

The extremes of the age group categories had the lowest proportions of Oropharyngeal cancer recorded on HES UK. Young patients (aged <40 years) had OPC fractions of 1.97% in males, 3.41% in females while the elderly (≥80 years) had 4.23% in males and 8.11% in females. There was however, no statistically significant difference between the two gender proportions for oropharyngeal cancer recorded on the UK HES database.

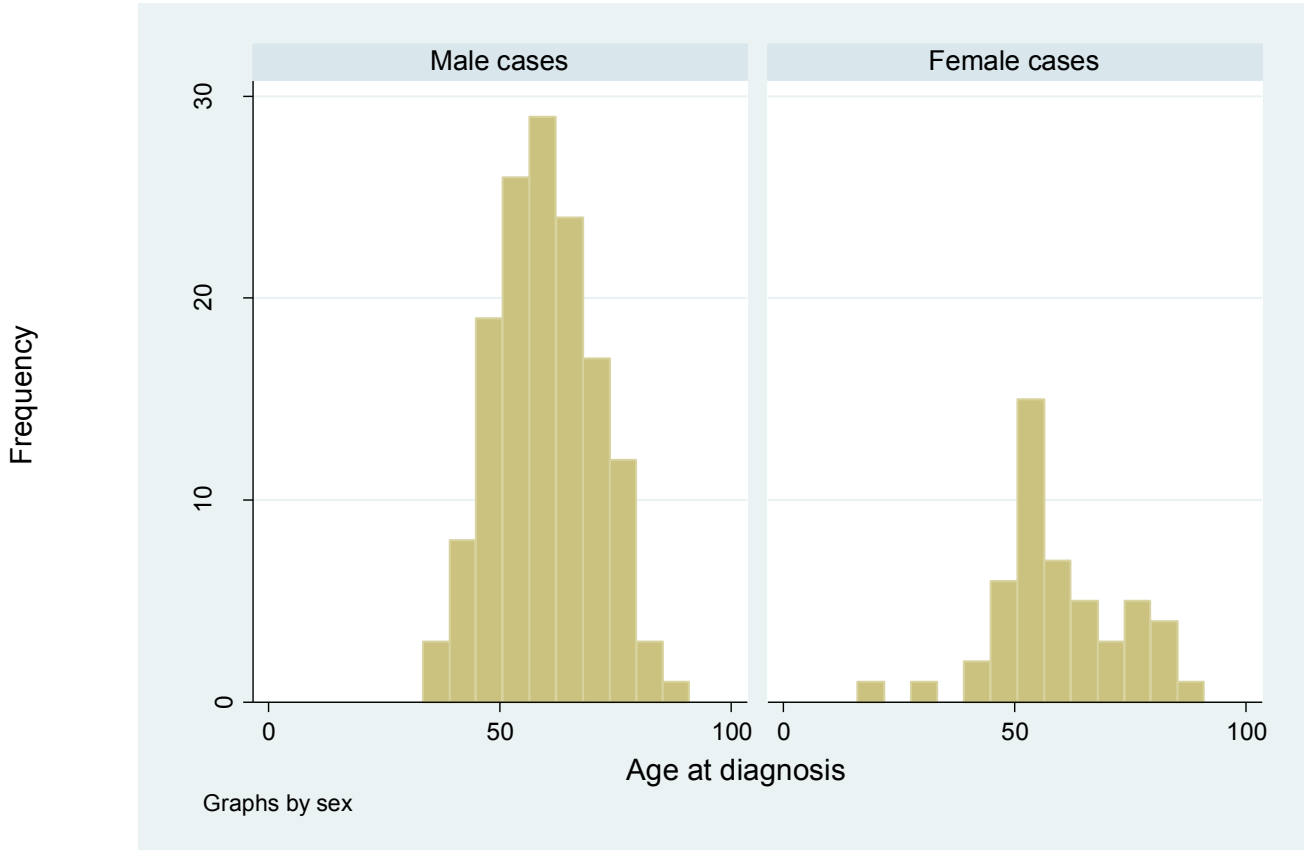
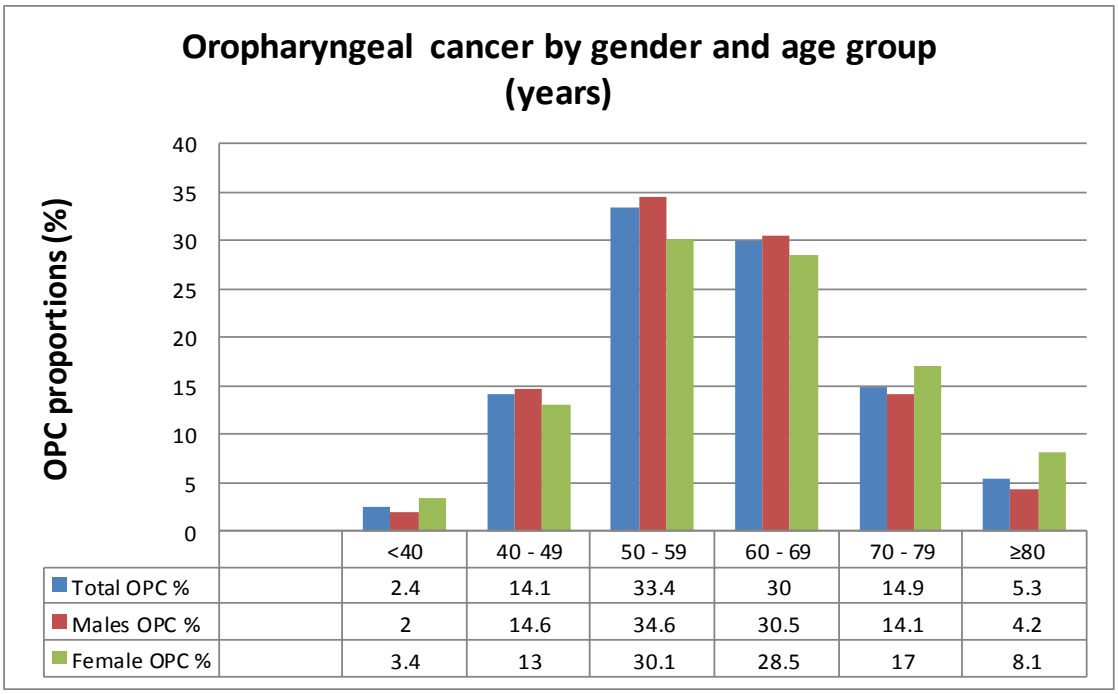


Figure 4.1: Oropharyngeal cancer by gender and age distribution on HES UK.
Histograms showing distribution of oropharyngeal cancer cases in the UK at the age of diagnosis by gender. The majority of OPC cases were recorded in males aged over 50 years.



	<40	40 - 49	50 - 59	60 - 69	70 - 79	≥80
Total OPC %	2.4	14.1	33.4	30	14.9	5.3
Males OPC %	2	14.6	34.6	30.5	14.1	4.2
Female OPC %	3.4	13	30.1	28.5	17	8.1

Figure 4.2: Oropharyngeal cancer distribution in years by sex and age. *Figure showing proportions of oropharyngeal cancers in each age group with total proportions in blue, male proportions in red and female proportions in green. The highest proportions were found in the age group 50 – 59 years.*

4.5.3 Ethnicity and oropharyngeal cancer:

81.8% of this oropharyngeal cohort had an ethnicity recorded as ‘White’ which is expected, given the underlying population is a UK mainly Caucasian population (Table 4.3). Males had higher frequency of oropharyngeal cancer diagnosis on HES UK for all the ethnic groups except Bangladeshi where it was even. These disparities between ethnicity by gender were not statistically significant on Chi-squared (X^2) testing. A baseline analysis using available UK population data on each of these ethnic groups is required to estimate disease prevalence in each group and this was not available.

Table 4.3: Impact of ethnicity on oropharyngeal cancer proportions in UK HES. *Table shows proportions of OPC with ethnicity data available on UK HES. The majority of cases were in the 'White' ethnic group consistent with the background population. There were small numbers in the other ethnic groups available for comparison.*

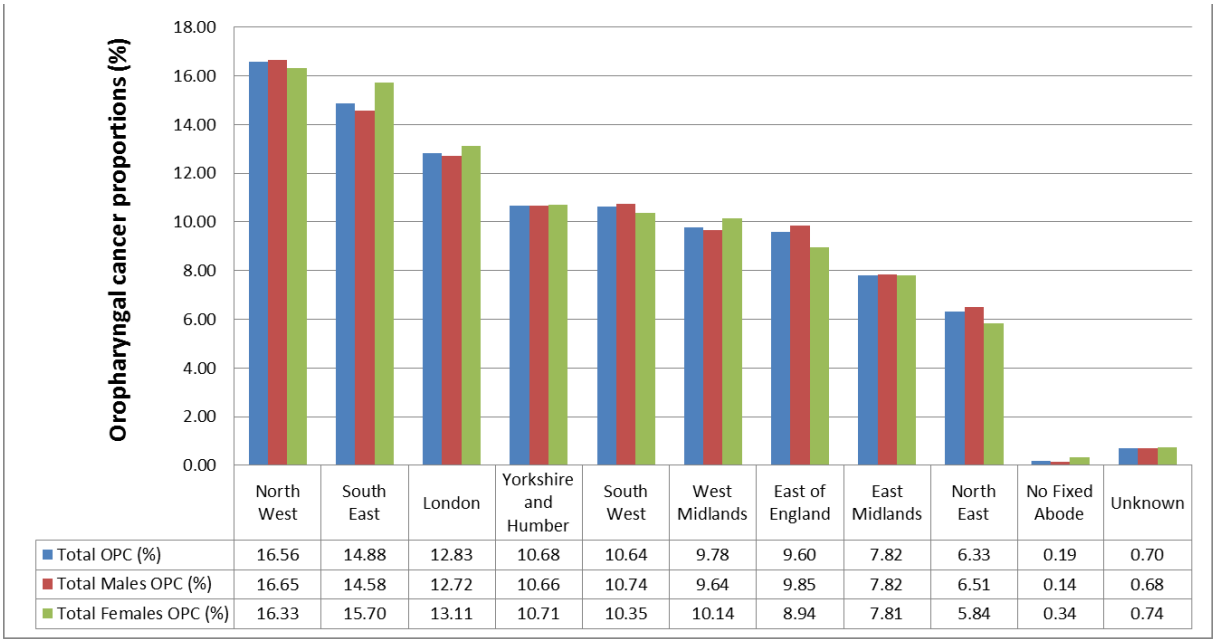
Ethnic Group	Total OPC N (%)	Males N (%)	Females N (%)
White	23,597 (81.80%)	17,332 (82.01%)	6265 (81.25%)
Unknown	4349 (15.08%)	3196 (15.12%)	1153 (14.95%)
Black or Black British	240 (0.83%)	160 (0.76%)	80 (1.04%)
Other Ethnic Group	179 (0.62%)	128 (0.61%)	51 (0.66%)
Indian	170 (0.59%)	118 (0.56%)	52 (0.67%)
Mixed	91 (0.32%)	60 (0.28%)	31 (0.40%)
Pakistani	73 (0.25%)	51 (0.24%)	22 (0.29%)
Bangladeshi	67 (0.23%)	33 (0.16%)	34 (0.44%)
Other Asian Background	60 (0.21%)	46 (0.22%)	14 (0.18%)
Chinese	20 (0.07%)	11 (0.05%)	9 (0.12%)
Total	28,846	21,135	7,711

4.5.4 UK regional variations in oropharyngeal cancer epidemiology:

An analysis of the regional variations in the incidence of oropharyngeal cancers in the UK based on the HES database was performed using the total number of oropharyngeal cancers in each UK region recorded on HES (Table 4.4) and population data obtained from the Office of National Statistics (ONS). The regional proportions of oropharyngeal cancers in the UK are shown in Figure 4.3, with the highest proportions recorded in the North West region.

Table 4.4: Regional variations in Oropharyngeal cancer proportions on UK HES. *OPC cases recorded on HES by UK region and by gender.*

UK Region	Total OPC		Males		Females	
	Total OPC N	Total OPC (%)	Males OPC N	Males OPC (%)	Females OPC - N	Females OPC (%)
North West	4778	16.56	3519	16.65	1259	16.33
South East	4293	14.88	3082	14.58	1211	15.70
London	3700	12.83	2689	12.72	1011	13.11
Yorkshire and Humber	3080	10.68	2254	10.66	826	10.71
South West	3068	10.64	2270	10.74	798	10.35
West Midlands	2820	9.78	2038	9.64	782	10.14
East of England	2770	9.60	2081	9.85	689	8.94
East Midlands	2255	7.82	1653	7.82	602	7.81
North East	1825	6.33	1375	6.51	450	5.84
No Fixed Abode	56	0.19	30	0.14	26	0.34
Unknown	201	0.70	144	0.68	57	0.74
Total	28,846	100.00	21,135	100.00	7,711	100.00



	North West	South East	London	Yorkshire and Humber	South West	West Midlands	East of England	East Midlands	North East	No Fixed Abode	Unknown
Total OPC (%)	16.56	14.88	12.83	10.68	10.64	9.78	9.60	7.82	6.33	0.19	0.70
Total Males OPC (%)	16.65	14.58	12.72	10.66	10.74	9.64	9.85	7.82	6.51	0.14	0.68
Total Females OPC (%)	16.33	15.70	13.11	10.71	10.35	10.14	8.94	7.81	5.84	0.34	0.74

Figure 4.3: OPC proportions in the UK HES by region. *Figure shows proportions of OPC in each UK region with total % in blue, male proportions in red and female proportions in green.*

4.5.5 Overall incidence of OPC in the UK:

The incidence of oropharyngeal cancer in the UK was calculated from the recorded yearly cases of oropharyngeal cancer retrospectively retrieved from the HES UK database between 2001 – 2014 as well as the UK yearly population estimates (overall and by sex) obtained from the Office of National Statistics (2016). The results are presented as per 100,000 population for each year, by sex and by UK region.

There has been an increasing incidence of oropharyngeal cancer in the UK over the study period. OPC incidence has risen steadily from 2.7 cases per 100,000 in 2001 to a peak of 5.8 per 100,000 in 2014 (Figure 4.4). This represents a doubling of the incidence rate over the study period which is consistent with published UK literature.

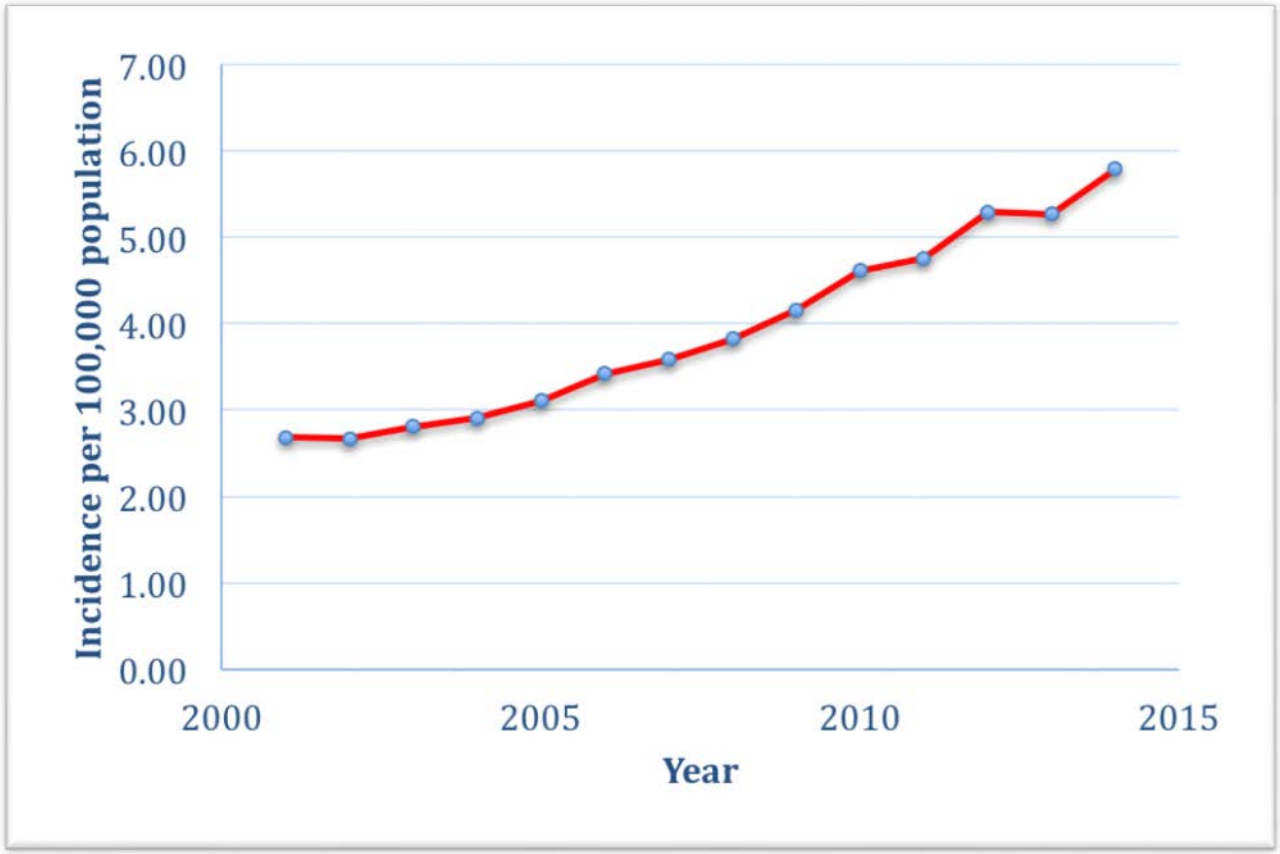


Figure 4.4: Oropharyngeal cancer increasing incidence trend in England. *Year on year incidence of OPC in England shows there has been an increasing trend over the study period*

4.5.6 OPC incidence by gender:

The incidence of oropharyngeal cancer in the UK was further stratified by gender. The data shows that although there has been an increase in OPC incidence in both males and females, the overriding driver in the observed increasing trend is shown in the male incidence. There has been a remarkable increase in OPC incidence among males in the UK with a doubling of the rate in a steady rise from 4.0 per 100,000 population in 2001 to 8.3 per 100,000 in 2014. (Figure 4.5).

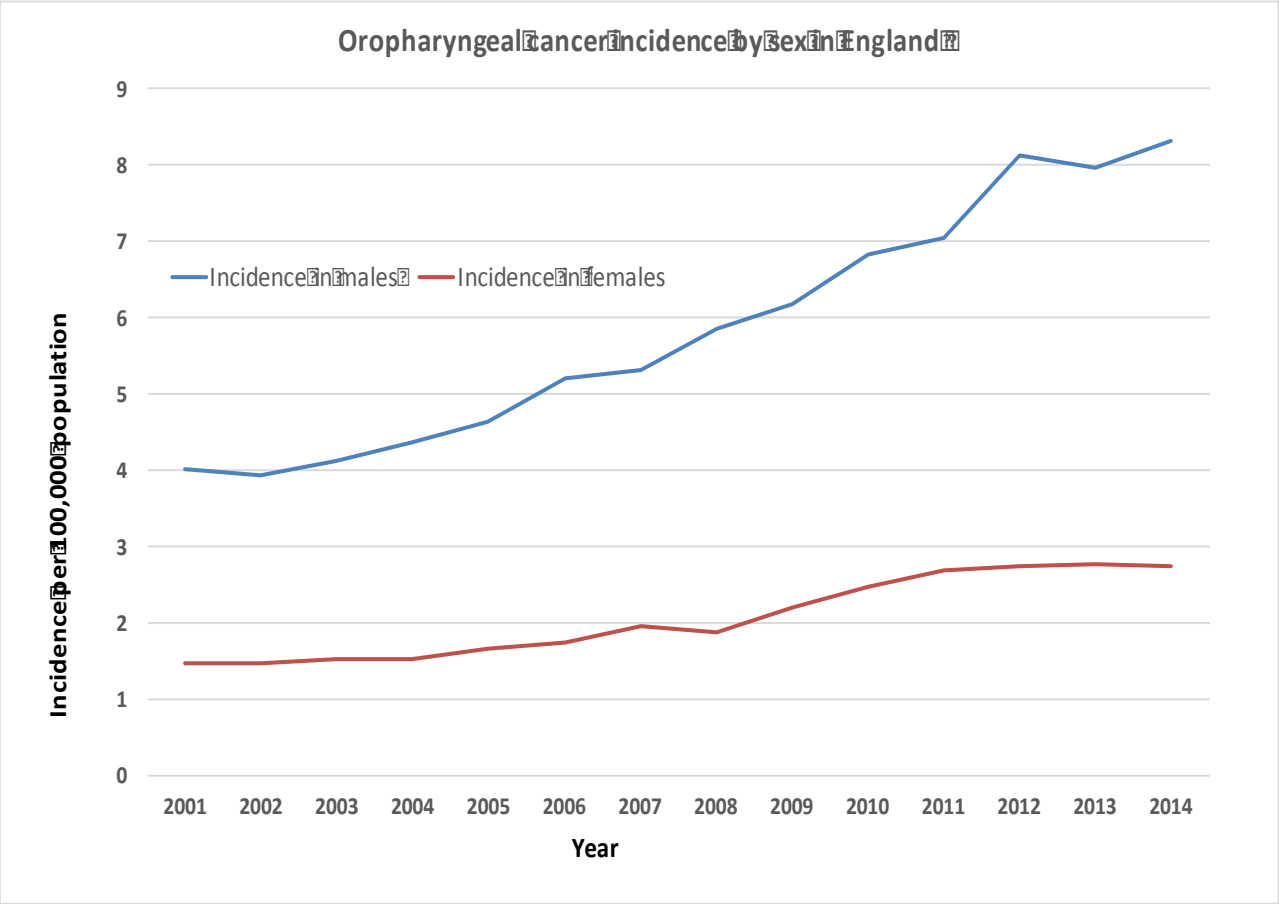


Figure 4.5: Incidence of oropharyngeal cancer in England by gender. *OPC incidence has increased in both males and females over the study period but the greatest increase is observed in male incidence (per 100,000).*

4.5.7 OPC incidence by UK region:

OPC incidence by UK region was also analysed in this cohort. The regional incidence rates in all UK regions studied showed an increasing incidence trend regardless of location (Figure 4.6). However, when compared with other UK regions, the North East and North West regions had the highest incidence rates of OPC at any given time in the study period. By 2014, the highest incidence was reported in the North East at 7.3 per 100,000 population. (Figure 4.6). The underlying factors and population dynamics responsible for the regional

variation needs further investigation which is not possible with the data available from the HES database or the ONS.

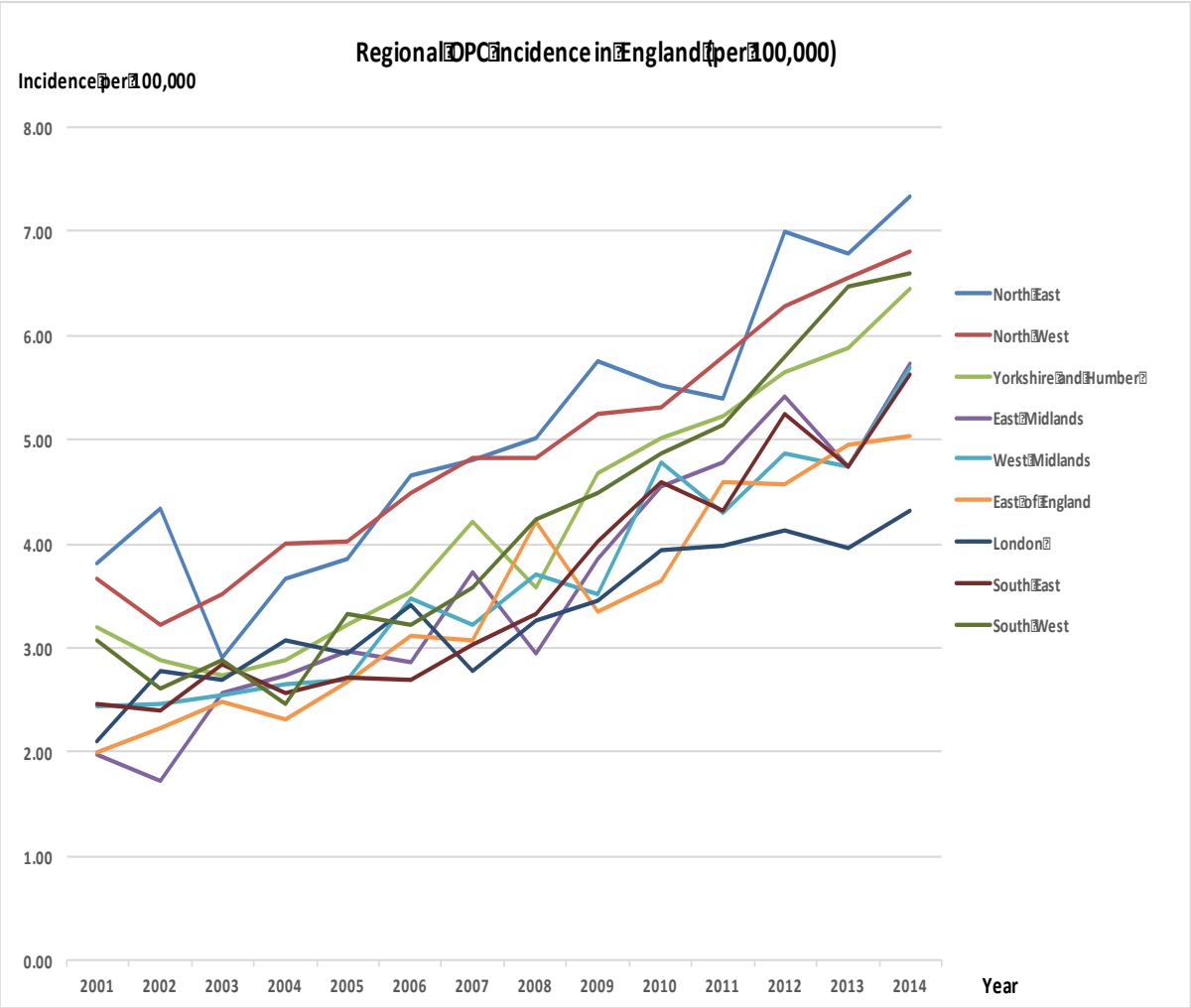


Figure 4.6: Oropharyngeal cancer incidence in England by region. *There has been a steady increasing incidence in OPC across all UK regions with the highest incidence in the North East and North West.*

4.5.8 Anatomic sub-sites of oropharyngeal cancer:

The anatomic sub-sites for oropharyngeal cancer include tonsil, base of tongue, and other oropharyngeal sites (namely soft palate, uvula, posterior and lateral pharyngeal walls). In the UK HES database overall, there were predominantly tonsil cancers with 11,924 cases (41.34%), followed by base of tongue cancers with 9922 cases (34.40%) and other oropharyngeal sites at 7,000 cases (24.27%) as shown in Table 4.5. This trend was maintained when the data was stratified by gender with tonsil cancer being the most frequent anatomic site, followed by base of tongue and lastly other oropharyngeal sites in both males and females (Table 4.5).

4.5.9 Risk factors:

An analysis of known risk factors for the development of oropharyngeal cancer included variables such as smoking history, alcohol consumption, HPV status, Co-morbidity scores and Deprivation scores. Almost 2/3rd of this dataset on oropharyngeal cancer did not have a history of ever smoking recorded on HES. There were 18,408/28,846 OPC cases (63.81%) who did not smoke tobacco and 10,438 cases (36.19%) who were smokers. Amongst males, 7705/21,135 cases were smokers (36.46%) while amongst females, there were 2733/7,711 (35.44%) smokers (Table 4.5). Similarly, a history of alcohol consumption was recorded on the UK HES in 5,840/28,846 OPC cases (20.25%) overall and in 4,613/21,135 males (21.83%) and 1,227/7,711 females (15.91%) (Table 4.5).

Table 4.5: Anatomic sites and risk factors for Oropharyngeal cancer. *Table showing the total numbers and proportions by gender for cancer subsite, smoking status, alcohol consumption, co-morbidity scores, and deprivation scores for all OPC on the HES database.*

	Total OPC N	Total OPC (%)	Total Males OPC N	Total Males OPC (%)	Total Females OPC N	Total Females OPC (%)
Cancer Subgroup						
Tonsil	11,924	41.34	8930	42.25	2994	38.83
Base of Tongue	9922	34.40	7356	34.80	2566	33.28
Other Oropharyngeal site	7000	24.27	4849	22.94	2151	27.90
Smoking						
Smoker	10,438	36.19	7705	36.46	2733	35.44
Non Smoker	18,408	63.81	13,430	63.54	4978	64.56
Alcohol						
Alcohol	5840	20.25	4613	21.83	1227	15.91
No Alcohol	23,006	79.75	16,522	78.17	6484	84.09
Comorbidity Score						
< 5	9457	32.78	6534	30.92	2923	37.91
5 to 10	1249	4.33	955	4.52	294	3.81
10 to 15	10,100	35.01	7628	36.09	2472	32.06
15 to 20	3388	11.75	2503	11.84	885	11.48
20 +	4652	16.13	3515	16.63	1137	14.75
Deprivation Score						
1	6956	24.11	5202	24.61	1754	22.75
2	6049	20.97	4402	20.83	1647	21.36
3	5692	19.73	4114	19.47	1578	20.46
4	5171	17.93	3766	17.82	1405	18.22
5	4689	16.26	3452	16.33	1237	16.04
Unknown	289	1.00	199	0.94	90	1.17
Total	28,846	100.00	21,135	100.00	7,711	100.00

4.5.10 Survival analyses:

Survival analyses of the oropharyngeal cancer cohort from the UK HES database was undertaken using Cox regression and Kaplan Meier survival analyses techniques. Demographic and risk factor variables were included in the analyses and hazard ratios (with 95% confidence intervals) provided as an estimate of the effect size. Statistical significance was set at a p-value less than 0.05. All OPC and by gender stratification were analysed using their age group, ethnicity, UK region of residence, cancer subgroup, deprivation score, Charlson co-morbidity index, smoking history, and alcohol consumption (Table 4.6).

4.5.10.1 Gender impact on survival:

There was no statistical difference by gender observed on survival in this oropharyngeal cancer cohort. With males as the reference group, the hazard ratio for females was 0.96 (95% C.I. 0.91 – 1.00) and a p-value of 0.079. (Table 4.6 and Figure 4.7).

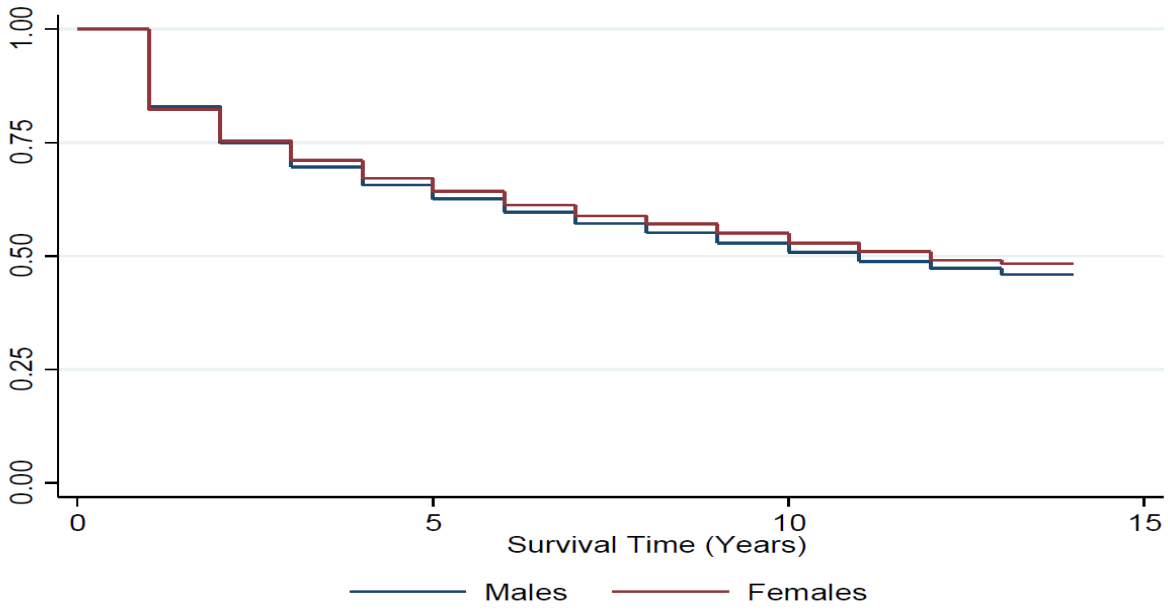


Figure 4.7: ***Kaplan Meier survival of oropharyngeal cancer by gender.*** *Survival shown in blue for males and in red for females.*

Table 4.6: Cox Regression analyses of oropharyngeal cancer and survival determinants.
Results of the analysis on prognostic factors for survival in OPC using UK HES data.

		All OPC		Males Only		Females Only	
		Hazard Ratio	P-Value				
Sex	Males	1 (baseline group)					
	Females	0.96 (0.91,1.00)	0.079				
Age Group	Under 40	1 (baseline group)		1(baseline group)		1 (baseline group)	
	40 to 49	1.25 (1.04,1.51)	0.019	1.07 (0.85,1.33)	0.576	1.71 (1.21,2.42)	0.002
	50 to 59	1.84 (1.54,2.20)	<0.001	1.59 (1.28,1.97)	<0.001	2.42 (1.74,3.37)	<0.001
	60 to 69	2.42 (2.02,2.90)	<0.001	2.18 (1.75,2.70)	<0.001	2.79 (2.01,3.89)	<0.001
	70 to 79	3.65 (3.04,4.39)	<0.001	3.33 (2.67,4.14)	<0.001	4.25 (3.04,5.94)	<0.001
	80 +	5.15 (4.24,6.25)	<0.001	4.61 (3.65,5.83)	<0.001	6.39 (4.52,9.04)	<0.001
Ethnic Group	White	1 (baseline group)		1 (Baseline Group)		1 (baseline group)	
	Mixed	1.45 (0.80,2.66)	0.223	1.22 (0.63,2.37)	0.558	2.73 (0.63,11.88)	0.18
	Indian	1.11 (0.54,2.27)	0.774	0.94 (0.40,2.21)	0.882	1.98 (0.42,9.30)	0.389
	Pakistani	1.41 (0.78,2.54)	0.262	1.37 (0.71,2.63)	0.346	1.87 (0.43,8.04)	0.403
	Bangladeshi	1.68 (0.67,4.20)	0.271	1.32 (0.42,4.16)	0.631	3.31 (0.55,19.84)	0.189
	Other Asian Background	1.77 (0.97,3.25)	0.065	1.84 (0.95,3.57)	0.072	1.81 (0.40,8.26)	0.444
	Black or Black British	0.89 (0.44,1.79)	0.735	1.00 (0.46,2.17)	0.992	0.83 (0.15,4.51)	0.826
	Chinese	0.82 (0.38,1.78)	0.624	0.72 (0.30,1.74)	0.464	1.40 (0.26,7.66)	0.696
	Other Ethnic Group	1.39 (0.81,2.40)	0.235	1.21 (0.67,2.20)	0.519	2.35 (0.59,9.39)	0.228
	Unknown	1.34 (0.78,2.32)	0.29	1.18 (0.65,2.13)	0.593	2.25 (0.56,9.05)	0.252
Region of Residence	North East	1 (baseline group)		1 (Baseline Group)		1 (baseline group)	
	North West	0.92 (0.84,1.01)	0.099	0.92 (0.82,1.02)	0.123	0.94 (0.78,1.13)	0.532
	Yorkshire and Humber	0.83 (0.75,0.92)	<0.001	0.83 (0.74,0.94)	0.003	0.82 (0.67,1.01)	0.06
	East Midlands	0.75 (0.67,0.84)	<0.001	0.76 (0.67,0.86)	<0.001	0.72 (0.58,0.90)	0.004
	West Midlands	0.88 (0.79,0.97)	0.013	0.90 (0.80,1.01)	0.073	0.83 (0.68,1.02)	0.073
	East of England	0.80 (0.72,0.88)	<0.001	0.81 (0.72,0.91)	<0.001	0.76 (0.62,0.95)	0.013
	London	0.92 (0.84,1.02)	0.113	0.92 (0.82,1.03)	0.136	0.94 (0.78,1.14)	0.546
	South East	0.77 (0.70,0.85)	<0.001	0.78 (0.69,0.87)	<0.001	0.76 (0.63,0.93)	0.006
	South West	0.80 (0.72,0.89)	<0.001	0.82 (0.73,0.92)	0.001	0.75 (0.61,0.93)	0.008
	No Fixed Abode	0.13 (0.04,0.42)	0.001	0.19 (0.05,0.76)	0.019	0.09 (0.01,0.61)	0.014
	Unknown	0.13 (0.07,0.25)	<0.001	0.16 (0.08,0.32)	<0.001	0.08 (0.02,0.32)	<0.001
Deprivation Score	1	1.64 (1.53,1.75)	<0.001	1.75 (1.61,1.89)	<0.001	1.36 (1.19,1.55)	<0.001
	2	1.32 (1.23,1.42)	<0.001	1.36 (1.25,1.48)	<0.001	1.23 (1.07,1.41)	0.003
	3	1.16 (1.08,1.25)	<0.001	1.21 (1.11,1.32)	<0.001	1.04 (0.91,1.20)	0.57
	4	1.05 (0.97,1.13)	0.238	1.08 (0.98,1.18)	0.112	0.98 (0.84,1.13)	0.75
	5	1 (baseline group)		1 (Baseline Group)		1 (baseline group)	
Cancer Subgroup	Tonsil	1.13 (1.02,1.25)	0.019	1.15 (1.02,1.30)	0.019	0.56 (0.50,0.62)	<0.001
	Base of Tongue	1.12 (1.07,1.17)	<0.001	1.09 (1.04,1.15)	0.001	0.86 (0.78,0.95)	0.003
	Other	1 (baseline group)		1 (Baseline Group)		1 (baseline group)	
Smoking	Smoker	1.57 (1.51,1.64)	<0.001	1.62 (1.54,1.70)	<0.001	1.45 (1.33,1.58)	<0.001
	Non Smoker	1 (baseline group)		1 (Baseline group)		1 (Baseline group)	
Alcohol	Alcohol	1.64 (1.57,1.72)	<0.001	1.64 (1.55,1.73)	<0.001	1.67 (1.51,1.84)	<0.001
	No Alcohol	1 (baseline group)		1 (Baseline group)		1 (Baseline group)	
HPV	HPV	0.39 (0.24,0.64)	<0.001	0.43 (0.24,0.78)	0.005	0.34 (0.14,0.82)	0.016
	No HPV	1 (baseline group)		1 (Baseline group)		1 (Baseline group)	
Comorbidity Score	< 5	1 (baseline group)		1 (Baseline group)		1 (Baseline group)	
	5 to 10	1.48 (1.33,1.65)	<0.001	1.44 (1.28,1.63)	<0.001	1.56 (1.25,1.94)	<0.001
	10 to 15	1.27 (1.20,1.34)	<0.002	1.22 (1.14,1.30)	<0.001	1.41 (1.26,1.56)	<0.001
	15 to 20	1.64 (1.53,1.76)	<0.003	1.59 (1.47,1.73)	<0.001	1.75 (1.52,2.00)	<0.001
	20 +	2.23 (2.10,2.37)	<0.004	2.13 (1.99,2.29)	<0.001	2.50 (2.22,2.80)	<0.001

4.5.10.2 Age impact on survival:

The OPC cohort was divided into age groups as follows: <40 years, 40 – 49 years, 50 – 59 years, 60 – 69 years, 70 – 79 years, ≥ 80 years. The age groups <40 years was used as the reference group. Survival analyses showed a statistically significant increase in the hazard for mortality with increasing age across all age groups. The hazard ratio increased steadily from 1.25 (95% C.I. 1.04 – 1.51) in the 40 – 49 years age group to four times the hazard at 5.15 (95% C.I. 4.24 – 6.25) in the ≥ 80 years age group (p-value <0.001) (Table 4.6 and Figure 4.8). This impact of increasing age was replicated regardless of gender with similar trends observed in both males and females (p<0.001).

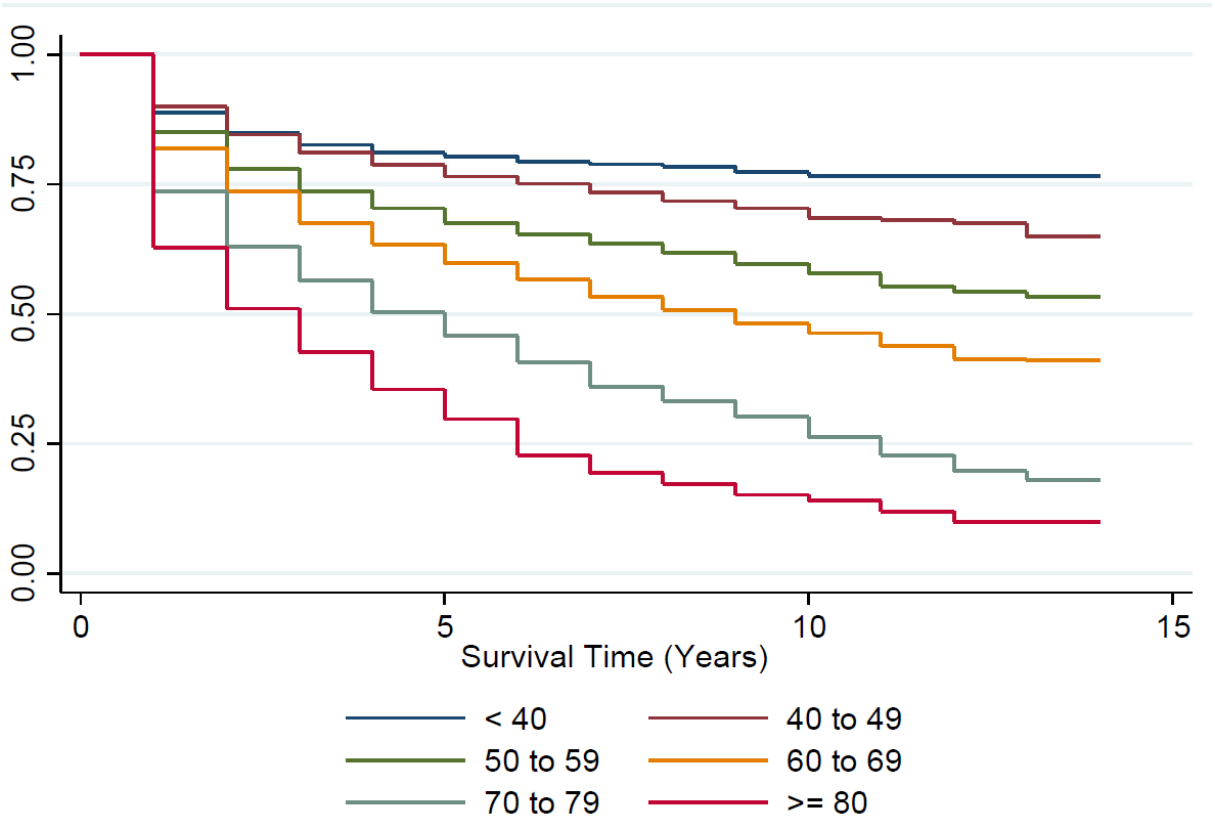


Figure 4.8: Kaplan Meier survival of oropharyngeal cancer by age groups. *Age is shown to be a significant prognostic factor in survival from OPC.*

The hazard ratio for mortality did not show a statistically significant increase in males aged 40 – 49 years at 1.07 (95% C.I. 0.85 – 1.33); p=0.576 but did show a consistent increase in males aged 50 – 59 years from 1.59 (95% C.I. 1.28 – 1.97) to a hazard ratio of 4.61 (95% C.I. 3.65 – 5.83) in males aged ≥ 80 years; p<0.001. The effect size was even larger in females with OPC with worsening survival with increasing age. The hazard ratio for females aged 40 – 49 years was 1.71 (95% C.I. 1.21 – 2.42) with a consistent increase in mortality hazard to a H.R of 6.39 (95% C.I. 4.52 – 9.04) in females aged ≥ 80 years; p<0.001 (Table 4.6 and Figure 4.8).

4.5.10.3 Ethnicity impact on survival:

The impact of ethnicity on survival in patients with oropharyngeal cancer was explored using the HES UK database. The 'White' ethnicity was used as the reference group. As expected, this was the predominant ethnicity with 23,597 cases (81.80%), mirroring the background UK caucasian population. There was no statistically significant impact of ethnicity detected on survival using Cox regression and Kaplan-Meier analyses. None of the p-values reached the 0.05 threshold. This was a consistent finding for both males and females (Table 4.6 and Figure 4.9). The number of patients in the other ethnic groups were small and so there is a high risk for a type II error and so we cannot draw any firm conclusions.

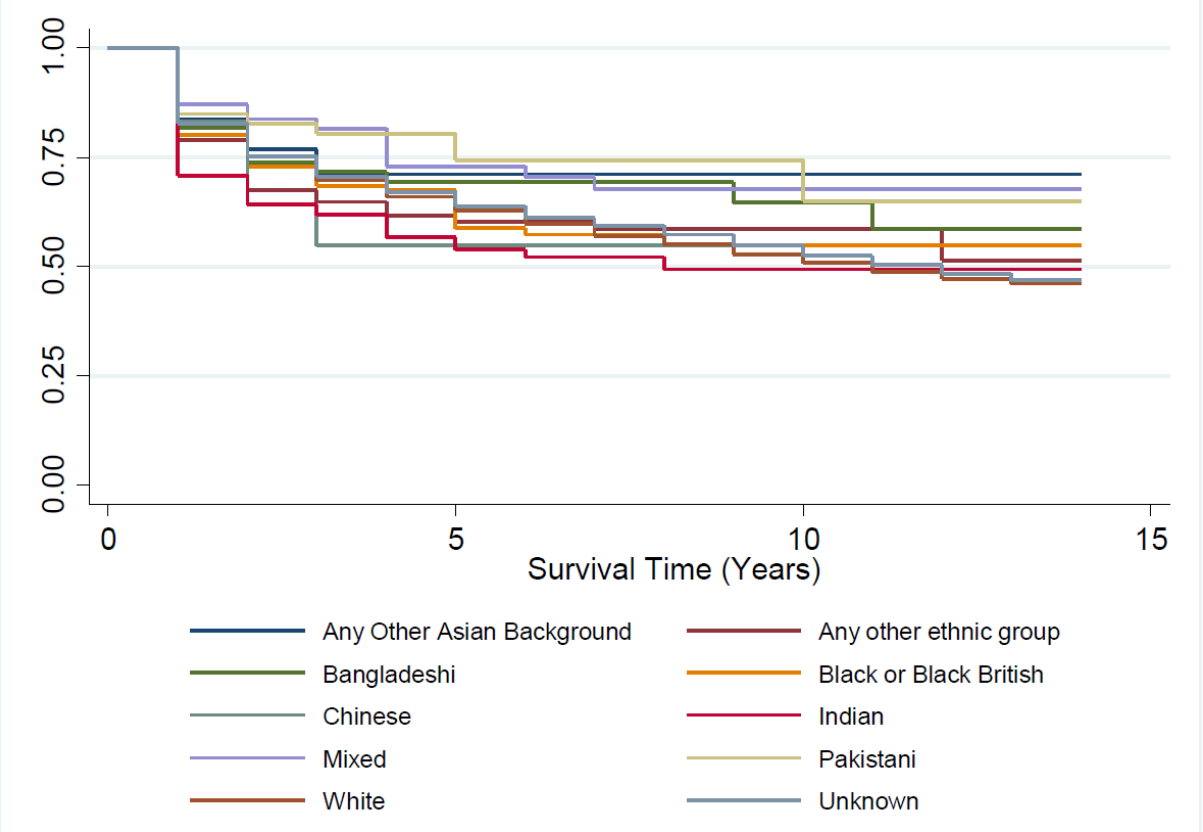


Figure 4.9: Kaplan-Meier survival of Oropharyngeal cancer by ethnicity. *There was no significant impact of ethnicity demonstrated due to small numbers in each ethnic group.*

4.5.10.4 Region of residence and impact on survival:

The impact of the region of residence of patients with OPC on their survival was investigated. The North East of England was designated the reference group. The North East had the lowest frequency of OPC over the study period with 1,825 cases (1,375 males and 450 females) representing 6.33% of the HES UK database (Table 4.6). The North East however had the highest incidence of OPC in the UK (Figure 4.6). The UK region previously shown to account for the highest proportion of oropharyngeal cancer in this HES UK database was however the North West with 4,778 cases (16.56%) (Table 4.4). The region of residence was 'unknown' in 201 OPC cases (0.7%) and 'no fixed abode' in 5cases (0.19%). Cox regression and Kaplan-

Meier survival analysis showed some disparity in survival across UK regions in comparison to the North East.

The UK regions with better survival outcomes in comparison to the North East region with statistical significance demonstrated at $p<0.001$ for hazard ratios included Yorkshire and Humber [H.R: 0.83 (95% C.I: 0.75 – 0.92)], East Midlands [H.R: 0.75 (95% C.I: 0.67 – 0.84)], East of England [H.R: 0.80 (95% C.I: 0.72 – 0.88)], South East [H.R: 0.77 (95% C.I: 0.70 – 0.85)], and South West [H.R: 0.80 (95% C.I: 0.72 – 0.89)]. West Midlands [H.R: 0.88 (95% C.I: 0.79 – 0.97)], also had a better survival outcome but with a p-value of 0.013. There was no statistical difference in survival in London [H.R: 0.92 (95% C.I: 0.84 – 1.02)] or in the North West [H.R: 0.92 (95% C.I: 0.84 – 1.01)].

In males the best statistically significant survival outcomes ($p<0.05$) were recorded in the East Midlands, South East, East of England, South West, Yorkshire and Humber (Table 4.6). There was no difference in survival amongst males in North West ($p=0.123$), West Midlands (0.073), and London ($p=0.136$).

Similarly in females, survival was better in the following UK regions ($p<0.05$) when compared to the North East: East Midlands, South West, South East, and East of England. There was no difference observed in survival in the North West ($p=0.532$), Yorkshire and Humber ($p=0.06$), West Midlands ($p=0.073$), and London ($p=0.546$) (Table 4.6).

4.5.10.5 Survival outcomes and deprivation scores:

This analysis crucially investigated the impact of deprivation on survival in patients with Oropharyngeal cancer. The degree of deprivation was categorised based on deprivation scores ranging from deprivation score 1 (most deprived) to deprivation score 5 (least deprived) in a

step-wise fashion with decreasing deprivation from score 1 to 5. Patients with oropharyngeal cancer identified on the HES database were uniquely stratified based on their deprivation scores on Kaplan-Meier and Cox-regression analyses. The survival outcome uniquely showed that survival worsens with increasing levels of deprivation (measured as deprivation scores). The least deprived patients with a score of 5 had the best survival outcome and were thus designated the reference group for univariate regression analyses.

When compared to the least deprived group (score 5), patients with the most deprivation (score 1) had a hazard ratio of 1.64 (95% C.I: 1.53 – 1.75) (Table 4.6). This was strongly statistically significant with $p<0.001$ and indicates that patients with the highest deprivation scores were 64 times increased hazard for dying when compared to those patients with the least deprivation scores; with 95% confidence that the real increase in this hazard lies between a range of 53 times to 75 times risk (Table 4.6). Similarly, patients with a deprivation score of 2 had an increased hazard of death when compared with the least deprived group [H.R: 1.32 (95% C.I: 1.23 – 1.42)] with a $p<0.001$. This trend continues in patients with a deprivation score of 3 albeit with a lesser hazard of dying at 16% compared with the 32% hazard in the group with a deprived score of 2. A deprivation score of 3 was associated with a H.R: 1.16 (95% C.I: 1.08 – 1.25) at a significant $p<0.001$ (Table 4.6 and Figure 4.10). There was no statistically significant difference between patients who had a deprivation score of 4 when compared with the reference group that had a deprivation score of 5. Deprivation score 4 was associated with a hazard ratio of 1.05 (95% C.I: 0.97 – 1.13) at $p=0.238$. Males had higher hazard ratios compared to females with the same deprivation scores indicating gender contributes to the observed effect.

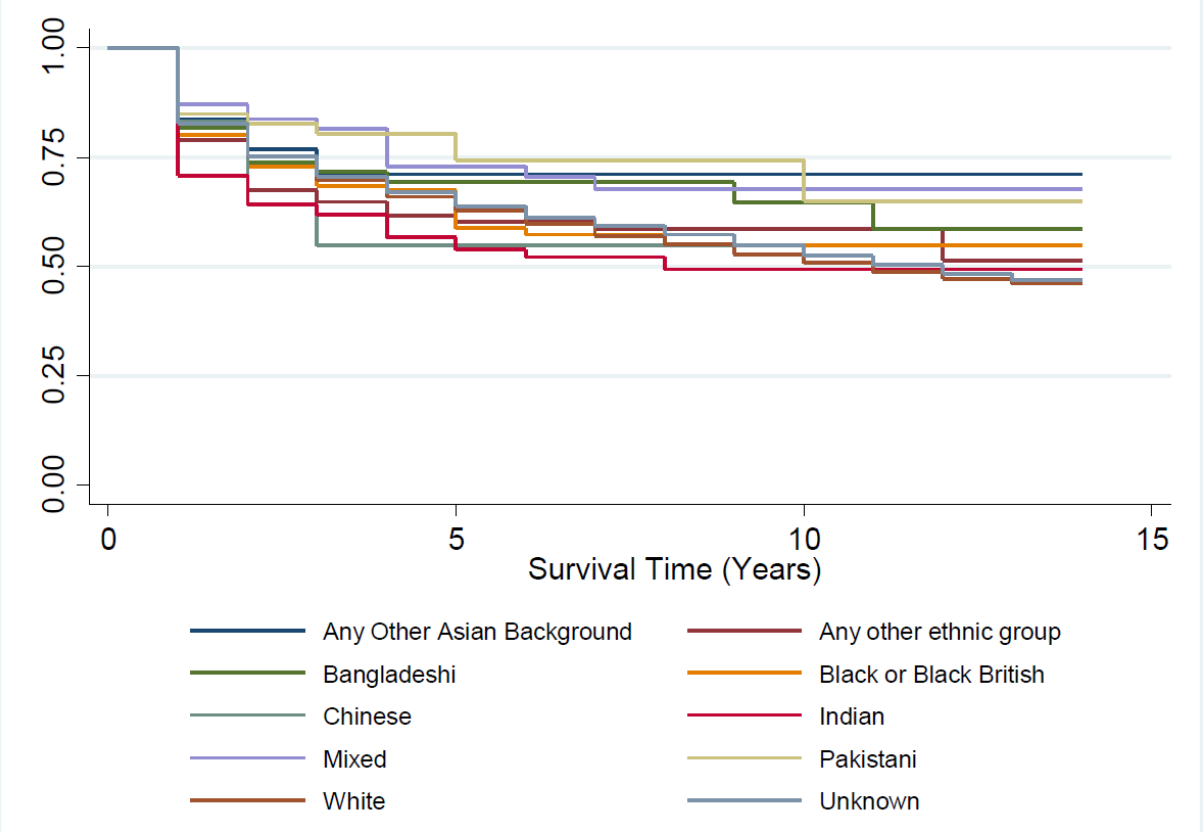


Figure 4.10: Kaplan Meier analysis of oropharyngeal cancer by deprivation scores. *Deprivation was shown to have a significant impact on survival from OPC in the UK HES.*

4.5.10.6 Impact of co-morbidity on survival:

The simultaneous presence of other conditions or diseases alongside the oropharyngeal cancer patients were documented as having on the UK HES was considered 'co-morbidity'. Co-morbidity was shown to have a significant effect on the survival of patients with oropharyngeal cancer, with increasing co-morbidity scores associated with worsening survival (Figure 4.11). The co-morbidities in this survival analyses have been previously discussed under the descriptive results section above and include presence of metastatic cancer, pulmonary disease, diabetes, acute myocardial infarction, cerebrovascular accidents, peripheral vascular disease, renal disease, and peptic ulcers. The *'Charlson's co-morbidity index'* was calculated for each patient as previously described with a score of 1, 2, 3 or 6

assigned to specific co-morbidities. Further categorisation into co-morbidity index groups depending on the aggregate scores per patient was as follows: [<5], [5 – 10], [11 – 15], [16 – 20], and [>20]. Unsurprisingly, patients with low levels of co-morbidity on the Charlson's index with a score <5 had the best survival in this oropharyngeal cancer cohort and were therefore designated the reference group for univariate regression analyses.

The group with co-morbidity index score [5 – 10], when compared with the low co-morbidity index group [<5] had an increased hazard of dying with H.R: 1.48 (95% C.I: 1.33 – 1.65) which was significant at $p<0.001$. In patients with co-morbidity index score between [10 – 15], there was also an increased hazard of dying of 27% observed when compared with the low co-morbidity reference group [H.R: 1.27% (95% C.I: 1.20 – 1.34)]; $p<0.002$. This trend increased exponentially in those patients with a co-morbidity index score >15. The group of patients with a co-morbidity index [15 – 20], had a 64% increased hazard of dying with H.R: 1.64 (95% C.I: 1.53 – 1.76); $p<0.003$ (Table 4.6 and Figure 4.11).

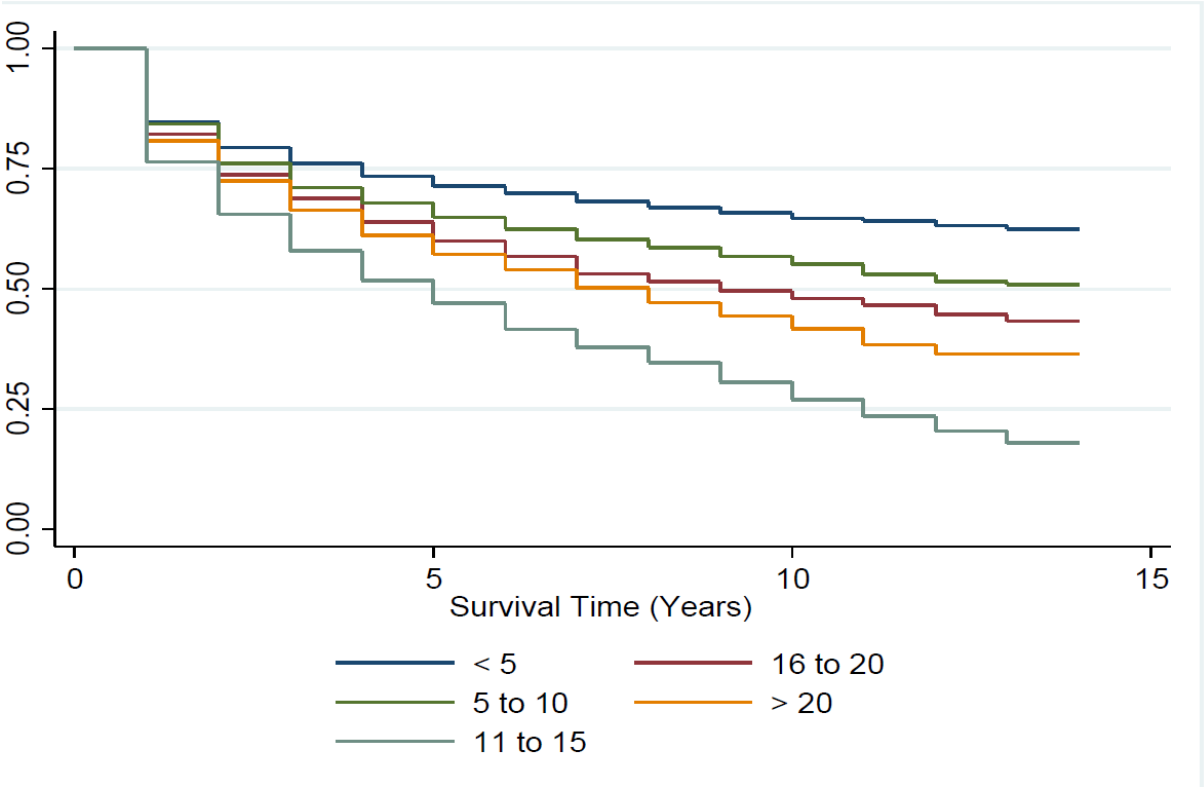


Figure 4.11: Kaplan Meier analysis of oropharyngeal cancer by co-morbidity index. *Co-morbidity had a significant impact on survival from OPC with mortality more likely with increasing number of co-morbidities present.*

The greatest effect on survival was observed in the patient group with a Charlson's co-morbidity index [>20] as they had a 123% increase in their hazard for dying when compared to patients with a low co-morbidity index [H.R: 2.23 (95% C.I: 2.10 – 2.37); p<0.004.

4.5.10.7 Behavioural risk factors impacting on survival:

Smoking is a recognised risk factor for head and neck cancers. In this UK HES cohort of oropharyngeal cancer, smoking maintained an expected significant effect on patient survival with smokers having a worse survival outcome when compared to non-smokers (Figure 4.12). Smokers had a 57% increased hazard of death in comparison to non-smokers (H.R: 1.57 (95% C.I: 1.51 – 1.64); a statistically significant finding with p<0.001 (Table 4.6).

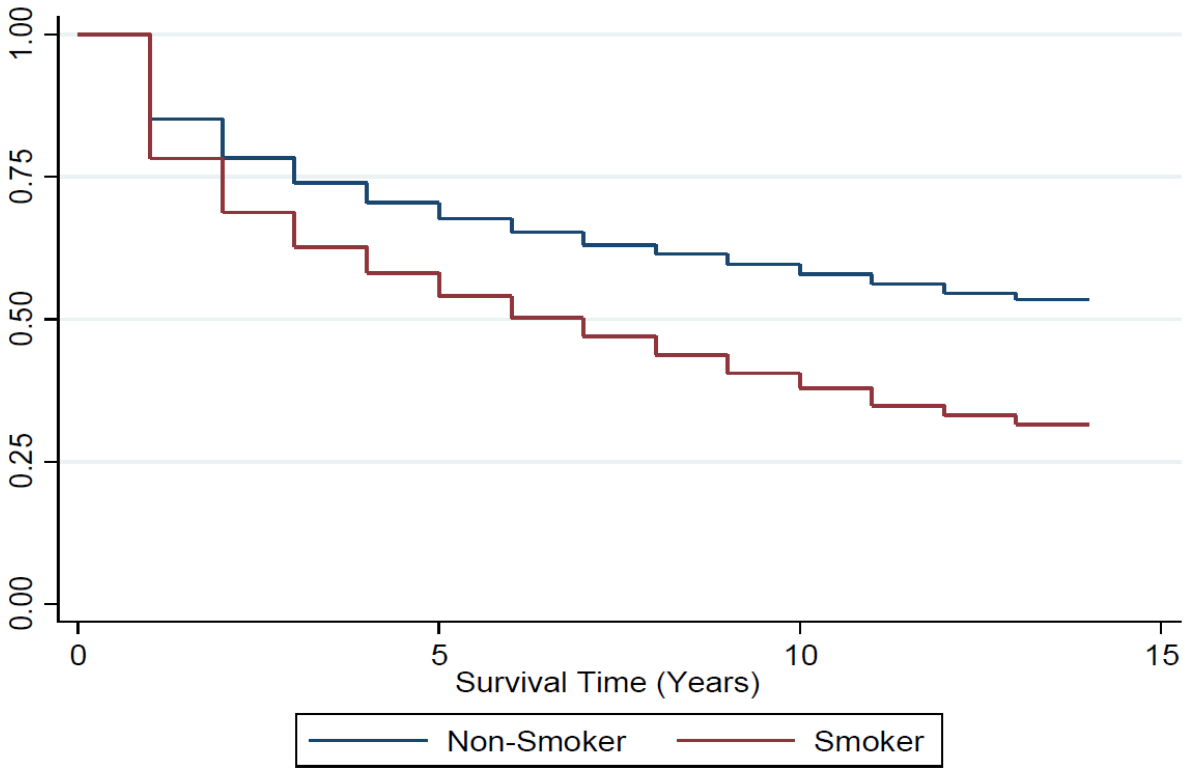


Figure 4.12: Kaplan Meier analysis of oropharyngeal cancer by smoking. *Smoking had a significant impact on survival with worse survival observed in smokers.*

The effect of smoking on worsening overall survival was maintained across gender divides with males [H.R: 1.62 (95% C.I: 1.54 – 1.70); p<0.001] and females [H.R: 1.45 (95% C.I: 1.33 – 1.58); p<0.001] (Table 4.6 and Figure 4.13).

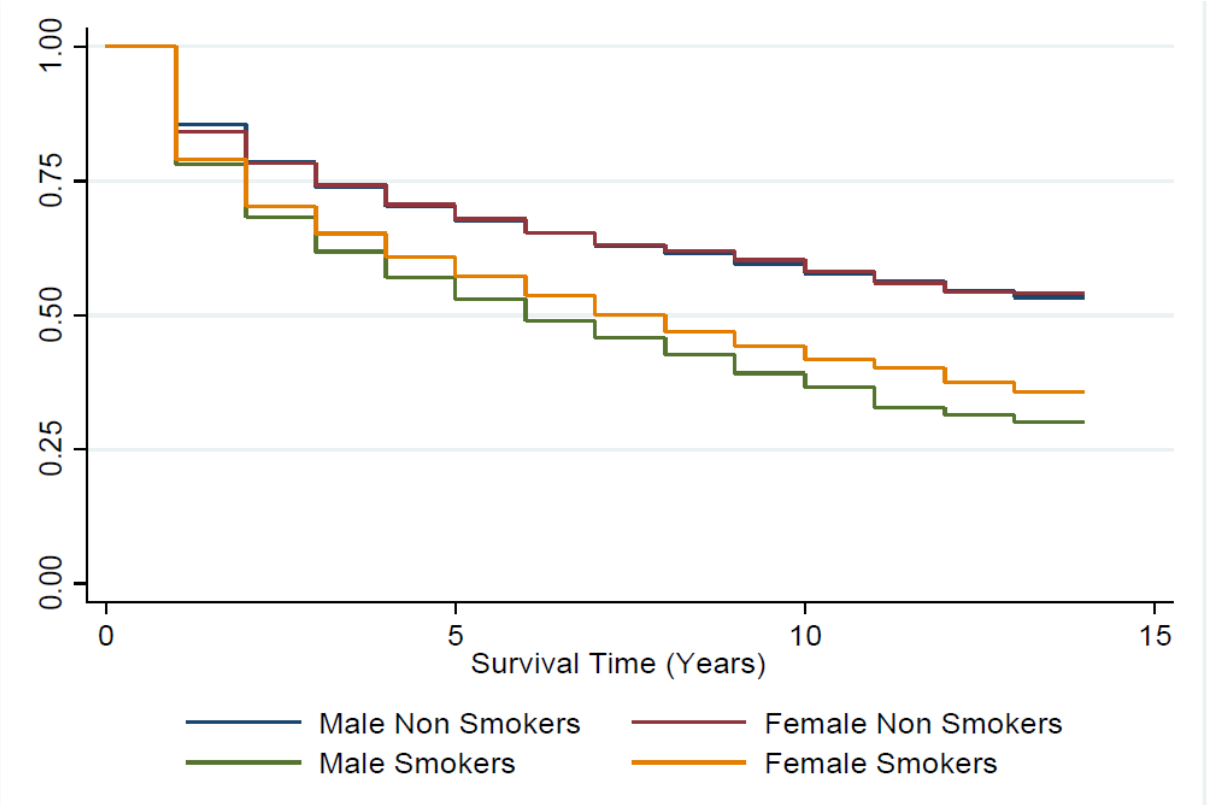


Figure 4.13: Kaplan Meier analyses by smoking and gender. *Smoking worsened survival in both males and females.*

Alcohol is also a risk factor known to work synergistically with smoking in the aetiology of head and neck cancers. Patients with oropharyngeal cancer and a record of alcohol intake on the UK HES database had a worse survival when compared with those patients who did not drink alcohol (Figure 4.14). Alcohol intake was associated with a 64% increased hazard of death in patients who drank compared to non-drinkers [H.R: 1.64 (95% C.I: 1.57 – 1.72); $p<0.001$] (Table 4.6).

The alcohol intake recorded on the HES database was however not quantified so a sub-analysis was not possible based on increasing alcohol intake or stratifying into mild, moderate or heavy drinkers. It was not also possible to establish the type of alcohol patients ingested in the wide spectrum from spirits to beers with little ethanol proportions.

Alcohol intake was further analysed based on gender and a clear demarcation does exist between alcohol drinkers and non-drinkers in both males and females (Figure 4.14 and 4.15).

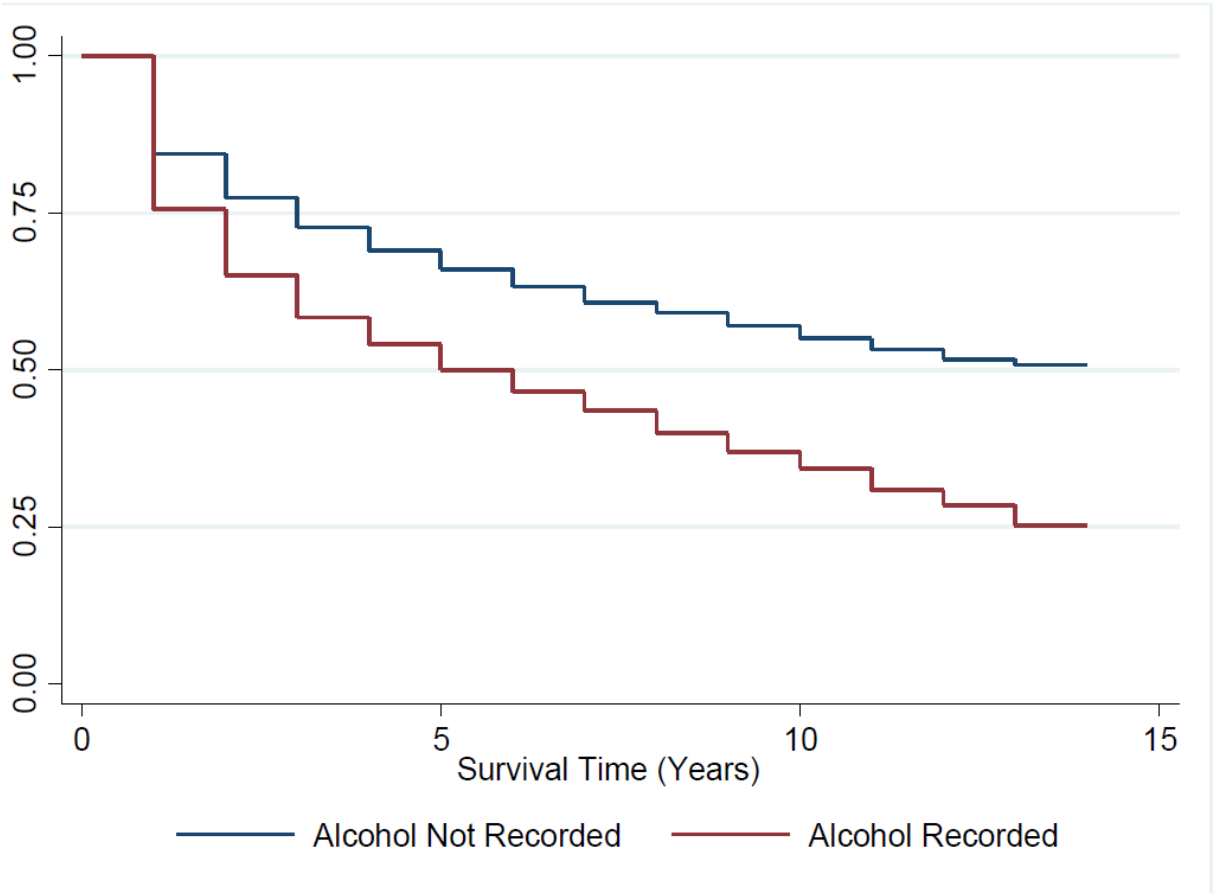


Figure 4.14: Kaplan Meier analyses by alcohol consumption. *Alcohol consumption was associated with worse survival in the OPC cases on HES. The alcohol consumption levels were however not quantified so clinical relevance unclear.*

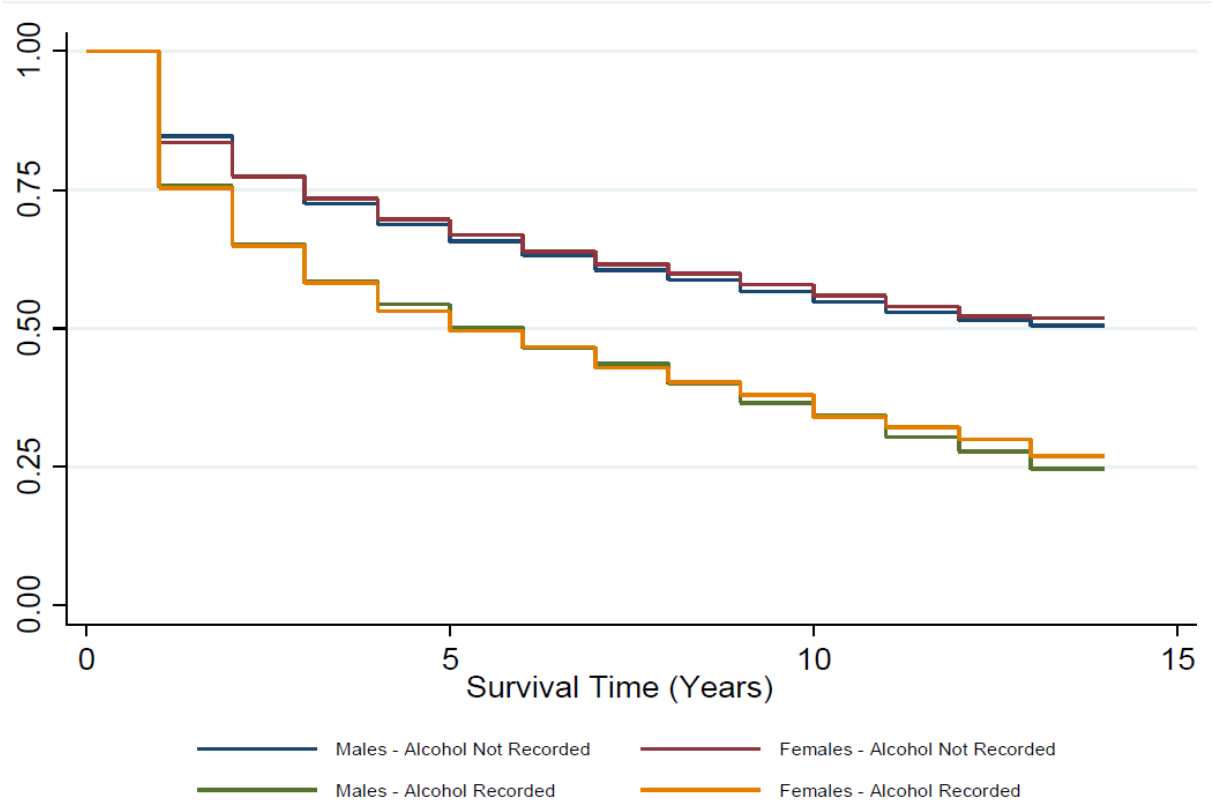


Figure 4.15: Kaplan Meier analyses by alcohol and gender. *Alcohol consumption was associated with worse survival in OPC cases on HES. However the quantities of alcohol consumed was not recorded.*

In male patients who also had a history of alcohol ingestion, there was a 64% increased hazard of dying in comparison to male non-drinkers [H.R:1.64 (95% C.I: 1.55 – 1.73); p<0.001]. Similarly, females who had alcohol intake recorded, had a 67% increased hazard of dying when compared to females without alcohol intake recorded on HES [H.R: 1.67 (95% C.I: 1.51 – 1.84); p<0.001] (Table 4.6 and Figure 4.15).

4.5.10.8 Oropharyngeal cancer subsites and survival:

The HES UK database was duly interrogated for the impact of anatomic subsite of oropharyngeal cancer on overall survival. Oropharyngeal cancer subsites principally include the palatine tonsils, lingual tonsils (or base of tongue), and other sites such as soft palate, uvula, and pharyngeal walls. This analysis broadly looked at comparative survival at those cancers primarily in the tonsils, base of tongue in relation to other oropharyngeal cancer subsites (as the reference group for regression). Generally, the risk was shown to be higher in tonsil cancers and base of tongue cancers (Table 4.6 and Figure 4.16).

In comparison to other OPC subsites, tonsil cancers had an increased hazard of death with H.R: 1.13 (95% C.I: 1.02 – 1.25); p=0.019. There was however a significant dichotomy in this trend when the analysis was stratified by gender. Males with tonsil cancer had a 15% increased hazard of dying when compared to males with OPC from other anatomic sites with a hazard ratio of 1.15 (95% C.I: 1.02 – 1.30) and this was significant at p=0.019. In sharp contrast, females with tonsil cancer had a reduced hazard of dying when compared with females with OPC from other anatomic sites at a hazard ratio of 0.56 (95% C.I: 0.50 – 0.62) and strongly significant at p<0.001 (Table 4.6).

Similarly, patients with base of tongue cancers had a poorer prognosis in overall survival when compared with the reference group of patients with cancers from other anatomic subsites in the oropharynx. Overall, those patients with tongue base cancers when compared to the reference group of other subsites, had a hazard ratio of 1.12 (95% C.I: 1.07 – 1.17); p<0.001. By gender however, there was a reversal of this trend in an identical fashion to the observed gender impact on tonsil cancers. Males with base of tongue cancers had a 9%

increase in the hazard of dying when compared with males with cancers of other oropharyngeal subsites [H.R: 1.09 (95% C.I: 1.04 – 1.15); p=0.001. In contrast, females with base of tongue cancers had a 14% reduction in their risk of dying when compared with females with cancers of other subsites [H.R: 0.86 (95% C.I: 0.78 – 0.95); p=0.003]. Therefore, although there was an impact of anatomic subsites on the overall survival from oropharyngeal cancer, this effect was sharply modified by gender, with females with either tonsil or tongue base cancers having a better prognosis than females with cancers of other subsites and conversely males with tonsil or tongue base cancers doing worse.

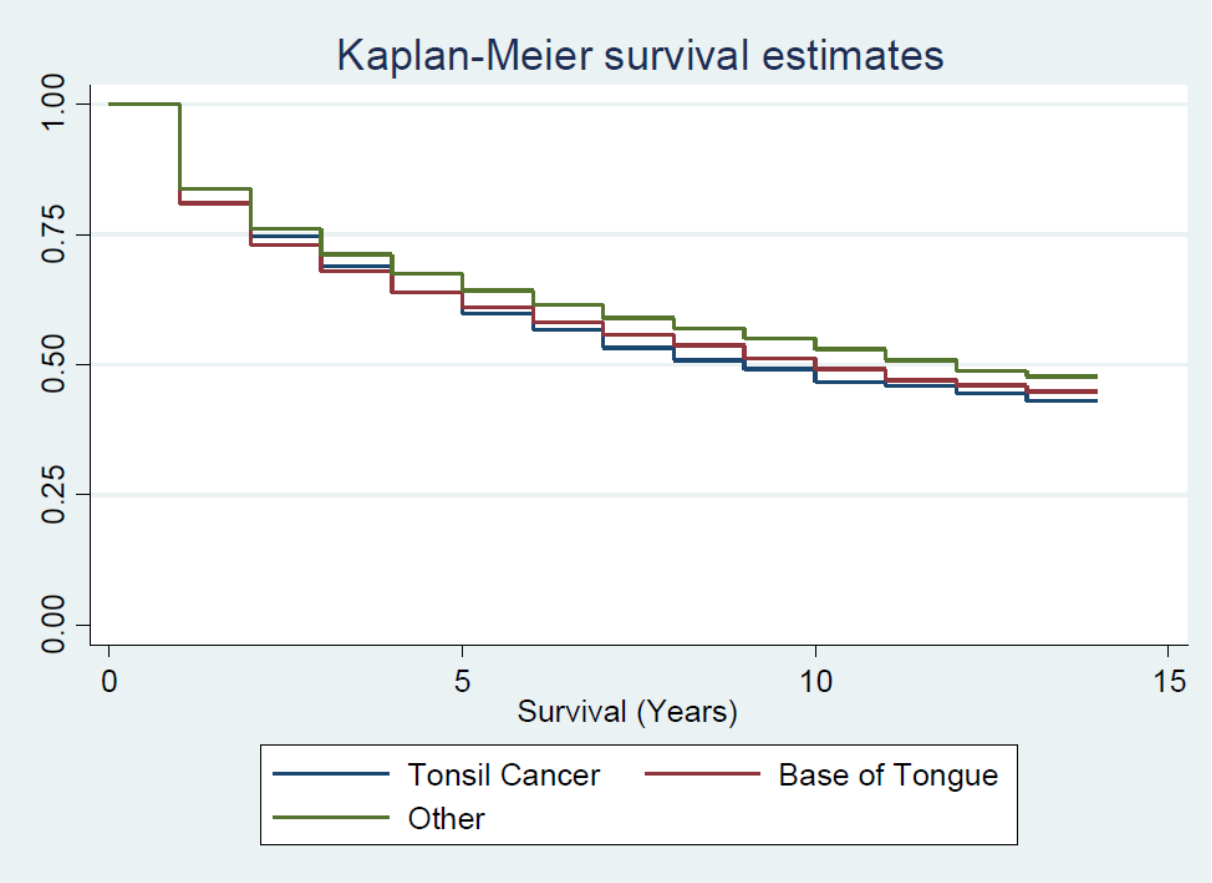


Figure 4.16: Kaplan Meier analyses by cancer subgroup. *Tonsil cancer shown in blue, base of tongue cancer in red and the other oropharyngeal subsites in green. There was no signficant difference observed in survival in OPC by subsite.*

4.6 Discussion:

The characterisation of oropharyngeal cancer using the UK HES database is important in broadening our understanding of the demographic factors associated with cancer states, the predisposing risk factors and the various determinants of survival. This UK-wide dataset uniquely covers a good proportion of oropharyngeal cancer (approximately 29,000 cancer cases) recorded over the last decade. Oropharyngeal cancer still remains a relatively rare cancer in the overall league tables of cancer diagnoses in the UK after lung cancer, breast cancer, prostate cancer, with a prevalence of 2.3 per 100,000 in 2010 (NCIN 2010). There is however documented evidence of the rising incidence of OPC in the developed world with a recent meta-analysis showing significant increase in America and across Europe in the last 30 years (Mehanna et al., 2010). Similarly, in the UK, the incidence of OPC doubled between 1996 and 2006 from 1.3 to 2.3 per 100,000 with this trend continuing in the last decade (NCIN 2010; Schache et al., 2016). Human papillomavirus 16 in particular has been identified as responsible for the vast majority of HPV-associated head and neck cancers (specifically oropharyngeal cancer). Whilst tobacco smoking and alcohol consumption still work in tandem as recognised risk factors for cancer development in the aerodigestive tract, HPV is now known to be causal for a subset of oropharyngeal cancer (IARC 2006).

4.6.1 OPC and increasing age:

The mean age at diagnosis of oropharyngeal cancer on the UK HES dataset was 60years with the entire dataset following a normal distribution. It is clearly a cancer that predominantly affects people aged over 40 years as these individuals accounted for 28,166 cancer cases (97.6%) of the entire cohort. The age group with the highest proportion of OPC was specifically the ‘50 – 59 years’ age group accounting for 9,642 cancer cases (33.4%). The

diagnosis of oropharyngeal cancer in this age group has been previously recognised. Gillison et al. (2016) described a bi-modal peak for oral HPV infection with the first peak observed when individuals become sexually active in their teenage years to early twenties, and the second peak around 50 – 60 years of age. It is instructive that this second peak of oral HPV infection coincides with the peak age of diagnosis for oropharyngeal cancer. A number of hypothesis have been put forward to explain the association between the two peaks of oral HPV infection and ultimately how this relates to cancer development. It is important to note that the latent period between oral HPV infection and OPC development is about 25 years so this crucially has meant a lack of longitudinal studies that can make concrete conclusions on the natural history of oral HPV infection. Two pathways of explanation are possible: the two oral HPV infection peaks as described by Gillison et al. (2016) are connected and represent latent infection that then overwhelms the host equilibrium in later life or the two peaks are independent and the second peak of infection represents 'new infection' in middle aged people at a time when their immune system is waning. The theory of 'new infection' with HPV in later life is predicated on changes in the sexual history of individuals with new partners, divorce, mid-life crises etc.

The literature available on HPV causation of cervical cancer is however unified on the observation that it is persistence of high-risk HPV infection that leads to irreversible cellular changes and then progression through carcinoma in-situ stages and invasive cervical cancer. Transferring this plausible argument to oropharyngeal carcinogenesis would suggest that persistent infection with high-risk HPV genotypes (particularly HPV-16) would be the precursor to HPV-associated oropharyngeal cancer. There is of course, also an on-going discourse around the tumour microenvironment and the role of the host immune system in HPV OPC and the truth is likely a combination of various changes that then allow HPV

replication with significant changes to the underlying transitional epithelium of the tonsillar crypts and invasive disease. The immune system is either overwhelmed by viral replication or the human papillomavirus reactivates after a latency period. The theory of reactivation is one that is recognised in other virally-mediated cancers especially Epstein-Barr virus (EBV) and nasopharyngeal cancers.

4.6.2 OPC and male predisposition:

The UK HES dataset shows a significant male predisposition for oropharyngeal cancer.

Males accounted for 73.3% of this UK cohort of oropharyngeal cancer, with a male: female ratio of 2.7: 1. This male preponderance is a feature of head and neck cancers generally and is partly because males tended to have exposure to tobacco smoking and excessive alcohol consumption – both of which act synergistically in cancer causation. The recent reality though is that HPV affectation in OPC tends to be in patients who are younger than HPV negative OPC and those affected often do not engage in smoking or alcohol ingestion. HPV is sexually transmitted with oral sex identified as an independent risk factor for HPV OPC alongside multiple lifetime sexual partners (D'Souza et al., 2009). Males are more likely to engage in 'high-risk' sexual practices that involve oro-genital contact and / or have multiple lifetime sexual partners. This could explain the disparity in gender predisposition for HPV OPC. Some authorities have explored the possibility that hormonal factors could be protective in females but no substantial conclusions have been published.

In the UK HES data, this male predisposition was maintained across all age groups analysed, with a proportion range between 58.9% - 75.9%.

4.6.3 OPC and ethnicity:

The UK HES dataset was unsurprisingly predominantly white caucassian (81.8%). There was no data on ethnicity in 15.08% of the dataset leaving all other ethnicity groups accounting for only <1% of the OPC cohort. Evidently, there was insufficient data to rigorously interrogate any differences that may exist by ethnicity or to calculate incidence but of note was the fact that the male predisposition in OPC persisted in the various ethnic groups included. Males accounted for the following higher proportions in the ethnic groups studied: Black or Black British (160/240 – 66.7%), Indian (118/170 – 69.4%), other ethnic group (128/179 – 71.5%), Mixed (60/91 – 65.9%), Pakistani (51/73 – 69.9%), other Asian background (46/60 – 76.7%), and Chinese (11/20 – 55.0%). The only ethnic group with equivocal OPC male proportions was Bangladeshi (33/67 – 49.3%).

Current world literature does indicate oropharyngeal cancer is prevalent across not just the developed world but also across the Asian continent. Crucially however, the HPV epidemic seems to have been more pronounced in the developed world. It is unclear how much HPV contributes as an aetiologic agent for OPC in Asia or Africa. The underlying risk factors in individuals from the Indian Sub-continent include the chewing of betel quid or areca nut alongside tobacco smoking. This has been put forward as a possible explanation for OPC prevalence in individuals who are resident in the UK but of Asian origins (NCIN 2010). These risk factors predispose to both oropharyngeal and oral cavity cancers. There is published evidence though on the risk of excessive alcohol consumption and cancer development in people from the Asian continent. There is a genetic susceptibility to glucose 6-dehydrogenase deficiency which results in an accumulation of toxic metabolites of ingested ethanol. Small quantities of alcohol can therefore cause significant risk for cancer when there is simultaneous exposure to tobacco smoking. American studies have shown that Black

Americans are at an increased risk of developing oropharyngeal cancer when compared to the Caucasian population. This is in part due to higher socioeconomic deprivation in Black American communities and behaviours that put them at increased risk. It is important to think about access to health care and the prevalent practices in the diverse ethnic groups in the UK as some communities may in fact be at higher risk for cancer development. It was not possible to calculate the incidence of OPC by ethnicity as the population data by ethnicity and region was not available.

4.6.4 OPC and regional variation:

Regional OPC proportions:

The UK HES dataset on OPC showed regional variation in the oropharyngeal cancer proportions documented in each UK region. OPC frequencies had a range of 6.35% - 16.56% over the timeframe recorded in the UK HES. The UK regions with the highest proportions of OPC (in descending order) were North West, South East and London. In contrast, the North East had the lowest proportion of OPC diagnoses. This finding is corroborated by the National Cancer Intelligence Network (NCIN 2010) report on head and neck cancers in England that indicated the highest prevalence of Oropharyngeal cancer was in the Cheshire Cancer Networks in North West England. Previous epidemiologic studies have confirmed an association between socioeconomic deprivation and high cancer prevalence rates. High levels of socioeconomic deprivation in parts of North West England may explain the high proportions of oropharyngeal cancer from the region. This unfortunately occurs simultaneously with exposure to tobacco smoking and alcohol consumption as poor communities are susceptible to these risk factors on the backdrop of malnutrition. The South East and London also had high rates of OPC documented in the UK HES cohort. These

regions are some of the most ethnically diverse regions of the United Kingdom with various communities having their own genetic profiles of cancer susceptibility as well as risk behaviours. For example, patients with origins in the Indian sub-continent have higher risks of oral cavity and oropharyngeal cancer due to betel quid chewing or ingestion of areca nut – risk factors for oropharyngeal carcinogenesis.

Regional OPC Incidence:

The incidence rates for oropharyngeal cancer in England, calculated from the HES database and population estimates from the Office of National Statistics clearly show an increasing incidence trend in England. The rising incidence of oropharyngeal cancer is also mainly driven by the higher rising incidence rates among males. This disparity in OPC incidence has been well-reported in the literature (Gillison et al., 2000; D,Souza et al. 2007). In general terms, head and neck cancers occur more frequently in men but it is currently unclear why oropharyngeal cancer in particular, with the HPV association in mind, has a predisposition for male affectation. Some authorities have suggested the observed differences could be due to high risk sexual practices that predispose males more to acquisition of oral high-risk HPV but hormonal and dietary factors need to be investigated further. The data from this study also shows that the OPC incidence in England has been increasing across all regions studied. The North East and North West of England however did consistently show the highest incidence rates year-on-year. The underlying reasons why this UK region is particularly at-risk for OPC is beyond the reach of the databases utilised in this study. The data however is corroborated by published studies from the National Cancer Intelligence Network (NCIN 2010) that show the highest rates of oropharyngeal cancer in the North of England and Cheshire cancer networks. A lower socio-economic balance with exposure to the risk factors of tobacco smoking and alcohol consumption are also important variables in this regional predisposition.

The age demographic is also important as shown in this data with London having comparatively lower incidence rates partly due to the higher proportions of young individuals in the nation's capital.

4.6.5 OPC and co-morbidity:

The risk of developing oropharyngeal cancer increases with increasing co-morbidity scores with attendant worse outcomes. The Charlston' Co-morbidity scoring system was useful in categorization this OPC cohort. Incredibly, 67.22% of this UK HES patient cohort had a co-morbidity score ≥ 5 and included a range of medical conditions. 1 in 5 patients also had pulmonary disease and just over half the entire cohort had metastatic cancer – two factors which become significant in terms of overall survival and response to treatment long-term.

4.6.6 OPC and behavioural risk factors:

In this UK HES cohort of oropharyngeal cancer, the majority were diagnosed in the palatine tonsils (41.34%) while tongue base tumours accounted for 34.4%. Both the palatine tonsils and the lingual tonsils (tongue base) are lymphoid tissue that are part of the *Waldeyer's ring of lymphoid tissue* in the oropharynx which perform a role in host immunity from birth but this role regresses after childhood.

A history of ever smoking was recorded in the UK HES in 36.19% of this OPC cohort while a history of alcohol consumption was found in 20.25%. These figures are surprisingly lower than expected with several published series of head and neck cancers indicating smoking and alcohol as main risk factors for OPC development. The available data on HPV status on HES was equally inconclusive largely in part because HPV testing has only recently been incorporated into the routine diagnostic testing for OPC in the UK.

Ang et al. (2010) published seminal work that stratified oropharyngeal cancer into 3 groups based on exposure to the risk factors of smoking and positivity to HPV. 260 cases of oropharyngeal cancer were studied, of which 187 were p16 positive. Based on their smoking history (greater or lesser than 10 pack years), TNM stage, and p16 test result, these patients with OPC were divided into low-risk, intermediate risk and high risk groups. Remarkably, 3 year overall survival in the low-risk, intermediate risk and high-risk groups were 94%, 67%, and 42% respectively (Ang et al., 2010).

4.6.7 OPC and determinants of overall survival:

Kaplan Meier and Cox regression analyses of this UK HES dataset showed overall survival was worse with increasing age, worsening deprivation, smoking, alcohol consumption, and increasing co-morbidity scores. There was no demonstrable difference in overall survival by gender or ethnicity.

4.6.8 Study strengths:

This study using the UK HES database captures the burden of oropharyngeal cancer that require diagnosis and treatment in hospital. It has provided useful insights into the demographic variables that are associated with OPC including age at diagnosis, gender predisposition, and the impact of deprivation and multiple co-morbidities. Crucially, this study highlights significant determinants of overall survival in patients with oropharyngeal cancer. Survival worsens with increasing age of the patient, their socioeconomic status and worsening deprivation, a history of smoking and alcohol consumption and the presence of multiple co-morbidities.

4.6.9 Study limitations:

The main limitation of the UK HES database lies in the inherent challenges of obtaining adequate cancer records from each hospital with the required data including demographic variables, risk factors, and survival (disease-free and recurrence-free). For this investigation into oropharyngeal cancer nationwide, there was inadequate data available as regards risk factors such as smoking, alcohol consumption levels and HPV status. Smoking and alcohol consumption could not be stratified by the quantities of exposure that could have provided deeper insight into cancer association e.g. ex-smoker, on-going smoking, smoking greater or lower than 10 pack years, type of tobacco exposure, etc. Alcohol consumption could not also be subdivided into mild, moderate or heavy drinking or even into the type of alcohol ingested (spirits, beers, wines). A simplistic association with ever having had exposure to either smoking or alcohol consumption is flawed on these grounds and requires further analyses along the lines of quantification of exposure levels.

The UK HES database did not convincingly provide the required data on HPV status in this cohort of oropharyngeal cancer. HPV as an aetiology of a subset of oropharyngeal cancer has only recently been recognised with HPV testing incorporated into routine diagnostic work-up of oropharyngeal cancer in many NHS Trusts. The responsible professional umbrella associations for Otolaryngology, Head and Neck Surgery as well as Maxillofacial Surgery in the UK e.g. ENT UK, BAHNO (British Association of Head and Neck Oncologists) and BAOMS (British Association of Oral and Maxillofacial Surgeons), have now adopted the recommendation to ascertain HPV status for all oropharyngeal cancers (Paleri et al., 2016).

PhD CHAPTER 5

CHARACTERISATION OF THE HUMAN PAPILLOMAVIRUS 16 GENOME AND EPIGENOME IN OROPHARYNGEAL CANCER

5.1 Background and hypothesis:

We know from cervical cancer research that when the HPV E2 gene is disrupted in a mixed (intact and episomal) infection, the lower concentrations of E2 protein produced will preferentially bind to the high-affinity, promoter activating E2BS1 with over-expression of viral oncogenes E6 and E7 (Chaiwongkot et al., 2013). This partly explains tumorigenesis in oropharyngeal cancer however, in contrast to cervical cancers, HPV positive oropharyngeal cancer appear to be more episomally driven tumours with an intact E2 gene in a significantly larger proportion (Reuschebach et al., 2015; Barbieri et al., 2014). Therefore, it is plausible that other mechanisms are in play that mediate carcinogenesis aside from the viral physical status. Methylation is one such mechanism that merits investigation – in particular because increasing methylation has been shown to be associated with increasing grade of cervical cancer severity (Chaiwongkot et al., 2013). It remains largely unexplored in HPV associated oropharyngeal cancer with only a few studies currently available on the subject.

Given the limited literature available on the impact of the HPV physical status and the methylation profiles at the E2 binding sites of the HPV 16 URR, this research focused on investigating the hypothesis:

''Carcinogenesis in those oropharyngeal cancer with an intact E2 gene is mediated by differential methylation at the E2 binding sites in the HPV 16 upstream regulatory region''.

As described in the Introduction (sections 1.15 and 1.17) in Chapter 1, E2 binding to the HPV 16 upstream regulatory region (URR) occurs at four locations. E2BS1 is the high-affinity binding site at CpG positions 7453 and 7459 which stimulates the p97 promoter. In contrast, the low-affinity binding sites E2BS3 (CpG 37, 43) and E2BS4 (CpG 52, 58) are inhibitory to the p97 promoter. E2BS2 at CpG site 7460 is responsible for canonical replication of the viral

genome but has not been investigated in this thesis as it has not been shown to be significantly involved in regulation of transcription (Thierry 2009; Reuschenbach et al., 2015). CpG methylation has been shown however to directly inhibit binding of the human papillomavirus type 16 E2 protein to specific DNA sequences (Thain et al., 1996). The molecular differences in E2 binding sites seen in sequence variations could explain the differences in affinity with E2BS1 showing high-affinity and promoter stimulation while E2BS3 and E2BS4 are inhibitory.

5.2 Study aims:

To investigate the correlation of clinical outcomes (cancer overall survival, recurrence and behavioural risk factors) with HPV 16 viral physical status and aberrant methylation at E2 binding sites in a cohort of HPV positive oropharyngeal cancers.

5.3 Study objectives:

1. To establish the integrity of the HPV 16 regulatory E2 gene (mixed or disrupted) in HPV positive oropharyngeal cancers using type-specific PCR with overlapping primers to the HPV 16 E2 open reading frame and URR in a number of samples.
2. To determine the methylation profiles at E2 binding sites (E2BS1, E2BS3 and E2BS4) in the HPV 16 upstream regulatory region, using pyrosequencing of bisulfite-modified DNA from HPV positive oropharyngeal cancers.
3. To confirm the proportion of the HPV oropharyngeal cancer samples that are transcriptionally active with expression of E2, E6 and E7 mRNA using reverse-transcription real-time PCR on extracted RNA.

4. To correlate the HPV 16 E2 physical status (mixed or disrupted), the methylation profiles in each E2 binding site, and the RNA expression data with epidemiologic and clinical outcomes such as demographic variables, risk factors' data, oropharyngeal cancer recurrence and overall survival from HPV positive oropharyngeal cancer.

5.4 Methodology:

5.4.1 Patient samples:

The original samples were obtained from FFPE (Formalin-Fixed Paraffin Embedded) biopsies taken from patients with oropharyngeal cancer entered into a multicentre UK study investigating the utility of the positron emission computerised tomography scan in monitoring cancer recurrence (PET-NECK) (Mehanna et al., 2016). HPV status was determined using adherence to rigorous protocols in a centralised laboratory at the Newcastle University Hospital by two independent pathologists using p16 immunohistochemistry (IHC) (CINtec Histology kit, Roche) and high risk HPV DNA in-situ hybridisation. Samples were positive on p16 IHC if there were >70% of the malignant cells with strong diffuse nuclear and cytoplasmic staining. Samples were classified as HPV positive if they were positive on both p16 immunohistochemistry and in-situ hybridisation. These initial HPV tests (p16 immunohistochemistry and high risk in-situ hybridisation for HPV DNA) were not performed by myself as I was given FFPE samples for which the HPV status was already determined.

40 FFPE blocks with known status as HPV positive oropharyngeal cancer were provided to me from Prof Mehanna's laboratory for this genetic / epigenetic study. The workflow of laboratory techniques utilised in downstream applications is shown in Figure 5.1.

5.4.2 Setting:

The PET-NECK trial was a phase III, randomised controlled multicentre UK study incorporating 43 NHS hospitals from across the West Midlands, East Midlands, London, Wales and Scotland and funded by the NIHR. The PETNECK trial assessed the non-

inferiority of PET-CT surveillance when compared with neck dissection in patients with advanced head and neck cancer (Mehanna et al., 2016).

5.4.3 Ethical approval:

Ethical approval for the PET-NECK study and further genomic studies using patient samples with valid consent was granted by the Oxfordshire Research Ethics Committee A (ISRCTN: 13735240).

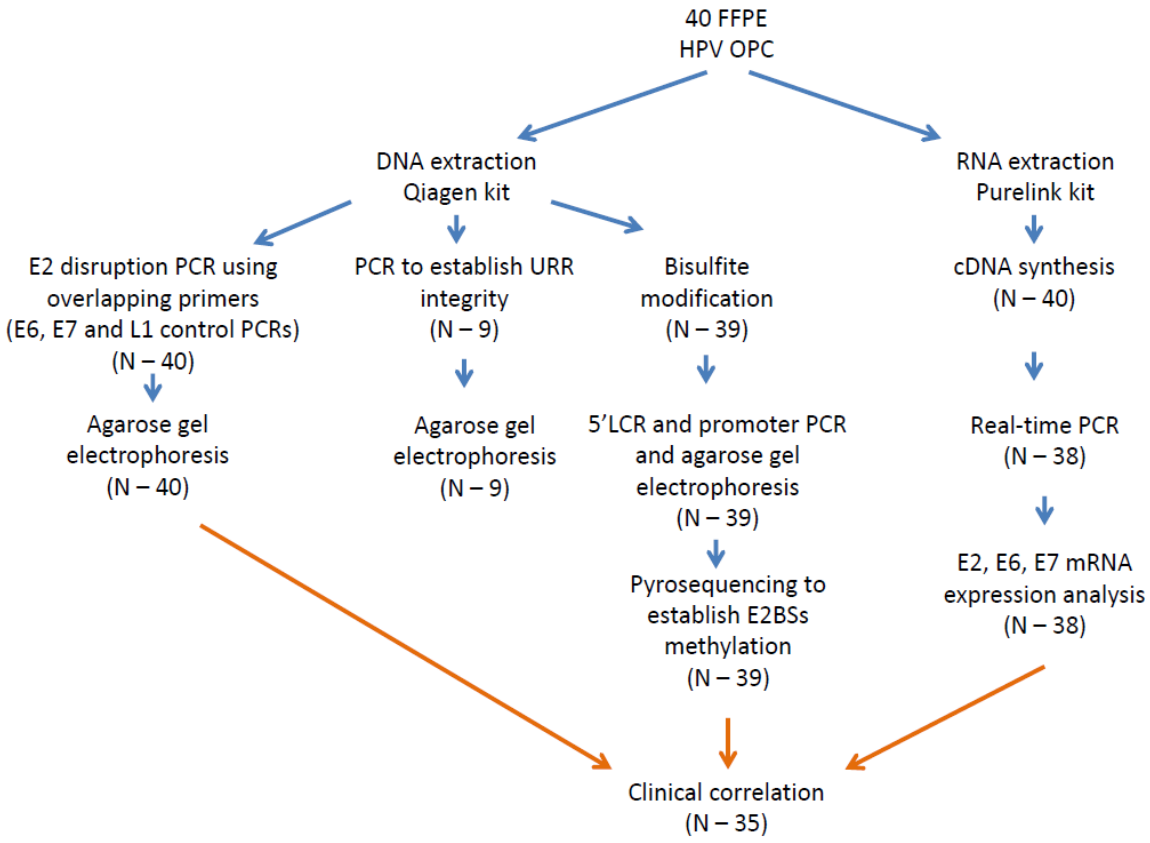


Figure 5.1: Workflow – Laboratory Techniques

5.4.4 Data analysis:

All data were entered into an Excel spreadsheet which included HPV 16 E2 status, methylation proportion at each E2 binding site, RNA expression data (E2, E6 and E7) and clinical outcomes. The viral physical status, methylation profiles and RNA expression were correlated with both demographic and cancer clinical outcomes (recurrence and survival) using STATA SE. Statistical analyses included Chi-squared test (X^2) for dichotomous variables such as E2 status with gender, smoking history, cancer recurrence. Student T-test was used to compare means of continuous variables such as age in the cohorts of E2 intact or disrupted. Time-to-event analyses using Kaplan-Meier curves were used for determining impact of E2 status and E2BS methylation on cancer recurrence or overall survival. Spearman's correlation coefficient was used to assess the degree of correlation in the methylation frequencies observed in the CpG site pairs in each E2 binding site. Mann-Whitney U test was used to check for any statistically significant differences between the mean methylation at each E2BS when comparing those cancer samples with disrupted E2 gene with those that had an intact E2 gene. Statistical significance was set at a value of $p<0.05$.

The flow chart for the entire study is depicted in Figure 5.2 below detailing the steps of enquiry.

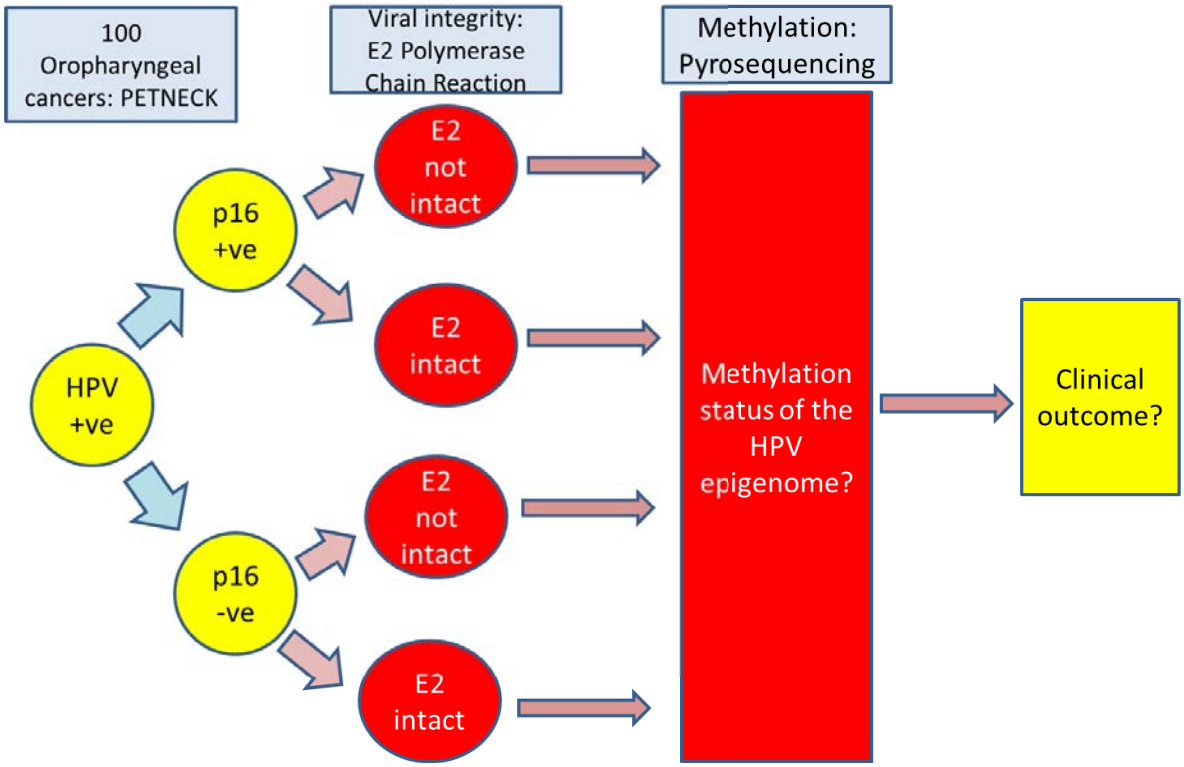


Figure 5.2: The flow chart of HPV16 genetics and epigenetics study in OPC

5.5 Laboratory methods:

5.5.1 FFPE sections using microtome:

Each FFPE section was suspended on the microtome and any excess paraffin trimmed. A fresh blade was used between samples to minimise the risk of cross-contamination. 5 – 8 scrolls per FFPE block were taken and each scroll measured 5–10 µm in thickness. The first 2 scrolls were discarded to avoid contamination especially if the blocks had been exposed to air. Each scroll was then transferred into a universal 1.5mL microcentrifuge PCR tube that was labelled.

5.5.2 DNA extraction from FFPE HPV OPC (Qiagen protocol adapted):

DNA extraction was performed using the Qiagen protocol for FFPE. As preparation for DNA extraction, 1ml of xylene solution was added to each microtube containing oropharyngeal cancer scrolls. The lid was closed and the sample mixed by vortexing for 10seconds. Each sample was centrifuged at 20,000 x g (14,000 rpm) for 2 minutes at room temperature (15–25°C). The solution was removed by pipetting and the pellet retained. 1ml of ethanol (96% - 100%) was added to the pellet and mixed by vortexing. In this step, the ethanol was used to extract residual xylene from samples. Each sample was placed in the centrifuge and spun at 20,000 x g (14,000 rpm) for 2 minutes at room temperature. The supernatant was removed by pipetting without removing any pellets. Residual ethanol was pipetted out using a fine tip. The samples were then incubated for 10minutes at room temperature or until all the residual ethanol had evaporated.

180 µl of Buffer ATL was used to resuspend the pellet obtained. 20 µl proteinase K was added and the sample vortexed and incubated at 56°C for 1 h (or until the sample has been completely lysed). The samples were centrifuged briefly to remove any residual drops from the lid and allowed to cool to room temperature. To obtain RNA-free genomic DNA, 2µl RNase A (100 mg/ml) was added and incubated for 2 minutes at room temperature. 200 µl Buffer AL was next added to the samples, and mixed thoroughly by vortexing. Subsequently 200 µl of ethanol (96–100%) was added, and vortexed to obtain a homogenous solution. A white precipitate may form on addition of Buffer AL and ethanol but it does not interfere with the extraction process.

After briefly centrifuging to remove liquid droplets from the lid, the entire lysate was transferred carefully to the QIAamp MinElute column (attached to a 2ml collection tube),

without wetting the rim. The samples were centrifuged at 6000 x g (8000 rpm) for 1 min. The QIAamp MinElute column was placed in a clean 2 ml collection tube while the flow-through was discarded. Repeat centrifugation can be performed at a higher speed if there is incomplete passage of the lysate through the membrane.

In sequence, each QIAamp MinElute column was opened and 500 µl Buffer AW1 was added without wetting the rim. The samples were centrifuged at 6000 x g (8000 rpm) for 1 min. The column was placed in a clean 2 ml collection tube and the flow-through discarded. This process was then repeated using 500 µl Buffer AW2. Care must be taken to avoid contact between the QIAamp MinElute column and the flow-through (which contains ethanol). This can occur if the centrifuge rotors vibrate when decelerating or when the collection tube is being retrieved from the centrifuge.

After adding Buffer AW2, the samples were centrifuged at full speed (20,000 x g; 14,000 rpm) for 3 min to dry the membrane completely. This ensured residual ethanol did not interfere with downstream applications. The QIAamp MinElute column was then placed in a fresh 1.5ml microcentrifuge tube and the collection tube with the flow-through was discarded. 20–100 µl Buffer ATE (at room temperature) was then carefully applied to the centre of the membrane to facilitate complete elution of bound DNA. The volume of eluate was determined by the downstream application but keeping in mind the final volume of eluate retrieved is approximately 5 µl less than that applied. The column lid was closed, incubated at room temperature for 1 minute and then centrifuged at full speed (20,000 x g; 14,000 rpm) for 1 min. Incubating with Buffer ATE for 5 min at room temperature before centrifugation generally increases DNA yield so should be considered.

5.5.3 RNA extraction from FFPE HPV OPC: (The PureLink™ FFPE Total RNA Isolation protocol adapted):

The paraffin from thin sections of a FFPE specimen was melted using heat in the presence of a specially designed Melting Buffer. The tissue was separated from the melted paraffin by centrifugation and digested with Proteinase K. The tissue lysate was further processed by selective binding of RNA to a silica-based membrane in the spin cartridge (PureLink™). Impurities were removed by thorough washing with Wash Buffer. The total RNA was eluted in RNase-Free Water. The expected total RNA yield exceeds 1µg from 3-8 FFPE sections of 10 µm each. The RNA yield is dependent on tissue type and quality of the specimen.

De-paraffinization and lysis:

3 – 8 scrolls of 10 µm FFPE samples were placed in a sterile RNAase-free 1.5ml microcentrifuge tube. 300µl of the melting buffer was added to the tube and centrifuged at maximum speed for 10 – 20 seconds. The samples were then incubated at 72°C for 10 minutes with intermittent gentle mixing and tapping every 2 – 3 minutes to ensure paraffin melting and that the scrolls remained under the melting buffer. 20 µl Proteinase K is then added and incubated at 60°C for 10-60 minutes with occasional mixing (Figure 5.3). On occasion, it was necessary to extend the incubation time by an additional 30-60 minutes and up to 3 hours, until lysis is complete. After this step, the samples were centrifuged at 20,000 x g (14,000 rpm) for 1 minute with a thin paraffin layer forming at the top, separate from the lysate below. The lysate was aspirated by pushing through this paraffin layer using a clean 1ml pipette tip, and then transferred to a clean RNAase-free tube.

Purification procedure (binding step):

400µl Binding Buffer (L3) and 800µl 100% ethanol was added to each sample (after de-paraffinization and lysis) and mixed well. 700µl from this sample mix was then transferred to a spin cartridge in a collection tube. This was centrifuged at 800 x g for 1 minute, the flow-through was discarded and the cartridge re-inserted into the collection tube. The remaining unspun sample (approximately 700µl) was then transferred to the spin cartridge in the collection tube, the centrifugation process repeated and the flow-through discarded (Figure 5.3).

Purification procedure (washing step):

20 ml Wash Buffer (W5) was added to 80ml of 100% ethanol and stored. 500µl Wash Buffer (W5) (with ethanol) is added to the cartridge. The cartridge was centrifuged at maximum speed for 1 minute and the flow through discarded. This step was repeated twice for a total of 3 washes. Each time the flow-through was discarded and the cartridge reinserted into the wash tube. Finally the cartridge was centrifuged at maximum speed for 1 minute to ensure all residual wash buffer with ethanol was removed. The wash tube was then discarded.

Elution step:

An aliquot of RNase-free water was pre-heated to 65°C. The spin cartridge was placed in a clean 1.7ml recovery tube and 50µl of the pre-heated RNase-Free water added to the centre of the cartridge. The cartridge was incubated for 1 minute and centrifuged at maximum speed for 1 minute. The recovery tube ultimately contained the total RNA. A second elution with 50µl of RNase-free water (if desired) can be performed as it may increase total RNA yield by 20 – 30%. The cartridge was removed and discarded. The tube containing the total RNA was

placed on ice prior to commencing downstream applications or stored at –80°C until needed (Figure 5.3).

DNase I digestion (protocol adapted from 5 – 8µl RNA sample):

DNase I digestion can be performed either with an aliquot of each of the purified RNA samples or with the entire amount of the purified RNA samples. In this study, a sterile tube was placed on ice and the following components added:

- Purified RNA sample: 28uL
- DNase I, Amplification Grade (1 U/ µl): 3.5uL
- 10X DNase I Buffer: 3.5uL
- Total volume: 35uL

The components were mixed by vortexing, spun briefly and incubated at room temperature for 5 – 15 minutes. Samples were centrifuged briefly and add 3.5uL of 25mM EDTA added to each tube on ice. Following further mixing by gently pipetting up and down, the samples were centrifuged and incubated at 65°C for 10 minutes. On completion of incubation, a final centrifugation was performed and each sample tube placed on ice prior to reverse transcription or storage of the RNA at –80°C until needed.

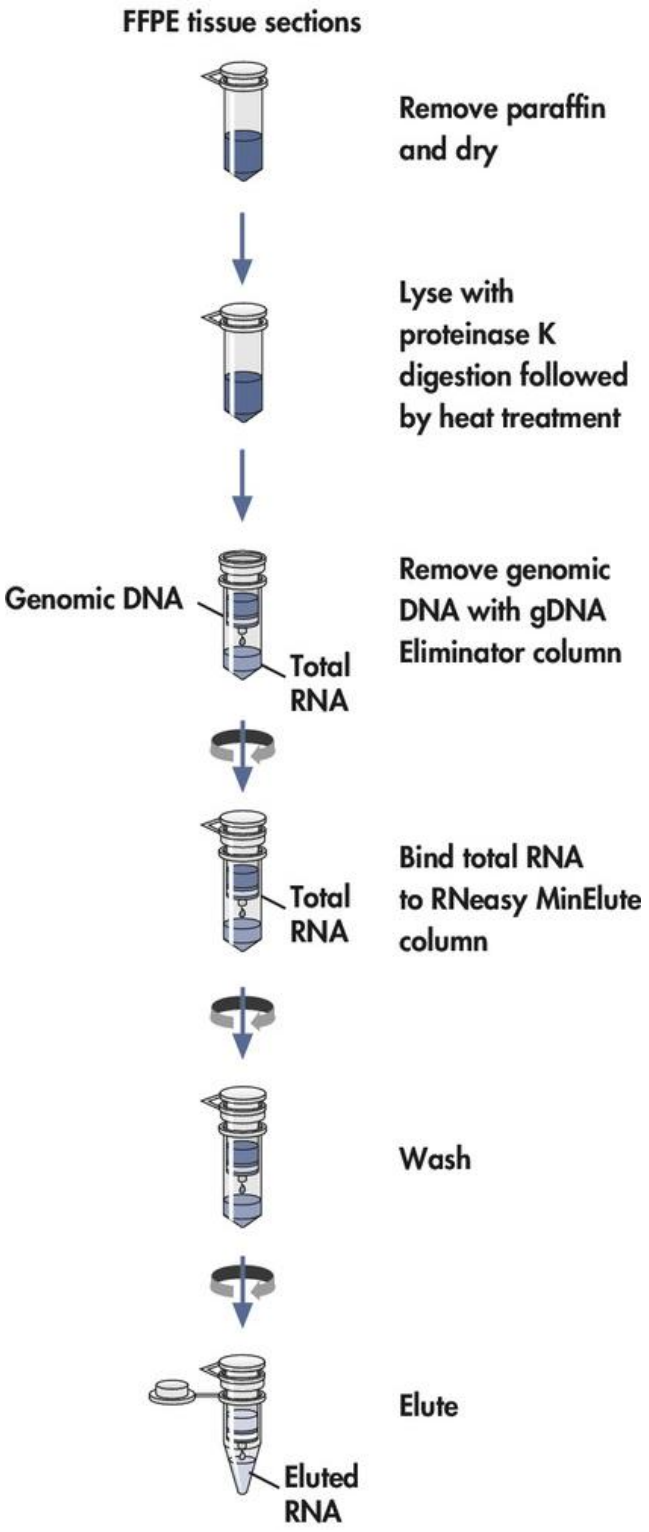


Figure 5.3: Isolation of total RNA from FFPE samples (taken from Qiagen handbook)

5.5.4 DNA quantification using Qubit fluorimeter:

DNA concentrations were assessed using the Qubit DNA broad range assay kit (Life Technologies). For those samples with DNA concentrations below detectable levels using the broad range kit, the Qubit DNA High Sensitivity Assay kit (Life Technologies) was used instead according to the manufacturer's instructions.

5.5.5 RNA quantification using Qubit fluorimeter:

Similarly, RNA concentrations were assessed using the Qubit RNA broad range assay kit (Life Technologies). This broad range kit was able to quantify RNA concentrations between 20 – 1000ng/µL. For those samples with RNA concentrations below detectable levels using the broad range kit, the Qubit RNA High Sensitivity Assay kit (Life Technologies) was used instead according to the manufacturer's instructions.

5.5.6 Polymerase chain reaction of HPV 16 E2, E6, E7 genes and URR:

DNA was analysed for viral E2 gene integrity using PCR and the results were repeated independently in duplicates. The assay utilised overlapping primers to the E2 gene with the forward and reverse nucleotide sequences shown in Table 5.1.

Table 5.1: PCR primers for HPV 16 E2, E6, E7, and L1 genes, the URR regions as well as host GAPDH

Target	Nucleotide sequence	Melting temp (T_m)
HPV 16 E2 Set 1	5'-AGGACGTGGTCCAGATTAAG -3' 5'-TCAAACTGCACTTCCACTGT -3'	54^{o}C
HPV 16 E2 Set 2	5'-TAACTGCACCAACAGGATGT -3' 5'-GCCAAGTGCTGCCTAATAAT -3'	54.3^{o}C
HPV 16 E2 Set 3	5'-ATCTGTGTTTAGCAGCAACG -3' 5'-TAAATGCAGTGAGGATTGGA -3'	54^{o}C
HPV 16 E2 Set 4	5'-ACAGTGCTCCAATCCTCACT -3' 5'-TCACGTTGCCATTCACTATC -3'	54^{o}C
HPV 16 E2 Set 5	5'- GGCATTGGACAGGACATAAT -3' 5'-CAAAAGCACACAAAGCAAAG -3'	54^{o}C
HPV 16 E6	5'-GAACAGCAATACAACAAACC -3' 5'-GATCTGCAACAAGACATACA -3'	55^{o}C
GAPDH	5'-GCTCAAGGGAGATAAAATTC-3' 5'-CGACCAAATCTAAGAGACAA-3'	55^{o}C
HPV 16 E7	5'-ATATATGTTAGATTTGCAACCAGAGACAAC-3' 5'-GTCTACGTGTGTGCTTTGTACGCAC-3'	55^{o}C
HPV 16 L1	5'-TTGTTGGGGTAACCAACTATTTGTTACTGTT-3' 5'-CCTCCCCATGTCTGAGGTACTCCTTAAAG-3'	55^{o}C
HPV 16 URR - LCR	5'-ACACCTACTAATTGTGTTGTGGTTA-3' 5'-GCCAAAAAGCATGCAACCGA-3'	65^{o}C
HPV 16 URR - Promoter	5'-AAAACTAAGGGCGTAACCGA-3' 5'-TAACTTTCTGGGTCGCTCCTG-3'	65^{o}C

These 5 sets of overlapping primers span the entire length of the HPV 16 E2 gene and have previously been validated in a study on viral E2 gene integrity in cervical cancer specimens (Woodman et al., 2013). The melting temperature for these primers are shown in Table 5.1. The overlapping nucleotide positions are as shown in Figure 5.4.

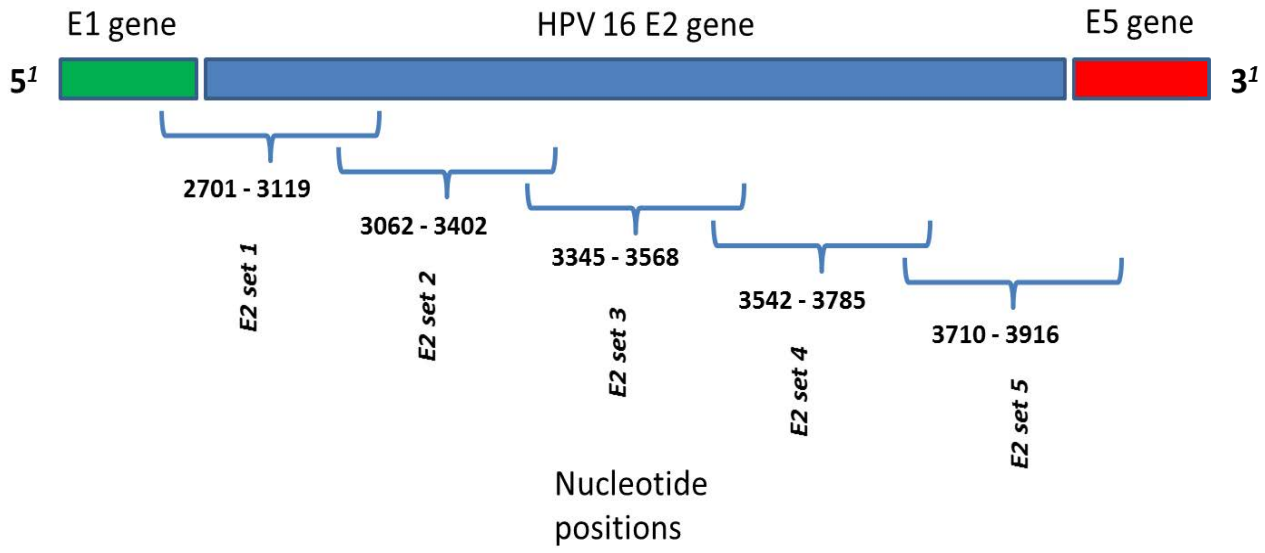


Figure 5.4: Primers for analysing HPV 16 E2 gene integrity with overlapping nucleotides

The HPV 16 E6, E7 and L1 genes were used as positive controls for viral DNA presence while GAPDH (Glyceraldehyde 3-phosphate dehydrogenase) is the house-keeping gene used for data normalization. Any samples not expressing GAPDH were excluded from further analysis. The respective primer sequences and melting temperatures are shown in Table 5.1. DNA from two cell lines, Archie and SiHA was used as controls. Archie was obtained from HPV positive keratinocytes and contains intact E2 gene while SiHA contained disruption of the E2 gene detectable by the E2 primer sets 2 and 3. The PCR process is described below for E2, E6, E7 on all 40 samples using the respective primers shown in Table 5.1. URR PCR was only performed on 10 samples where it was not possible to establish the methylation profiles at the E2 binding sites following pyrosequencing.

Each reaction contained 20 – 25ng of DNA in 5 µL from each patient sample. However, it was not possible to achieve this concentration in 16 samples with DNA concentrations less

than 5 ng/µl (Table 5.5). In these cases, 5µL of undiluted DNA was added to the reaction. Each reaction contained 10µL of PCR MasterMix (Promega), 5µL of DNA sample (or nuclease-free water for negative control), 3.5 µL of nuclease-free water and 1.5µL of each forward and reverse primer separately. The 96 well plate containing the primers and nuclease-free water was prepared in a separate laboratory in a designated hood that had been sterilised prior by meticulous cleaning with alcohol and UV light for 20 minutes. This was to minimise the risk of contamination from DNA fragments. The plate was then transferred to the main laboratory bench where the DNA samples were added, the plates were sealed and labelled appropriately. The PCR with the overlapping E2 primers, E6, E7, L1 primers were run as follows: 95°C for 10 minutes, followed by 50 cycles of 95°C for 20 seconds, 54°C for 40seconds, and 72°C for 50 seconds. On completion of PCR, the PCR products were analysed using agarose gel electrophoresis for the presence of bands indicative of the presence of the amplified HPV segment.

To investigate the integrity of the HPV 16 URR in those samples that failed amplification prior to pyrosequencing, PCR reactions were set up in duplicate using 10µL of PCR mastermix, 3.5µL of water, 1.5µL of primers and 5µL of DNA. PCR was performed running denaturation step at 95°C for 15 minutes, followed by 50 cycles at 95°C for 20 seconds, 54°C for 40 seconds, 72°C for 50 seconds. Final step of 72°C for 1 minute and 4°C holding temperature. The PCR reactions included a positive control (CaSki), a negative control (DNase – free sterile water), a house keeping gene (GAPDH) and a viral positive control (HPV 16 E6 gene). The reactions were separated for both the Long Control Region (LCR) and the Promoter region of the viral HPV 16 URR.

5.5.7 1.5% agarose gel electrophoresis:

Agarose (2.25g) was weighed and mixed with 150mL of 1X TBE (Tris-Borate-EDTA). The 1X TBE was made by diluting 100mL of 10X TBE with 900mL of sterile water. The agarose and TBE mixture was then heated in a microwave for 1 minute and intermittently checked for complete dissolution of the agarose. The mixture was allowed to cool and then 5µL of ethidium bromide was carefully added to the 150mL of agarose gel. The required number of combs were inserted and the gel was allowed to set over a 30 – 40 minutes. The gel was subsequently placed in an electrophoresis tank and the DNA ladder and each PCR product was sequentially added using a micropipette. The DNA ladder (100 bp) was prepared by diluting 1µg of DNA ladder (concentration 1µg/µL) with 9µL of nuclease-free water. 2.5µL of orange / blue loading dye (5x) was then added to the ladder mix. 5µL of the 5x loading dye was added to each 20µL PCR product. The electrophoresis tank was filled with 1X TBE and the electrophoresis ran at 150 volts for 25 minutes (ensuring a cephalo-caudal direction of electricity i.e. black electrode at top and red at bottom). On completion of the electrophoresis run, the gel was gently retrieved from the tank and immediately transferred to the UV light box for assessment of the presence / absence of bands in each well. The presence of a band in each OPC sample indicates the presence of episomal HPV (although does not exclude the presence of integrated forms of the virus elsewhere in the HPV genome) while the absence of a band indicates disruption of the gene (A representative sample is shown in Figure 5.5). E2 gene disruption was recorded if there was an absence of a band in any of the wells containing the 5 overlapping primers to the HPV 16 E2 gene. All PCR reactions to assess HPV 16 E2 gene integrity in patient samples were performed in duplicates.

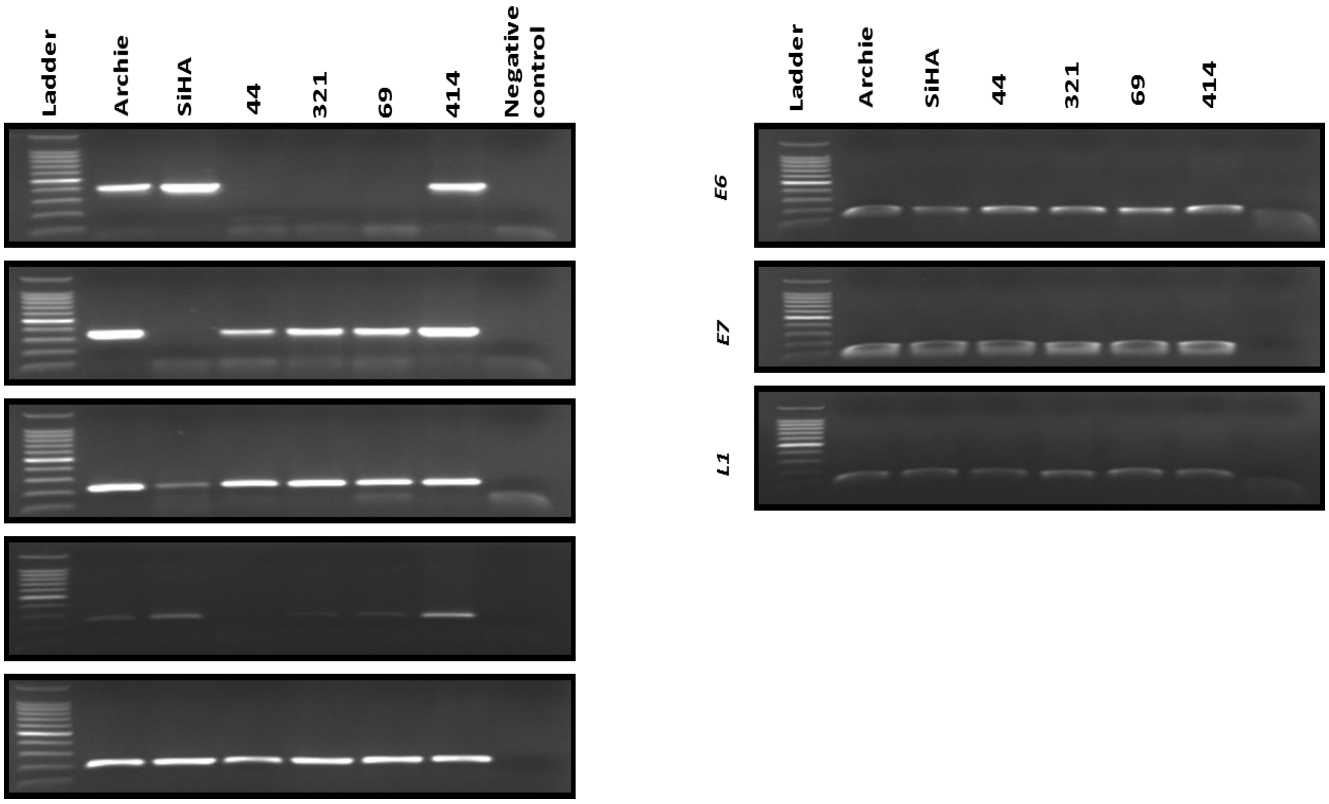


Figure 5.5: HPV 16 E2 gene integrity assessed by PCR using overlapping primers to E2

Figure shows a representative PCR with overlapping primers to the E2 gene as well as positive controls E6, E7 and L1 genes. The absence of bands in any of the E2 primer sets indicates a disrupted E2 gene and this is shown by primer sets 1 and 4 in patient samples 44, 321, 69 adjacent to a patient sample with an intact E2 gene (414). Control cell line Archie has intact E2 while SiHa has disrupted E2 in primer set 2.

5.5.8 Bisulfite-modification of DNA samples:

To evaluate the methylation profiles at the E2 binding sites in the HPV 16 URR, the extracted DNA from patients' samples had to undergo bisulfite modification prior to pyrosequencing amplification. In a result of bisulfite modification, the methylated cytosines remain intact while the unmethylated cytosines are converted into uracil. Subsequently, the uracil is then detected as thymine following PCR amplification.

In this study, bisulfite modification of DNA was performed using the EZ DNA Methylation-Gold™ kit (Zymo research) and manufacturer's protocol. The initial step was to prepare the Cytosine – Thymine (CT) conversion reagent by mixing 900 µL of nuclease-free water, 300 µL of M-Dilution buffer, 50 µL of Dissolving buffer and one tube of CT conversion reagent. This mixture was vortexed for 10 minutes. 130 µL of the prepared CT conversion mixture was then added to 20 µL of the extracted DNA sample and mixed. This was placed in a Thermal cycler and incubated at 98°C for 10 minutes, 64°C for 2.5 hours and then 4°C hold. Each sample was added to 600 µL of M-Binding buffer in a Zymo-Spin™ IC column and mixed. Each sample was centrifuged at full speed (>10,000 x g) for 30 seconds and the flow-through discarded. 100 µL of the M-Wash Buffer was added to the column and spun for 30 seconds. Subsequently, 200 µL of M-Desulphonation buffer was added to the column and left for 15 – 20 minutes incubation on the benchside in room temperature. After incubation, samples were spun at full speed for 30 seconds. This step was then followed by addition of 200 µL of M-Wash Buffer to the column and the solution was spun for 30 seconds. This particular wash step was repeated, the column changed to a new, dry one and the sample spun at full speed for 1 minute to dry. 10 µL of M-Elution Buffer was then added to the column matrix without making contact with the matrix in the column. The entire column was placed in a 1.5mL eppendorf tube and spun for 1 minute to elute the now bisulfite-modified DNA sample. Depending on the downstream applications, the bisulfite-modifided DNA can then be diluted and in this study, it was diluted to make up 20 µL of final sample needed for pyrosequencing. The samples were then labelled and stored in – 20°C.

5.5.9 PCR amplification of bisulfite-modified DNA:

Following bisulfite-modification of the DNA extracted from oropharyngeal cancer, PCR amplification and agarose gel electrophoresis were performed prior to any pyrosequencing. The primers used cover the location of the CpG sites in the E2 binding sites at the HPV 16 URR and are shown in Table 5.2.

The HPV 16 LCR primer covers E2BS1 (CpG positions 7452 and 7458) while the HPV 16 promoter primer covers E2BS3 (CpG positions 37, 43) and E2BS4 (CpG positions 52, 58). The 96 well plates were prepared in the dedicated hood in a separate laboratory to avoid contamination with PCR fragments. Each well contained 25 µL of PCR Mastermix (Promega), 18 µL of nuclease-free water, and 2 µL of each reverse and forward primers. 5 µL of nuclease-free water was added into the wells for negative controls. The plate was taken to the main laboratory where 5 µL of the bisulfite-modified DNA from each patient cancer sample was added. 5 µL of each control cell line DNA (SiHa and CaSki) were also added to their separate wells, the plate sealed and placed in the thermal cycler. PCR amplification was performed at the following settings: 95°C for 10 minutes, followed by 60 cycles of 95°C for 30 seconds, 50°C for 45 seconds, 72°C for 1 minute. After the 60 cycles were completed, the reaction ran at 72°C for 10 minutes and then 4°C holding temperature.

5.5.10 Agarose gel electrophoresis of PCR products before pyrosequencing:

On completion of PCR amplification, the 96 well plate was placed on a separate laboratory bench reserved for PCR products in preparation for 1.5% agarose gel electrophoresis. The agarose gel and ladder were prepared as previously described in section 5.13. 3µL of loading dye (6x) was added to 8 µL of each PCR product including controls (SiHa and CaSki). In

cases where PCR bands were either faint or absent following electrophoresis, those samples were not expected to amplify satisfactorily on subsequent pyrosequencing.

5.5.11 Pyrosequencing of bisulfite-modified DNA samples:

The pyrosequencing was performed using the Pyro Q-CpG^{TM} system (Qiagen, Uppsala, Sweden) and PyroMark reagents (Qiagen). A single-stranded template of HPV DNA specifically the HPV 16 upstream regulatory region containing the E2 binding sites, is hybridized to sequencing primers for the LCR and promoter regions separately. The pyrosequencing primers were optimised in duplicates using VU147 as the reference sample and normalized to β-actin, with clear water controls prior to any analysis of patient samples. Unmethylated cytosine is converted to uracil on bisulfite modification. Uracil is then converted on PCR amplification to thymine which is recognised during pyrosequencing.

5.5.11.1 Pyrosequencing methodology:

Pyrosequencing is based on 'DNA sequencing by synthesis'. It was developed by Mostafa Ronaghi and Pal Nyren in Sweden in 1996 at the Royal Institute of Technology. During DNA synthesis, only one nucleotide at a time can be added to the complementary single-stranded DNA. Pyrophosphate is emitted on nucleotide incorporation and this is detectable as a light signal on a pyrogram. Solutions of each nucleotide are added one after another and each is cleared from the reaction by enzymes before the next nucleotide is added.

The enzymes necessary for the reaction include DNA polymerase, ATP sulfurylase, luciferase and apyrase. The substrates required are adenosine 5'phosphosulfate (APS) and luciferin.

DNA polymerase attaches the complementary nucleotide to the HPV DNA template of the E2BS, with pyrophosphate release. Whilst utilising APS, the ATP sulfurylase converts the

released pyrophosphate to ATP. Luciferase requires ATP to convert luciferin to oxyluciferin. This process generates light that is captured on the PyroMark camera with the resulting analysis shown on a pyrogram. Any remaining nucleotides and ATP not used up in the preceeding reaction becomes degraded by the enzyme apyrase in preparation for the next reaction. The ratio of unmethylated cytosine (now thymine) to methylated cytosine (now cytosine) at each CpG site is recorded as the mean methylation at that CpG site (Figure 5.6 shows an example of the entire process with E2BS1 as an example). A cytosine that is not immediately followed by guanine on the HPV DNA sequence for amplification acts as the control for the efficiency of bisulfite modification.

5.5.11.2 Pre-pyrosequencing steps:

Prior to pyrosequencing a number of preparatory steps were undertaken. Firstly, the PyroMark reagents (Qiagen) including Binding buffer, Annealing buffer and Streptavidin were all retrieved from the refrigerator and left in room temperature for 15 – 20 minutes. The primers required were retrieved which include the LCR primer (CpGs 7452, 7458), promoter primer (CpGs 37, 43), the sequencing primer for the LCR and the sequencing primer for the promoter end (Table 5.2). Also required was nuclease-free water, a 96 well pyrosequencing plate and the actual PCR plate containing the PCR products of the bisulfite-modified DNA. The PCR plate contained 40µL of PCR products in each well. The heat block was set to 80˚C and allowed to reach this temperature. The pyrosequencing cartridges were cleaned by rinsing with water to ensure each needle channel was open and unblocked. This step is crucial as the entire experiment could fail if reagents are not dispensed due to cartridge blockage. Tap water is pressed through each channel with a finger without touching the needle ends. The pyrosequencing troughs were rinsed out with tap water as required (especially the denaturation buffer trough). Negative controls for the pyrosequencing were prepared with

38µL of nuclease-free water and 2µL of LCR and promoter primers (in separate wells) as well as 39 µL of nuclease-free water and 1µL of sequencing primers to both the LCR and promoter regions.

5.5.11.3 Preparation of plates:

A mastermix was produced containing 37 µL of Binding buffer and 3µL of Streptavidin beads for each sample. This makes up the 40µL of the mastermix pipetted into each well on the 96 well PCR plate already containing 40µL of PCR products from bisulfite-modified DNA. This plate was left to stand in room temperature on a bench for 10 – 15 minutes. After this incubation period, the plate contents were mixed using a multi-channel pipette or put on a plate shaker for 5 minutes at 1,300rpm. The pyrosequencing 96 well plate was then prepared by pipetting 38.5µL of annealing buffer into each well. 1.5µL of sequencing primers (LCR and promoter) were then added separately to each well containing the annealing buffer.

Table 5.2: Pyrosequencing primers for methylation study of HPV 16 E2 Binding sites

Target	Binding sites	CpG sites	Amplification primers (5` → 3')	Sequencing primers (5' – 3')
HPV 16 'LCR primer'	E2BS1	7452, 7458	FW: 5'- ATTGTGTTGTGGTTATTTATTGTA -3' RV: 5'-AAAAAACACATTTTATACCAAAAA -3' Annealing temp – 57°C Amplicon length 146bp	**GGTTTATTTTG TAGTTTTAA**
HPV 16 'Promoter primer'	E2BS3 E2BS4	37, 43 52, 58	FW: 5'-ATTTATGTATAAAATTAAGGGTGT -3' RV: 5'- TCACTCCTATAAATCCTAAAACAT -3' Annealing temp – 56°C Amplicon length 118bp	**ATGTATAAAA TTAAGGGTGT**

5.5.11.4 Pyrosequencing:

This process starts with a series of sequential wash steps in various solutions using a hedgehog (a hand-held device with 96 metallic tubes sealed with a permeable seive and connected to a suction machine). There are individual troughs filled separately with 180mL of 70% ethanol, 120mL of denaturation buffer (sodium hydroxide 8g), 180mL of Wash Buffer (1:10 dilution), 180mL of sterile water, and 180mL of sterile water in the parking station for the hedgehog.

Once the plate and samples were ready, the vacuum was switched on and the hedgehog placed in the sterile water in the parking station. The samples were then carefully suctioned using the hedgehog ensuring that the residue of beads were attached to each end of the hedgehog. With the beads attached, the hedgehog was placed in the ethanol trough, allowed to suction for 5 seconds, lifted out and this process repeated sequentially through the Denaturation Solution and Wash Buffer. The vacuum was then turned off after the beads have dried. This step is critical as omitting to turn the vacuum off will allow suctioning away of the pyrosequencing primers!. The whitish beads that formed on the ends of the hedgehog were placed into the pyrosequencing plate containing the annealing buffer and sequencing primers and gently dislodged into solution. The hedgehog was replaced into sterile water and the plate taken to the main laboratory and placed on the heat block. The plate was heated at 80°C on the heat block for 2 minutes to ensure complete denaturation of beads-bound DNA.

The pyrosequencing reagents and nucleotides were then pipetted into each well on the pyrosequencing plate which was then transferred into the PyroMark Instrument. The volume information required for each run is indicated on the instrument for example in a typical run, Reagents: E-mix (Enzyme) 137 µL, S-mix (Substrate) 137 µL; Nucleotides: adenosine 62 µL,

cytosine 62 µL, guanine 67 µL, thymine 71 µL. The nucleotides were pipetted into the pyrosequencing cartridge and placed into the PyroMark Instrument.

The 96 well pyrosequencing plate was labelled on the Instrument databoard appropriately for each sample indicating which sequences were to be analysed. Upon completion of the pyrosequencing reaction, the results were analysed by the Pyro Q-CpG^{TM} software (Qiagen). It calculated the proportion of methylated to unmethylated cytosine (after bisulfite-conversion) at each HPV 16 E2 binding site. Mean methylation at each E2BS was calculated from the individual readings for every separate sample. Pyrosequencing of OPC patient samples were performed in triplicates.

Pyrosequencing was chosen to carry out this project as it was the methylation determination technique available in our institution and had been previously validated on a cohort of cervical and vulval cancer FFPE specimens with optimised assays and algorithms for bisulfite modified DNA.

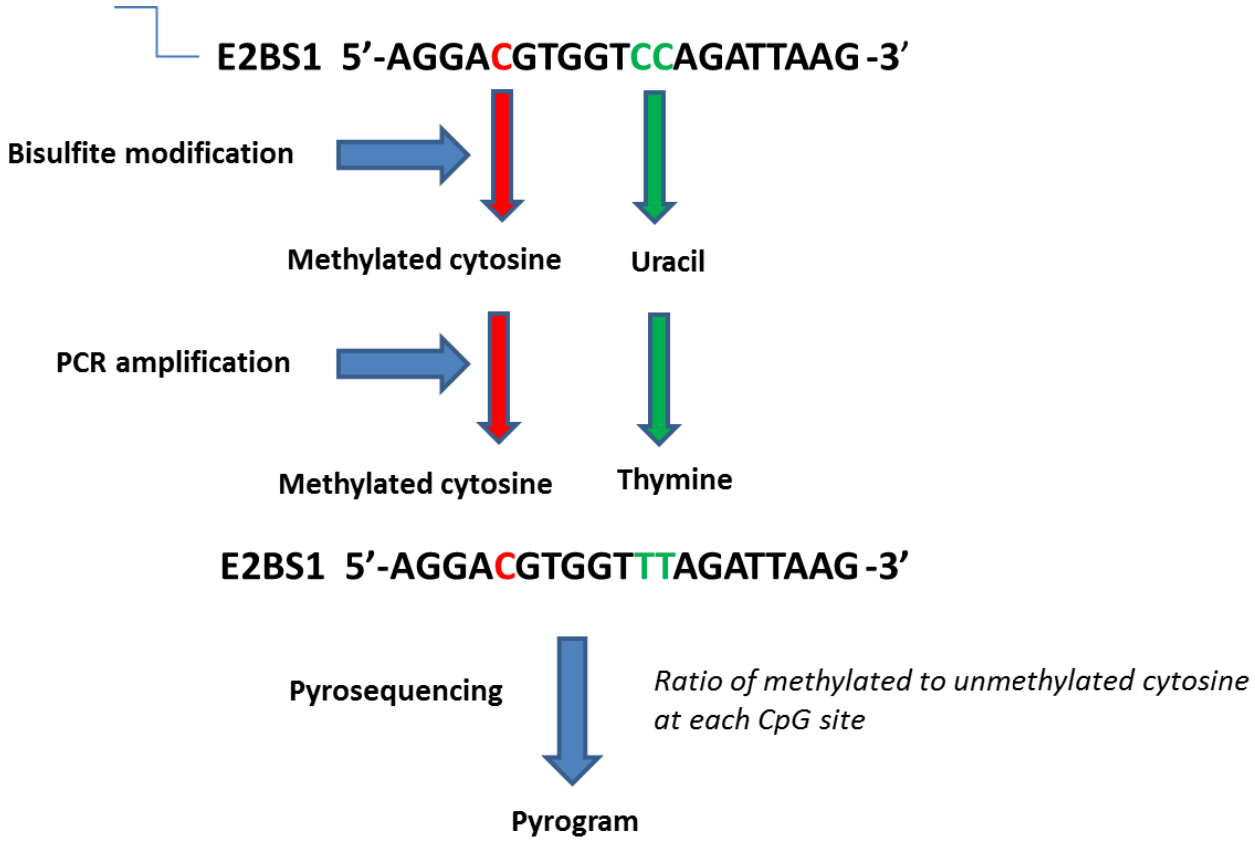


Figure 5.6: Pyrosequencing methodology. *Bisulfite modification converts the cytosines in the E2 binding sites of the HPV16 Upstream Regulatory Region into uracil which is then converted to thymine on PCR amplification. The cytosine / thymine ratio is then determined by the pyrosequencer software.*

5.5.12 cDNA synthesis:

RNA samples were obtained from FFPE of 40 oropharyngeal cancers as described above (section 5.9). RNA concentrations were measured using a Qubit fluorimeter in ng/µL and dilutions were made with nuclease-free water if necessary. 1µg of RNA was used for c DNA synthesis. The positive control used was RNA from VU147 HNSCC cell line which is HPV positive.

Reverse transcription was performed using the 'High-capacity cDNA reverse transcription kit' by Applied Biosystems. Each 20µL reaction consisted of 1µg of each RNA sample diluted with nuclease-free water to a volume of 10µL, 2µL of 10x RT buffer, 2µL of 10x Random primers, 0.8 µL of 25x dNTPs, 1µL of reverse transcriptase and 4.2µL of nuclease-free water. cDNA reaction was carried out in 0.2µL tubes using a thermocycler and the following settings: 25˚C for 10 minutes, followed by 37˚C for 120 minutes and 85˚C for 5 minutes and held at 4˚C.

5.5.13 E6 and E7 mRNA expression in OPC samples using real time quantitative PCR (qPCR):

Each cDNA sample was prepared for qPCR as follows: 2 µL of cDNA (1:3 dilution) was added to 2 µL of E6 or E7 probe (1:20 dilution), 3 µL of E6 or E7 primer, 3 µL of nuclease-free water, and 10 µL of Luminaris low ROX mastermix (Thermoscientific UK). The reaction was normalised to β-actin and this control was similarly prepared as follows: 1 µL of cDNA (1:3 dilution), 1 µL of β-actin primers, 8 µL of nuclease-free water and 10 µL of Luminaris low ROX mastermix (Thermoscientific UK). The qPCR primers and probes are shown in Table 5.3.

Table 5.3: qPCR primers and probes for E6 and E7 RNA expression

Primers and probes	Sequences ($5^1 – 3^1$)
HPV 16 E6 (nt 99 – 178)	Forward primer: CTGCAATGTTTCAGGACCCA Reverse primer: TCATGTATAGTTGTTTGCAGCTCTGT Probe: FAM-AGGAGCGACCCGGAAAGTTACCACAGTT-BHQ
HPV 16 E7 (nt 739 – 816)	Forward primer: AAGTGTGACTCTACGCTTCGGTT Reverse primer: GCCCATTAACAGGTCTTCCAAA Probe: FAM-TGCGTACAAAGCACACACGTAGACATTCGTA-BHQ

qPCR was performed on the samples using the '7500 Real time PCR system' by Applied Biosystems. The 'Quantitative comparative' set up was chosen for a reaction volume of 20 µL. Taqman reagents and a standard run were indicated on the set up screen. The targets recorded were β-actin, E6, E7 with the quencher for each indicated as NFQ-MGB, None, and None respectively. The following were recorded: negative control (nuclease-free water), reference sample (VU147: cDNA obtained from a HPV positive cell line), endogenous control (β-actin), passive reference dye (ROX), reaction volume (20 µL).

The qPCR conditions were: 50°C for 2 minutes in step 1, 95°C for 10 minutes in step 2, followed by 50 cycles of 95°C for 15 seconds in step 1 and then 55°C for 1 minute in step 2. The assay was first fully optimised using a control with known expression of E6 and E7 (VU147). This HPV positive control cell line also expressed β-actin. Any patient samples that did not express β-actin on duplicate qPCR were excluded from the analysis. The specificity of

the reactions was monitored by melting curves. The E6 and E7 qPCR on patient samples were performed in triplicates and the results recorded as normalised to β-actin.

5.5.14 E2 mRNA expression in OPC samples using real time quantitative PCR:

Expression of E2 m-RNA and E2 isoforms ('E2 880^2709' and 'E2 880^2582') was performed with a similar qPCR set up as the E6 and E7 assays described above. Primers for all isoforms of E2 investigated were prepared and their respective nucleotide sequences shown in Table 5.4. In each tube, 2µL of each E2 primer was separately added to 2µL of cDNA (1: 10 dilution), 6µL of nuclease-free water and 10µL of SYBR green Mastermix. Separately, 1µL of β-actin primers was mixed with 7µL of nuclease-free water.

Table 5.4: qPCR primers for HPV 16 E2 gene isoforms

	Primer Names	Nucleotide positions	Sequences
E2	AM771 (sense)	3302 – 3321	GTATTGGGAAGTTCATGCGGG
	AM772 (antisense)	3381 – 3359	TCAGGAGAGGATACTTCGTTGCT
E2 880^2709	AM1126 (sense)	866 – 880,	TGGCTGATCCTGCAGGACG
	AM1127 (antisense)	2778 – 2755	AAACGTTGGCAAAGAGTCTCCAT
E2 880^2582	AM1128 (sense)	866 – 880	TGGCTGATCCTGCAGATTC
	AM1129 (antisense)	2706 – 2680	TGAGAAAAAGGATTTCCAGTTCTTATC

Real-time PCR was performed using the '7500 Real time PCR system' by Applied Biosystems. The 'Quantitative comparative' set up was chosen for a reaction volume of 20 µL

in √7500 (96 wells). SYBR reagent and a standard run were indicated on the set up screen. The targets recorded were E2, 'E2 880^2709', 'E2 880^2582', and β-actin. All four targets had the reporter labelled as SYBR and the quencher as 'None'.

The qPCR settings were as follows: 95°C for 10 minutes, followed by 50 cycles of 95°C for 10 seconds (denaturation) and 60°C for 60 seconds (annealing and extension). After these stages, the melting curve was set at 95°C for 15 seconds, 60°C for 60 seconds, 95°C for 30 seconds and a final 60°C for 15 seconds. These experiments were optimised adequately for the assays using the positive control VU147 in duplicates. Any patient samples that did not express β-actin in two separate runs were subsequently excluded from the analysis. The E2 qPCR on patient samples were performed in triplicates and shown as normalised to β-actin. Expression of any of the three E2 isoforms investigated was regarded as satisfactory for E2 expression in that sample. Each isoform was however analysed separately and correlated with clinical outcomes.

5.6 Results:

5.6.1 DNA and RNA extraction from FFPE samples:

40 Formalin-Fixed Paraffin Embedded (FFPE) blocks of HPV-positive OPCs randomly selected were available for this study from the multicentre UK PET-NECK trial. DNA and RNA were extracted from the FFPE blocks using the standard protocols described in the Methods section. Concentrations for each DNA and RNA extracted sample were determined using the Fluorimeter and shown in Table 5.5.

DNA concentrations (ng/μL) obtained from the Qubit fluorimeter for the 40 HPV positive oropharyngeal cancer samples ranged between (0.01 – 54.9) ng/μL (Table 5.5). Similarly, the RNA concentrations for the 39 available samples ranged widely between (10.0 – 1,842.0) ng/μL. There was no tissue left on the FFPE block for one cancer sample (OPC 70) to allow further RNA extraction.

Table 5.5: DNA and RNA concentrations of 40 oropharyngeal cancer FFPE samples

Number	Tumour ID	DNA concentration (ng/µl)	RNA concentration (ng/µl)
1	48	54.90	600.00
2	414	44.00	*525.00**
3	400	38.00	450.00
4	321	32.10	317.00
5	141	28.10	902.00
6	22	26.00	1842.00
7	490	25.10	1200.00
8	418	18.20	900.00
9	57	17.40	1120.00
10	449	15.60	236.00
11	346	15.40	172.00
12	525	15.20	1040.00
13	69	14.20	347.00
14	450	13.70	1000.00
15	519	12.50	892.00
16	462	10.60	580.00
17	142	10.20	*530.00**
18	145	9.83	479.00
19	451	9.18	748.00
20	320	8.59	566.00
21	42	8.56	698.00
22	19	6.42	456.00
23	537	6.22	526.00
24	14	5.98	*590.00**
25	491	4.32	654.00
26	408	3.83	10.00
27	311	2.84	176.00
28	44	2.53	*170.00**
29	55	2.27	*170.00**
30	201	2.14	165.00
31	224	1.80	700.00
32	372	1.08	314.00
33	185	0.84	119.00
34	39	0.78	269.00
35	70	0.61	No sample
36	423	0.01	1200.00
37	488	0.01	388.00
38	214	0.01	412.00
39	350	0.01	172.00
40	330	0.01	498.00

**indicates concentrations obtained using the Qubit DNA and RNA high-sensitivity kit*

5.6.2 HPV 16 E2 gene disruption on type-specific PCR:

In the type-specific PCR using overlapping primers to the HPV 16 E2 open reading frame (ORF), the absence of a band on 1.5% agarose gel electrophoresis for one or more of the overlapping primers (E2 sets 1 – 5) was interpreted as 'E2 gene disruption'. All the samples had detectable bands at E6, E7 and L1 indicating the samples indeed contained HPV 16 viral DNA, and were of reasonable quality.

40% (16/40) of the HPV positive oropharyngeal cancer samples had HPV 16 E2 gene disruption shown by PCR (Table 5.6). HPV 16 E2 gene disruption was found with the overlapping primers in differing proportions. The most frequent E2 disruption was within 'E2 set 1 primer' as it identified 93.8% (15/16), in contrast to the least frequent location within primer (E2 Set 5) which was intact in all samples and did not pick up any of the 16 samples with E2 disruption (Table 5.6). It is worth mentioning that the nucleotide positions for the E2 set 5 primer overlap part of the E5 gene (Figure 5.7) and this may explain why this primer set was the least frequent location for E2 gene disruption. A representative sample with E2 gene disruption is shown in Figure 5.7.

OPC 346 showed absence of a band with E2 set 1 primer in two separate PCR reactions. This was classed as 'E2 gene disrupted' as one of the five overlapping E2 primers had absence of a band demonstrated on duplicate PCR. OPC 346 also showed intact positive controls with E6, E7 and L1 bands present. The cell line controls were also consistent with the HPV positive keratinocyte cell line (Archie) showing intact E2 gene and the presence of bands in all overlapping E2 primer sets. As expected the HPV positive cervical cancer cell line (SiHa) showed E2 gene disruption in the E2 primer set 2 and a faint band in E2 primer set 3. SiHa has 2 copies of integrated HPV and was used as a positive control to detect E2 gene

disruption. The negative control (nuclease-free water) was free of contamination and did not amplify any bands as expected.

The remaining 60% (24/40) of the HPV OPC samples had intact HPV 16 E2 gene. An example is shown in Figure 5.7 in the PCR done 22/2/16, where OPC 224 and OPC 70 both demonstrated the presence of bands in all the E2 primer sets as well as E6, E7 and L1. Both these patient samples were therefore classified as having an 'intact E2 gene'. The limitations of this assay have already been highlighted in the methods section. In particular, the presence of bands on electrophoresis over the span of the E2 gene did not exclude the presence of viral integrations elsewhere in the HPV 16 genome. In addition, the presence of an intact band does not rule out the presence of both episomal and integrated forms (mixed).

Samples that did not amplify the housekeeping gene (GAPDH) on duplicate PCR, such as OPC 81 were excluded from further analysis (see Figure 5.7). In every case of discordance in establishing E2 gene integrity following duplicate PCR i.e. a band was present on agarose gel electrophoresis but absent on the second gel, a third confirmatory PCR was performed and the sample classified as either E2 intact or E2 disrupted once there were two concordant PCR results.

Table 5.6: HPV 16 E2 gene disruption pattern at each primer site in 16/40 samples tested (all other samples had intact E2 gene)

Number	OPC ID	E2 set 1	E2 set 2	E2 set 3	E2 set 4	E2 set 5
1	519	x	x	x	x	present
2	462	x	x	x	x	present
3	423	x	x	x	x	present
4	350	x	x	x	x	present
5	330	x	x	x	x	present
6	490	x	x	present	x	present
7	449	x	x	present	x	present
8	488	x	x	present	x	present
9	22	x	x	present	x	present
10	14	x	x	present	x	present
11	408	x	present	present	x	present
12	346	x	present	present	present	present
13	44	x	present	present	x	present
14	321	x	present	present	present	present
15	69	x	present	present	present	present
16	214	present	present	present	x	present

x denotes disruption of the E2 gene at this primer set with absence of a band on 1.5% agarose gel electrophoresis

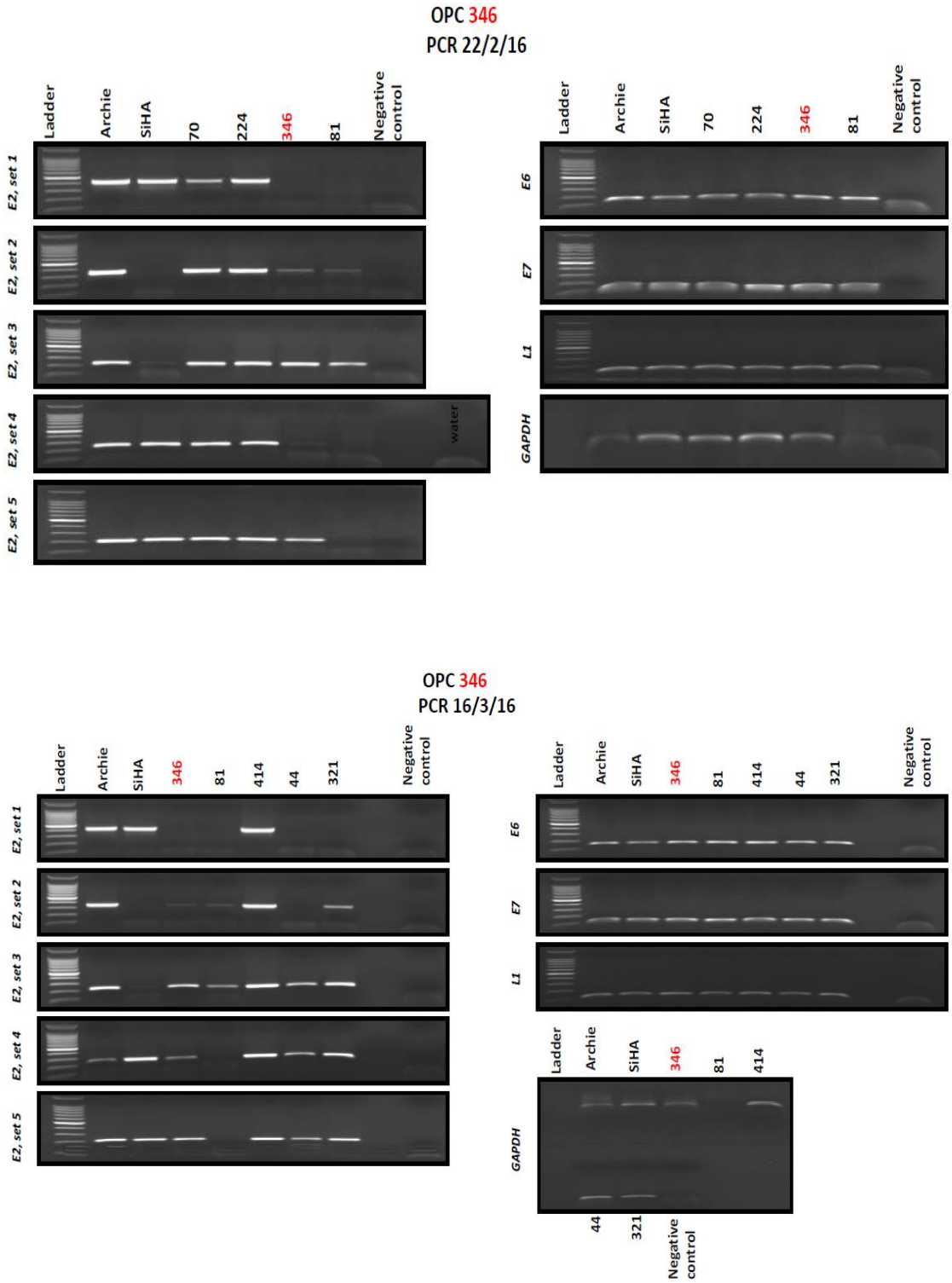


Figure 5.7: HPV 16 E2 gene disruption in patient sample 346 on duplicate PCRs. *A representative sample demonstrating E2 gene disruption by PCR and 1.5% agarose gel electrophoresis in patient sample 346 (in red) by duplicate PCRs done on separate dates (22/2/16 and 16/3/16). In both PCRs, there is absence of a band in E2 primer set 1.*

5.6.3 HPV 16 E2 gene status and clinical correlation:

The status of the HPV 16 E2 gene in each oropharyngeal cancer sample (intact or disrupted), as established by PCRs, was correlated with demographic and clinical data to investigate any differences between the two groups. These clinical correlates include anatomic subsite, age, gender, tumour size, smoking, alcohol consumption, overall survival and cancer recurrence.

Table 5.7: HPV 16 E2 gene integrity correlated with OPC anatomic site and gender

	Total OPC	**E2 gene Disrupted (N)**	***E2 gene Disrupted (%)***	**E2 gene Intact (N)**	***E2 gene Intact (%)***	**Statistical Test**
Anatomic Sub-site						
Total number	40	16	*40*	24	*60*	
Tonsils	25	9	*36*	16	*64*	
Tongue base	10	4	*40*	6	*60*	
Not known	5	3	*60*	2	*40*	
						Chi-square statistic = 0.049; p = 0.825
Gender Distribution						
Males	35	13	*37.1*	22	*62.9*	
Females	5	3	*60*	2	*40*	
						Chi-square statistic = 0.95; p=0.329

5.6.3.1 Anatomic subsite:

The anatomic subsite was available in 35/40 (87.5%) of oropharyngeal cancer in this study. Of the 35 samples with known subsite, the tonsil subsite accounted for 25/35 (71.4%) and the tongue base for 10/35 (28.6%). By subsite, the HPV 16 E2 gene was disrupted in 9/25 (36%) in the tonsil samples and 4/10 (40%) in the tongue base cancer samples. There was therefore no statistically significant difference between the tonsil and tongue base anatomic sites when considering E2 gene disruption (p=0.825) (Table 5.7). The finding of higher proportions of episomal HPV i.e. intact E2 gene, in oropharyngeal cancer was consistent regardless of whether the tumour was tonsil cancer (64%) or tongue base cancer (60%).

5.6.3.2 Gender:

The majority of the oropharyngeal cancer samples were obtained from males with 35/40 (87.5%) in this cohort. Only 5 samples (12.5%) were obtained from females. By HPV E2 gene status, 13/35 males (37.1%) had disruption of the E2 gene compared with 3/5 (60%) disruption in females (Table 5.7). The observed gender difference with females having higher proportions of OPC with HPV E2 gene disruption is however not statistically significant given the small numbers of females in the analysis (p=0.329). This merits further investigation.

Table 5.8: HPV 16 E2 gene integrity correlated with OPC tumour size

	Total OPC	E2 gene Disrupted (N)	*E2 gene Disrupted (%)*	E2 gene Intact (N)	E2 gene Intact (%)	Statistical Test
Tumour size						
T1 (≤ 2cm)	9	3	*33.3*	6	66.7	
T2 (> 2cm - < 4cm)	18	8	*44.4*	10	55.6	
T3 (> 4cm)	6	2	*33.3*	4	66.7	
T4 (Invasion)	7	3	*42.9*	4	57.1	
						p = 0.930

5.6.3.3 Tumour size:

HPV E2 gene status was investigated by oropharyngeal cancer tumour size based on the following divisions: T1 (≤2cm), T2 (>2cm - <4cm), T3 (>4cm), T4 (invasion). The majority of the tumours in this cohort were T2 (18/40), followed by T1 (9/40), T4 (7/40) and T3 (6/40). HPV E2 gene was disrupted in T1 (33.3%), T2 (44.4%), T3 (33.3%) and in T4 (42.9%) (Table 5.8). There was no statistically significant difference between tumour size categories and HPV E2 gene disruption (p=0.930) (Table 5.8).

5.6.3.4 Age:

Mean age for this OPC cohort of 40 patients was 56.98 years. The OPC samples with HPV 16 E2 gene disruption were from older patients with a mean age of 60.63 years when compared with the OPC samples with intact HPV 16 E2 gene that had a patient mean age of 54.54 years. This indicates that on average, patients who had tumours with intact HPV were approximately 6 years younger than those patients found to have disrupted HPV. This was a significant finding using student T-test (p=0.032) (Table 5.9).

Table 5.9: HPV 16 E2 gene integrity stratified by age and gender

		E2 gene disrupted	E2 gene intact	Statistical Test
Age (years)				
Overall	56.98	*60.63*	54.54	T-test: t = 2.223; p = 0.032
Males		*57.62*	54.5	T-test: t = 1.169; p = 0.251
Females		*73.63*	55	T-test: t = 2.592; p = 0.081

On further stratification by gender however, the data showed that although males with disrupted HPV 16 E2 gene were marginally older at a mean age of 57.62 years compared with the intact HPV 16 E2 group at 54.5 years, this effect was not statistically significant on student T-test (p=0.251). In contrast amongst females, there was a greater dichotomy shown on age stratification by E2 status albeit not statistically significant but trending towards significance. Females with OPC that had disrupted HPV 16 E2 gene had a mean age of 73.63 years in contrast to females with OPC that had intact HPV 16 E2 gene who were younger at a mean age of 55 years (Table 5.9). Student T-test showed a trend towards significance (p=0.081).

With increasing mean age, there appears to be a tendency towards increasing HPV 16 E2 gene disruption in OPC, with this age effect more marked amongst females. This observation though must be interpreted within the small groups of patients in this study, especially females.

5.6.3.5 Smoking:

Smoking is an established risk factor in the development of oropharyngeal cancer. Smoking status was defined by the following categories: 'current smoker', 'ex-smoker' and 'non-smoker'. There were 3 patients for whom their smoking status was not known. There were 26/37 (70.3%) of this OPC cohort who gave a history of ever having smoked (12/37 (34.4%) were current smokers, 14/37 (37.8%) were ex-smokers), compared to 11/37 (29.7%) non-smokers (Table 5.10).

In this cohort, smoking did not show a statistically significant impact on viral E2 status in OPC when stratified into current smoker, ex-smoker or non-smoker on X^2 test (p=0.651).

Table 5.10: HPV 16 E2 gene integrity and smoking status (N – 37)

Smoking status	Overall OPC	HPV 16 E2 disrupted	HPV 16 E2 gene intact	X^2 test
Current Smoker	12	5/12 (41.7%)	7/12 (58.3%)	
Ex-Smoker	14	4/14 (28.6%)	10/14 (71.4%)	
Non-Smoker	11	5/11 (45.5%)	6/11 (54.5%)	
Total	37	14/37 (37.8%)	23/37 (62.2%)	X^2 statistic=0.857 P=0.651

Even when analysed as 'Ever smoked' versus 'Non-smoker', smoking did not show a statistically significant difference in terms of E2 gene integrity on X^2 test (p=0.534) as shown in Table 5.11.

Table 5.11: HPV 16 E2 gene integrity and 'ever having smoked' status (N – 37)

Smoking status	Overall OPC	HPV 16 E2 disrupted	HPV 16 E2 gene intact	X^2 test
Ever smoked	26	9/26 (34.6%)	17/26 (65.4%)	
Non-Smoker	11	5/11 (45.5%)	6/11 (54.5%)	
Total	37	14/37 (37.8%)	23/37 (62.2%)	X^2 statistic=0.386 P=0.534

5.6.3.6 HPV 16 E2 gene integrity and overall survival:

The integrity of the HPV 16 E2 gene was next correlated with overall survival (in years). 39 patients in this cohort had data available for survival analyses i.e. number at risk (16 with disrupted E2 gene and 23 with an intact E2 gene). Median follow up time for this cohort was 35.5 months (approximately 3 years). Range of follow up was 0 – 67.6 months. The 25th November 2015 was the date the survival data was most up to date for this analysis, so time to event analyses were calculated to this fixed point. Kaplan-Meier survival analyses were performed for both the disrupted and intact HPV E2 groups. 1-year survival probabilities in the disrupted E2 group was 14/16 (87.5%) and in the intact E2 group was 21/23 (91.3%). Similarly, the 3-year survival probabilities in the disrupted E2 group was 5/16 (31.3%) compared with the intact group of 11/23 (47.8%). The 5-year Kaplan-Meier survival probability in the disrupted E2 group was 0/16 (0%) while in the intact group there was 1/23 (4.3%). The Kaplan-Meier probabilities show a trend of worse survival at 1, 3 and 5 years following diagnosis for oropharyngeal cancer in the cohort with disrupted HPV E2 gene but this trend was not statistically significant (p=0.395) on Log-rank test as shown in Figure 5.8.

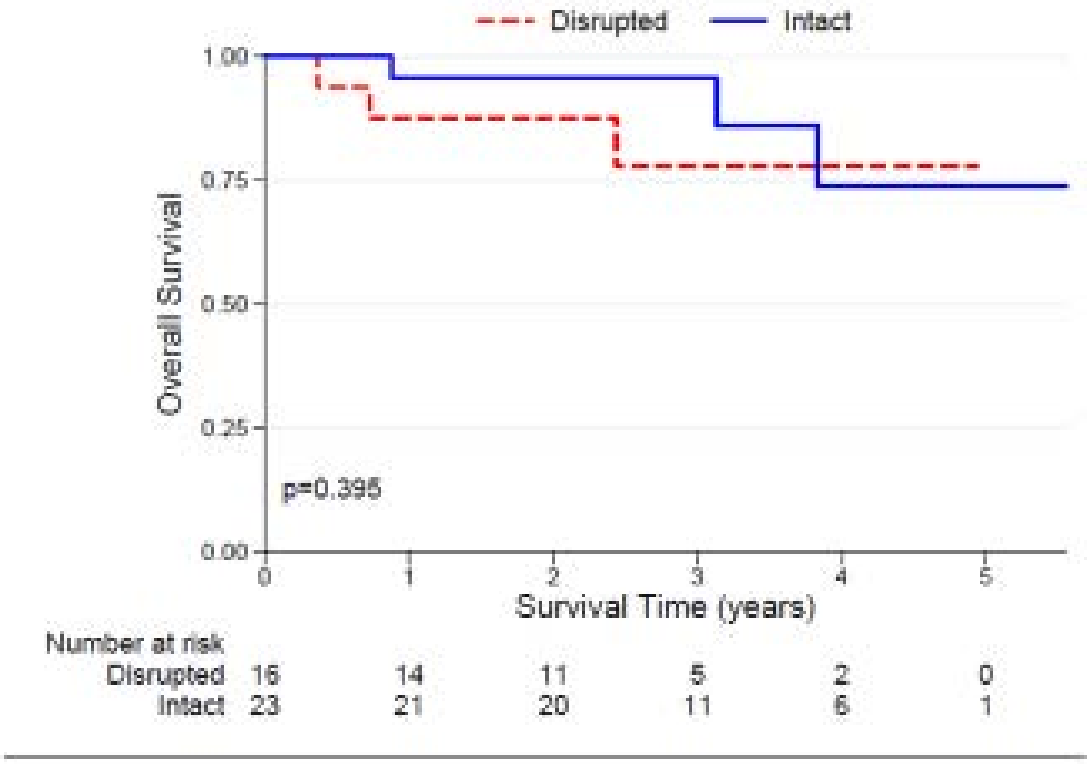


Figure 5.8: Impact of the HPV 16 E2 gene integrity on overall survival from OPC

5.6.3.7 HPV 16 E2 gene integrity and OPC recurrence:

The number of patients included in the Kaplan-Meier time-to-cancer recurrence analysis stratified by HPV 16 E2 gene status was 39 (16 disrupted and 23 intact). The number of events i.e. cancer recurrence was 6 with 2 recurrences in the E2 disrupted group and 4 recurrences in the E2 intact group. Time to recurrence (in months) in the 6 patients in ascending order was 1.1, 3.0, 4.5, 6.0, 10.5, 15.1. Oropharyngeal cancer recurrence in the 6 patients therefore had a median time to recurrence of 5.3 months and a range 1.1 – 15.1 months. There was no statistically significant difference observed between the E2 disrupted and E2 intact groups (p=0.767) on Log-rank test.

Cancer recurrence occurred in distant organs in 4 of the 6 patients with recurrent disease. 1 patient had a recurrence in the ipsilateral neck while 1 patient had recurrence in the contralateral neck. Significantly, Oropharyngeal cancer recurrence was associated with an increased likelihood of mortality on Kaplan-Meier survival analyses ($p<0.05$). 5 out of the 6 patients with oropharyngeal cancer recurrence (83.3%) died within the 1 year follow up period (Figure 5.9).

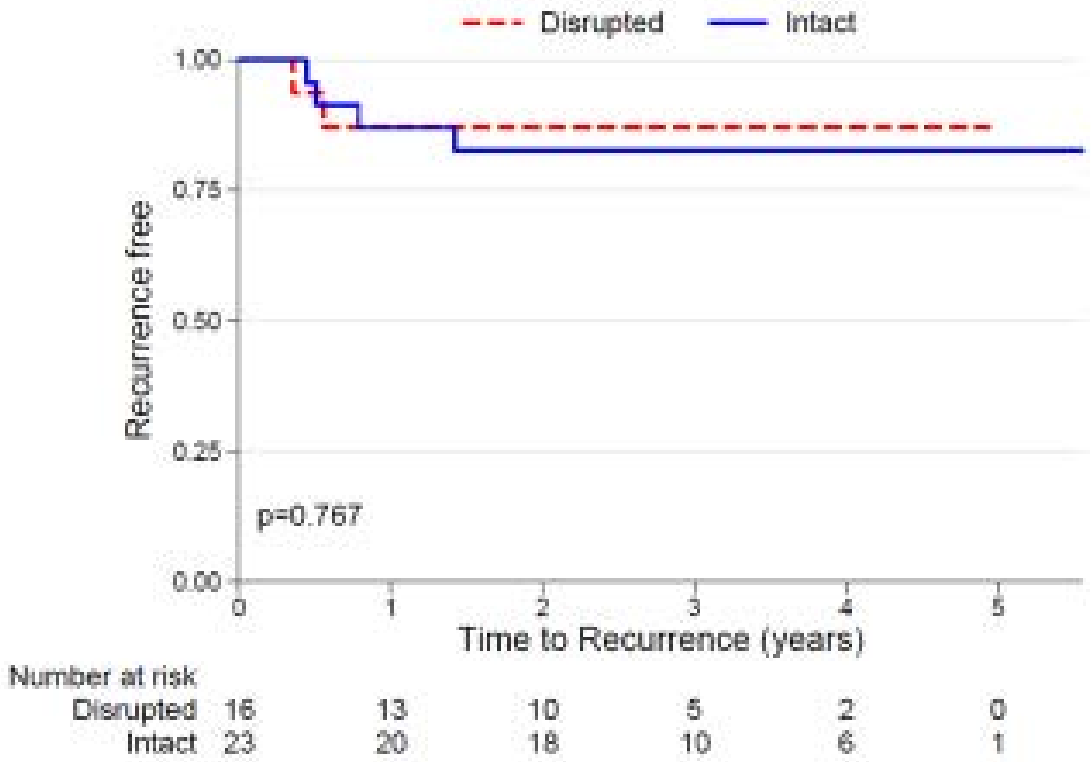


Figure 5.9: Kaplan-Meier analysis of the impact of HPV 16 E2 gene integrity on oropharyngeal cancer recurrence.

5.6.4 Methylation profiles of the E2 binding sites in the HPV 16 URR:

There were 39 oropharyngeal cancer DNA samples available for methylation studies in the HPV 16 URR. It was not possible to obtain any further DNA from one sample from the original 40 FFPE. A valid result for the E2BS methylation studies consisted of triplicate results that satisfied the thresholds set on the Pyro Q-CpG^{TM} software (Qiagen). Valid results did not include results with high background noise on pyrosequencing, unusual pattern sequences or results that lacked the internal controls for satisfactory bisulfite conversion (a cytosine that is not immediately followed by guanine on the HPV DNA sequence for amplification acts as the control for the efficiency of bisulfite modification). Valid results were depicted in colour blue on analysis with yellow and red coloured results indicative of intermediate and low confidence results, respectively. Only results that met the rigorous thresholds were included in methylation analyses as valid mean methylation at each E2BS.

Of the 39 OPC samples tested, 29 samples had valid results for investigating methylation profiles at the long control region (LCR) and 29 samples had methylation results for the promoter region of the HPV 16 URR. There was one sample (414) with the result for LCR only and one sample (201) with the result for the promoter region only. Unfortunately 10 samples did not amplify on pyrosequencing and these were investigated further to explain why this occurred. Of significance was that 9 out of the 10 samples that failed the amplification using pyrosequencing were in the HPV 16 E2 gene disrupted cohort. Primer-specific PCR for the HPV 16 URR confirmed the presence of disruptions within the URR (see section 5.6.8: Investigating HPV URR disruptions in select OPC samples).

The methylation thresholds were based on current literature in which similar studies were performed and include:

1. Low methylation: <20% methylation
2. Intermediate methylation: 20 – 80% methylation
3. High methylation: >80% methylation

Two cell line controls were used for pyrosequencing experiments and included a cell line from primary foreskin keratinocytes containing HPV and ‘CaSki’ which is a cervical cancer cell line and contains approximately 600 copies of integrated HPV with two transcriptionally active HPV. The HPV positive cell line from primary foreskin keratinocytes had low mean methylation at all the E2BS (Figure 5.10) while ‘CaSki’ had intermediate mean methylation levels at E2BS1 and high mean methylation at E2BS4 on pyrosequencing (Figure 5.11). The optimised assay was therefore able to detect the spectrum from low to high methylation and the examples of methylation results in tumour samples are shown in Figures 5.12, 5.13 and 5.14.

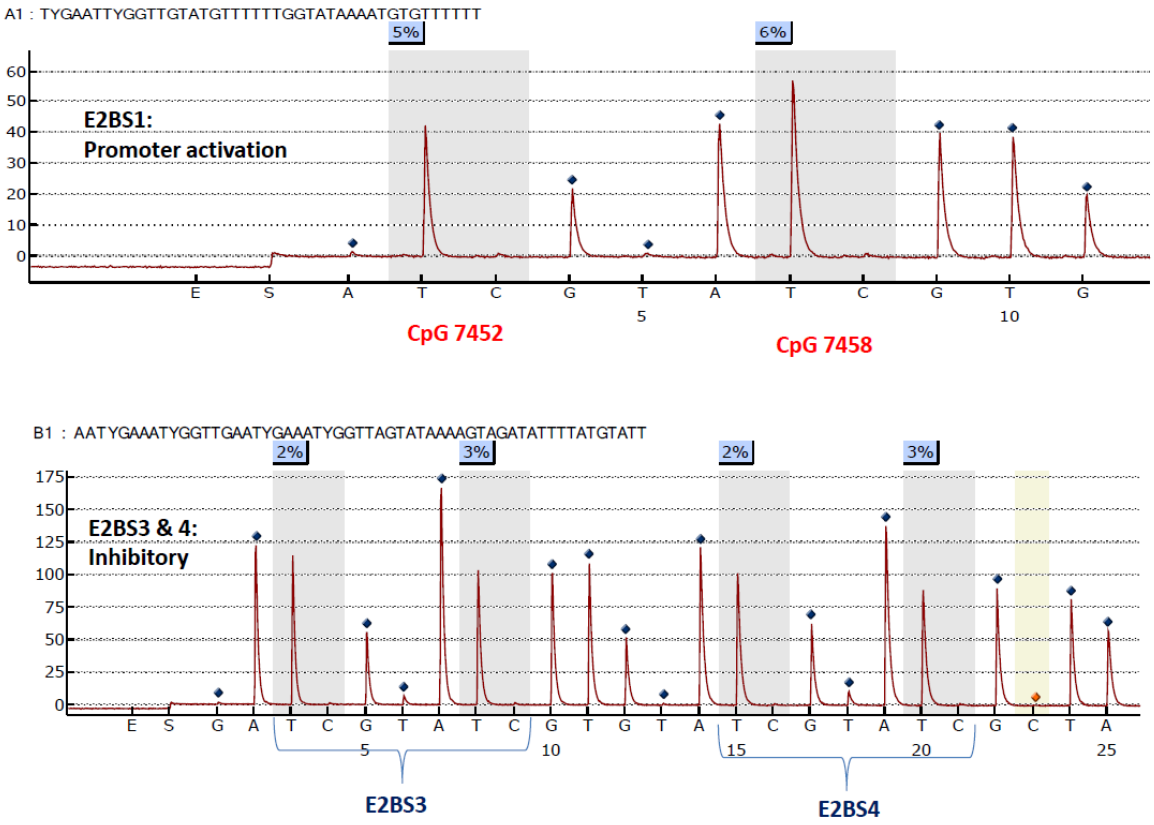


Figure 5.10: Low methylation at all E2BS in an HPV positive primary keratinocyte control with an intact E2 gene. *(A1) shows low methylation at the E2BS1 that modulates promoter activation with 5% and 6% mean methylation at CpG 7452 and CpG 7458, respectively. (B1) shows low mean methylation at both inhibitory E2BS3 and E2BS4. E2BS3 has mean methylation of 2% and 3% at CpG 37 and CpG 43, respectively while E2BS4 also has low mean methylation of 2% and 3% at CpG 52 and CpG 56, respectively. The 'C' not followed by a 'G' after the E2BS4 is the internal control for the efficiency of bisulfite conversion and is depicted as a red diamond.*

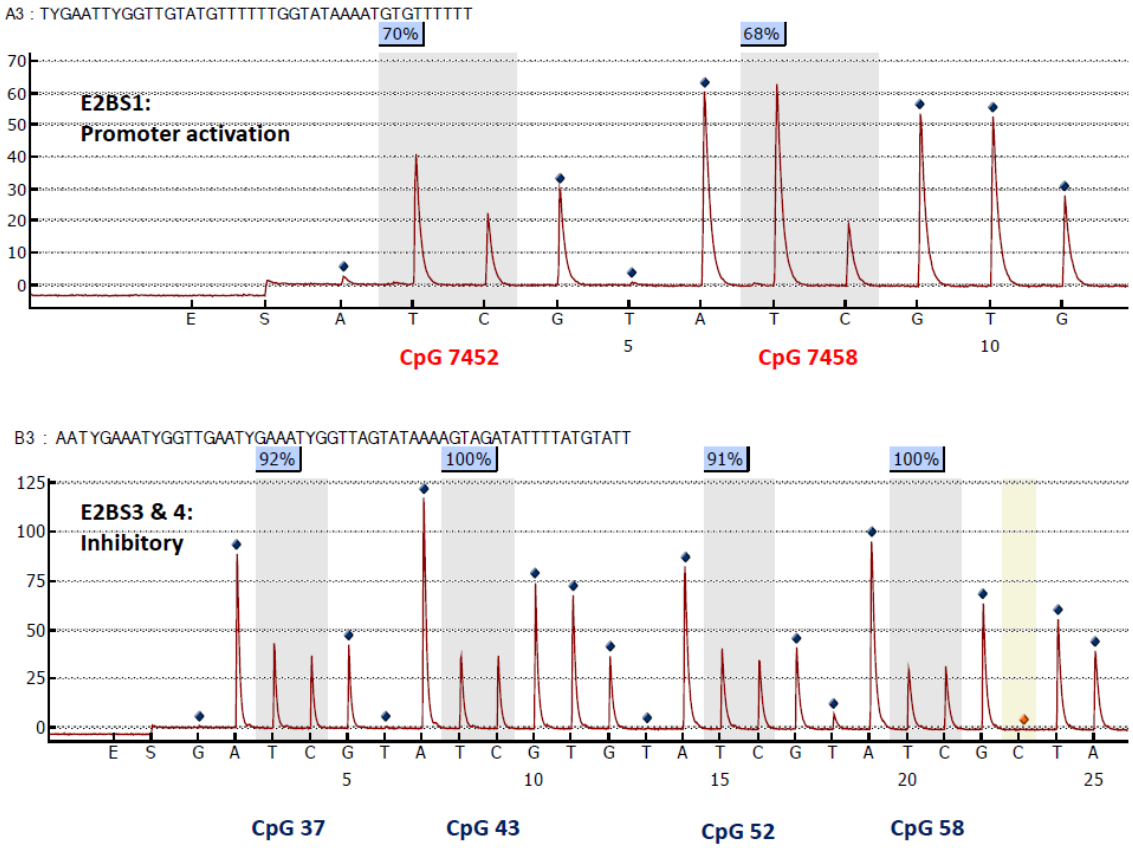


Figure 5.11: Methylation at E2BS in cervical cancer cell line 'CaSki'. *(A3) shows intermediate methylation at the E2BS1 that modulates promoter activation with 70% and 68% mean methylation at CpG 7452 and CpG 7458, respectively. (B3) shows high mean methylation at both inhibitory E2BS3 and E2BS4. E2BS3 has high mean methylation of 92% and 100% at CpG 37 and CpG 43, respectively while E2BS4 also has high mean methylation of 91% and 100% at CpG 52 and CpG 56, respectively. There was high correlation between mean methylation observed at both CpG sites within each E2 binding site which was statistically significant.*

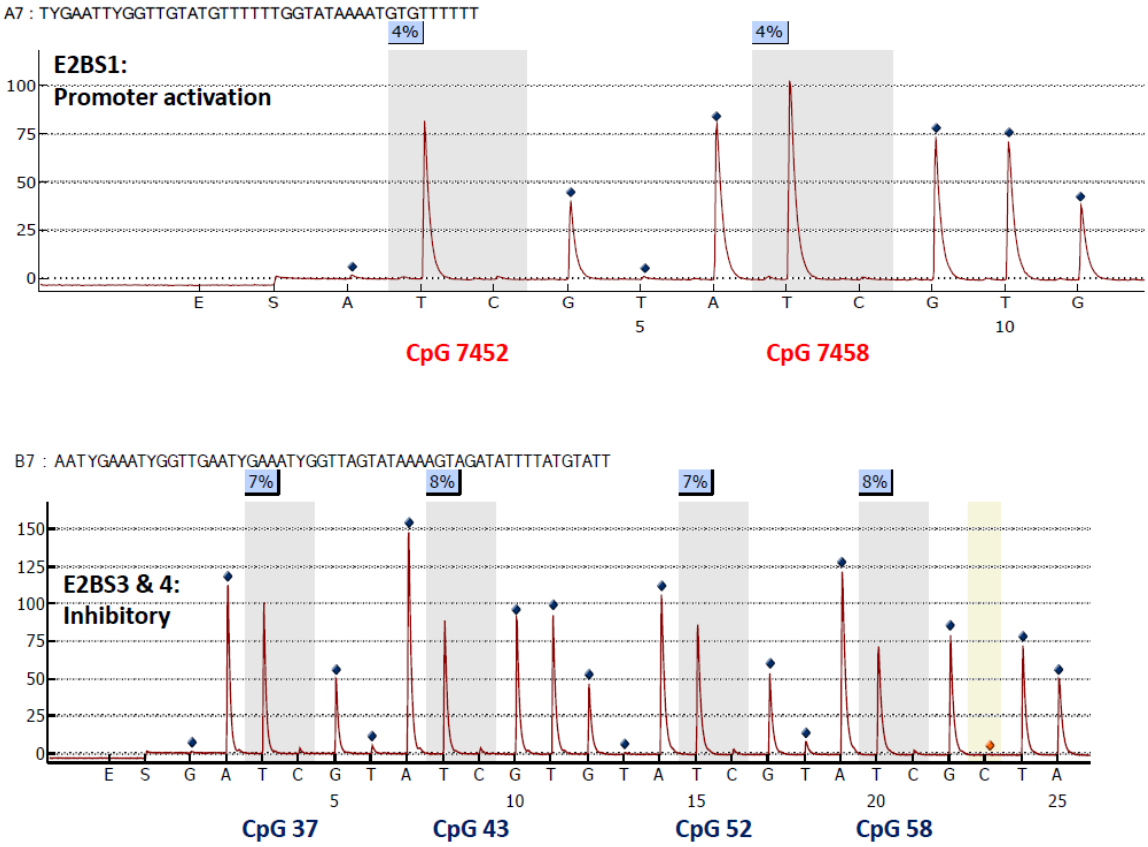


Figure 5.12: ***A representative patient sample (OPC 42) with an intact E2 gene and low mean methylation at all E2BS.*** *(A7) shows low methylation at the E2BS1 that modulates promoter activation with 4% and 4% mean methylation at CpG 7452 and CpG 7458, respectively. (B7) similarly shows low mean methylation at both inhibitory E2BS3 and E2BS4. E2BS3 has low mean methylation of 7% and 8% at CpG 37 and CpG 43, respectively while E2BS4 also has low mean methylation of 7% and 8% at CpG 52 and CpG 56, respectively.*

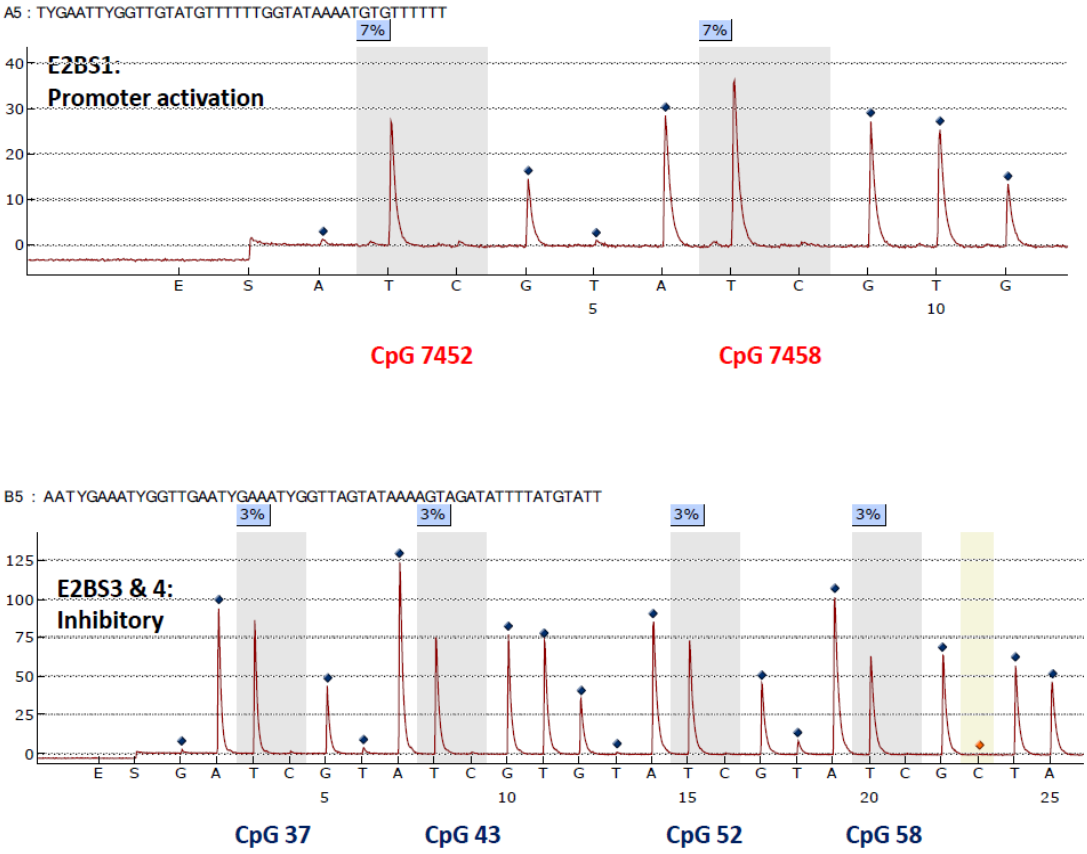


Figure 5.13: A representative patient sample (OPC 22) with a disrupted E2 gene and low mean methylation at all E2BS. *(A5) shows low methylation at the E2BS1 that modulates promoter activation with 7% and 7% mean methylation at CpG 7452 and CpG 7458, respectively. (B5) similarly shows low mean methylation at both inhibitory E2BS3 and E2BS4. E2BS3 has low mean methylation of 3% and 3% at CpG 37 and CpG 43, respectively while E2BS4 also has low mean methylation of 3% and 3% at CpG 52 and CpG 56, respectively.*

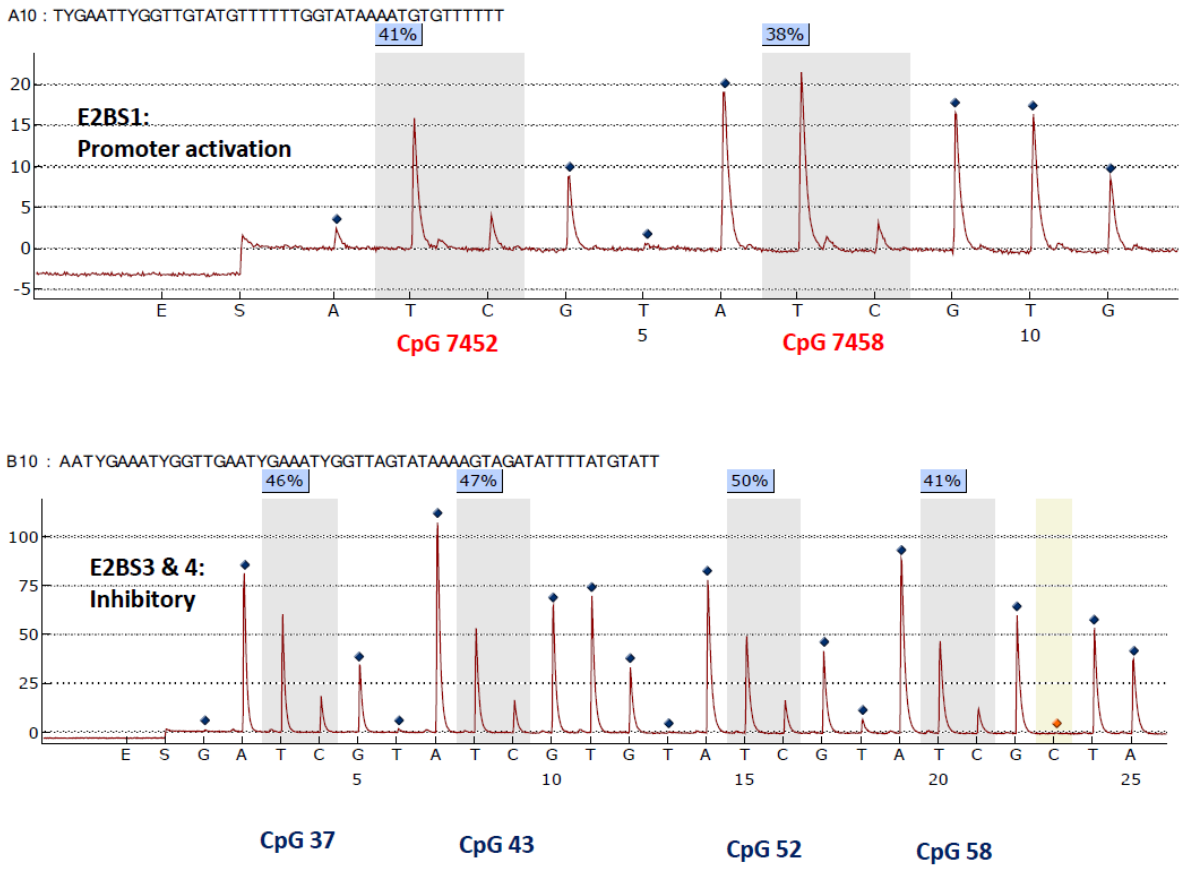


Figure 5.14: A representative patient sample (OPC 69) with a disrupted E2 gene and intermediate mean methylation at all E2BS. *(A10) shows intermediate methylation at the E2BS1 that modulates promoter activation with 41% and 38% mean methylation at CpG 7452 and CpG 7458, respectively. (B10) similarly shows intermediate mean methylation at both inhibitory E2BS3 and E2BS4. E2BS3 has intermediate mean methylation of 46% and 47% at CpG, 37 and CpG 43 respectively while E2BS4 also has intermediate mean methylation of 50% and 41% at CpG 52 and CpG 56, respectively.*

There were 29 samples from which mean methylation results were obtained in triplicates. Of the 29 samples, 7 had disruption of the HPV 16 E2 gene while 22 had intact HPV 16 E2 gene. There was no sample in the entire cohort (regardless of E2 status) that had high mean methylation i.e. methylation >80% at any E2 binding site.

In the E2 gene disrupted group (N – 7), at the promoter activating E2BS1, 14.3% (1/7) had intermediate mean methylation while the majority 85.7% (6/7) had low mean methylation. In the promoter inhibiting E2 binding sites 3 and 4, 28.6% (2/7) had intermediate mean methylation while 71.4% (5/7) had low mean methylation (Table 5.12).

Table 5.12: HPV 16 E2 gene status and methylation at E2BS in OPC (N – 29)

	E2 disrupted (N – 7)	E2 disrupted (%)	E2 intact (N – 22)	E2 intact (%)
Methylation (E2BS1) stimulatory				
High	0/7	0.0	0/22	0.0
Intermediate	1/7	14.3	1/22	4.6
Low	6/7	85.7	21/22	95.5
Methylation (E2BS3and4) Inhibitory				
High	0/7	0.0	0/22	0.0
Intermediate	2/7	28.6	5/22	22.7
Low	5/7	71.4	17/22	77.3

In the E2 gene intact group (N – 22), at the promoter activating E2BS1, 4.6% (1/22) had intermediate mean methylation while the vast majority 95.5% (21/22) had low mean methylation. In the promoter inhibiting E2 binding sites 3 and 4, 22.7% (5/22) had

intermediate mean methylation while 77.3% (17/22) had low mean methylation (see Table 5.12).

The mean of the methylation results obtained from each E2 binding site in the HPV 16 URR, in those patient samples that also had disruption of the regulatory E2 gene, is presented in Table 5.13.

Table 5.13: Methylation levels at E2BSs in OPC with disrupted E2 gene

This table shows OPC samples with E2 gene disruption. Individual methylation levels at each CpG site as well as the mean methylation levels of the CpG sites in each E2 binding site are shown. E2BS1 is mean methylation of CpG 7452, 7458. E2BS3 is mean methylation of CpG 37, 43. E2BS4 is mean methylation of CpG 52, 56. The overall inhibitory mean methylation was obtained from the average of methylation levels at E2BS3 and E2BS4. The samples highlighted in green (14, 22, 346, 214, 321) all have mean methylation less than 20% across all E2 binding sites. The sample highlighted in yellow (44) has mean methylation (≥ 20% to < 80%) in E2BS3 and E2BS4 but <20% mean methylation at E2BS1. The sample highlighted in red (69) has mean methylation levels (≥ 20% to < 80%) across all the E2 binding sites.

No	OPC ID	E2BS1	CpG 7452	CpG 7458	E2BS3	CpG 37	CpG 43	E2BS4	CpG 52	CpG 58	Overall Inhibitory (E2BS3 and4)
1	14	8.7	9.0	8.3	3.0	3.0	3.0	2.0	4.0	0.0	2.5
2	22	5.7	6.0	5.3	7.6	7.2	8.0	6.9	7.8	6.0	7.3
3	346	8.4	9.0	7.7	6.5	6.2	6.7	6.1	6.8	5.3	6.3
4	214	12.5	11.5	13.5	7.8	7	8.5	7.5	8	7	7.7
5	321	9.2	9.8	8.5	17.5	20.0	15.0	16.0	20.0	12.0	16.8
6	44	9.8	10.5	9.0	56.3	51.0	61.5	56.5	61.5	51.5	56.4
7	69	33.9	34.7	33.0	45.9	43.7	48.0	45.3	51.3	39.3	45.6
Mean		**12.6**	**12.9**	**12.2**	**20.7**	**19.7**	**21.5**	**20.0**	**22.8**	**17.3**	**20.4**

The pattern that emerged in the methylation results from the entire cohort (regardless of their E2 status) showed patient samples with a) low mean methylation (<20%) in all the E2 binding

sites (depicted in green), b) low mean methylation at E2BS1 (<20%) and intermediate methylation (20 – 80%) at E2BS3 and E2BS4 (depicted in yellow), and c) intermediate methylation at all the binding sites (depicted in red) (see Tables 5.13 and Table 5.14).

Specifically in the HPV 16 E2 gene disrupted cohort, 5/7 samples demonstrated low methylation in all the E2BS (Table 5.13). This pattern was similar to that observed in the majority of the E2 intact group where there was also low methylation across E2BSs (Table 5.14). One of the 7 samples had low methylation in E2BS1 and intermediate methylation in E2BS3 and E2BS4. There was also one sample with intermediate methylation in all E2BS sites. (Table 5.13). There was a high correlation observed in the mean methylation between the two CpG sites located in each E2 binding site regardless of E2 status. Overall, in the E2 gene disrupted group, the average of mean methylation readings at the promoter stimulating E2BS1 was 12.6% while it was 20.7% and 20.0% at the promoter inhibiting E2BS3 and E2BS4, respectively. Although the data show a slightly higher mean methylation at the inhibitory E2BS3 and 4, when compared to E2BS1 within the E2 disrupted group, this difference was not statistically significant given the methylation means were practically in the same 'low methylation' category and the observed difference could simply be due to background noise in pyrosequencing.

In the group with intact E2 gene (N – 23), there were 22 samples with methylation results in the promoter stimulating E2BS1 and 22 samples with methylation results in the inhibitory E2BS3 and 4 (Table 5.14). 18/22 samples had low mean methylation (<20%) across all the E2 binding sites. 4/22 samples had intermediate mean methylation (20 – 80%) at the inhibitory E2BS3 and E2BS4 alongside low mean methylation in their associated E2BS1. One sample (400) had intermediate mean methylation across all the E2 binding sites (Table 5.14). There was also a high correlation shown at each E2BS in the mean methylation data obtained from

each CpG site located therein. In terms of the overall average of the mean methylation observed at the specific E2 binding sites (regardless of the 3 methylation patterns just described), there was low mean methylation across all sites in the HPV 16 E2 intact group with 8.2%, 13.2% and 13.1% at E2BS1, E2BS3 and E2BS4 respectively (Table 5.14).

Table 5.14: Methylation levels at E2BSs in OPC with Intact E2 gene

This table shows OPC samples with an intact E2 gene. Individual methylation levels at each CpG site as well as the mean methylation levels of the CpG sites in each E2 binding site are shown. E2BS1 is mean methylation of CpG 7452, 7458. E2BS3 is mean methylation of CpG 37, 43. E2BS4 is mean methylation of CpG 52, 56. The overall inhibitory mean methylation was obtained from the average of methylation levels at E2BS3 and E2BS4. The samples highlighted in green all have mean methylation less than 20% across all E2 binding sites. The samples highlighted in yellow (201, 491, 141, 372)) have mean methylation (≥ 20% to < 80%) in E2BS3 and E2BS4 but <20% mean methylation at E2BS1. The sample highlighted in red (400) has mean methylation levels (≥ 20% to < 80%) across all the E2 binding sites.

No	OPC ID	E2BS1	CpG 7452	CpG 7458	E2BS3	CpG 37	CpG 43	E2BS4	CpG 52	CpG 58	Overall Inhibitory (E2BS3 and4)
1	320	2.5	2.7	2.3	3	3	3	2.8	3	2.5	2.9
2	42	4.5	4.7	4.3	6.9	6	7.7	7.4	7.7	7	7.2
3	145	5	5.3	4.7	3.1	4.3	1.8	3	4.3	1.7	3.1
4	39	5.1	4.8	5.3	6	5	7	6	6	6	6
5	185	5.4	6.7	4	3.3	4	2.5	2.8	2.5	3	3.1
6	537	5.4	4.7	6	4.5	4.3	4.7	3.7	3.7	3.7	4.1
7	414	5.5	6	5	N/A	N/A	N/A	N/A	N/A	N/A	N/A
8	48	5.5	5.3	5.7	6.3	6.8	5.8	6.8	6.5	7	6.6
9	19	5.9	6	5.8	4.8	5.3	4.3	4.2	4	4.3	4.5
10	450	6	6.3	5.7	7	8	6	7.3	6.5	8	7.2
11	142	6.3	5.3	7.3	4.7	5.7	3.7	4	3.7	4.3	4.4
12	311	6.4	6.7	6	7.2	7.3	7	7.2	7	7.3	7.2
13	224	7.4	7	7.7	3.5	2.5	4.5	2.5	3	2	3
14	451	7.4	7.3	7.5	6	6.7	5.3	6	6	6	6
15	525	8.4	8.7	8	4	4.3	3.7	4.4	4	4.7	4.2
16	57	8.5	8.6	8.4	11.3	10.25	12.25	10.3	10.75	9.75	10.8
17	55	10.4	10.7	10	6	5	7	4.5	9	0	5.3
18	418	12.2	11	13.3	7.7	8	7.3	8	7.7	8.3	8
19	201	N/A	N/A	N/A	22.9	21	24.7	23.7	25	22.3	23.3
20	491	11.6	11.8	11.3	25.9	28.7	23	25.8	28.3	23.3	25.9
21	141	11.9	12.7	11	40.3	38.5	42	38.7	39.3	38	39.5
22	372	17.3	17.3	17.3	61.2	59.5	62.8	60.2	60	60.3	60.7
23	400	20.8	21.3	20.3	48.5	47.7	49.3	48.7	49	48.3	48.6
Mean		**8.2**	**8.2**	**8.0**	**13.4**	**13.3**	**13.4**	**13.1**	**13.5**	**12.6**	**13.3**

A comparative analysis of mean methylation recorded at each E2 binding site stratified by their E2 gene status was undertaken in line with the set objectives of this study with results detailed in Table 5.13 and Table 5.14. The mean methylation data for the entire OPC cohort was non-parametric with a left skew (see Figures 5.15 and 5.16), therefore a Mann-Whitney U-test was used to assess for any differences between the two independent groups. Significance level was set at $p<0.05$.

At E2BS1, the overwhelming majority of the samples were hypomethylated with mean methylation <20% in 27/29 samples (93.1%) and intermediate methylation in only 2/29 (6.9%) (Figure 5.15 and Tables 5.13, 5.14).

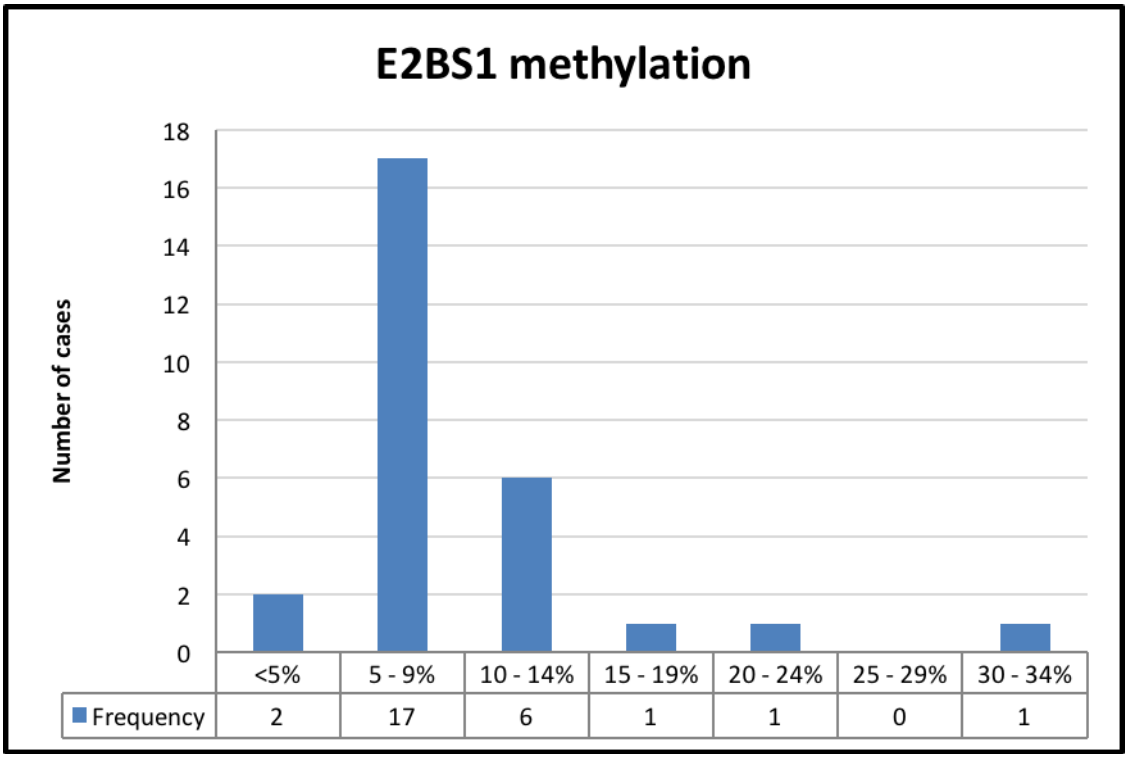


Figure 5.15: Methylation distribution at E2BS1 for all samples (N – 29)

Similarly, at the inhibitory E2BS3 and E2BS4, there were 21/29 samples (72.4%) with low mean methylation and 8/29 (27.6%) with intermediate methylation (Figure 5.16 and Tables 5.13 and 5.14).

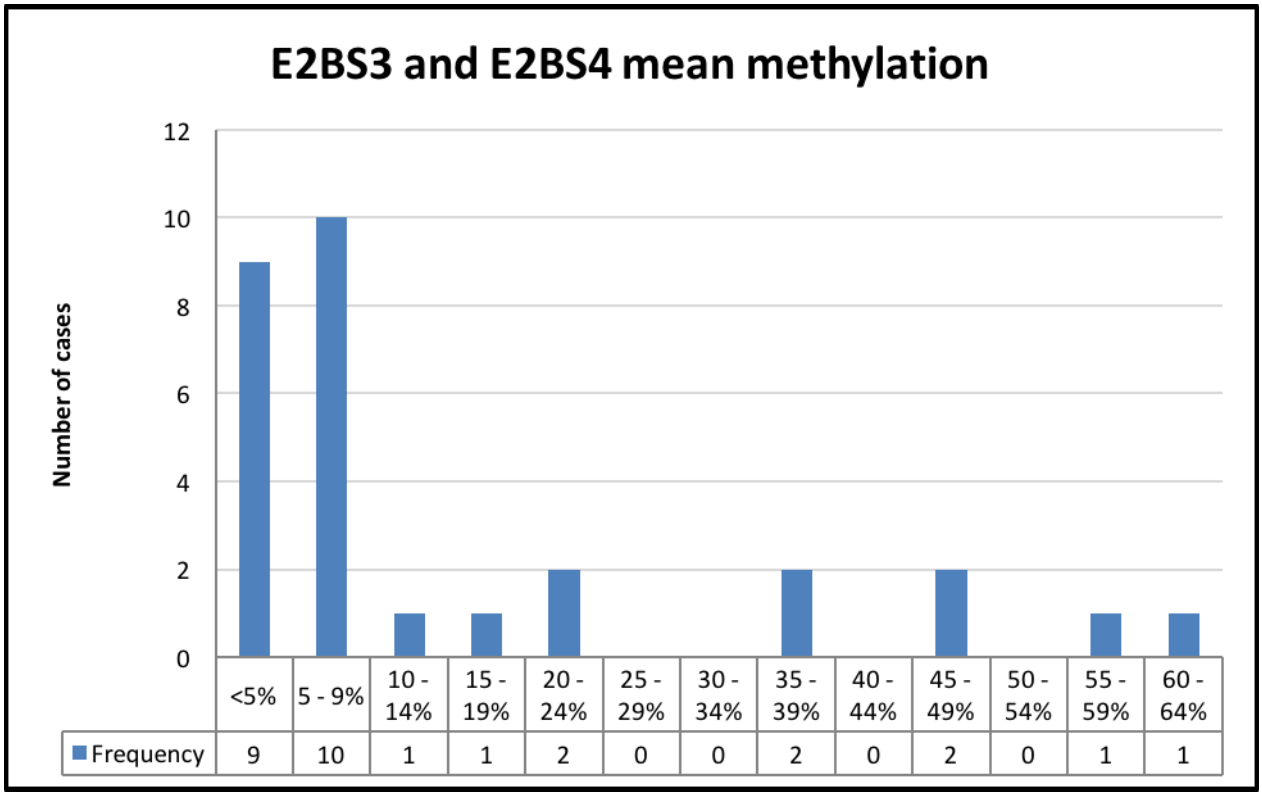


Figure 5.16: Methylation distribution at E2BS3 and 4 for all samples (N – 29)

The promoter stimulating E2BS1 had higher mean methylation in the disrupted E2 group (12.6%) when compared with the intact E2 group (8.2%). This difference was tending towards significance at p=0.061. There was however no conclusive difference as the two independent means were both <20% which puts them in the same 'low methylation' category. There was no statistically significant difference when the individual CpG sites at E2BS1 (CpG 7452 and 7458) were assessed with p=0.064 and p=0.093, respectively (Table 5.15).

Table 5.15: Mean methylation by E2 status at each E2BS in the HPV 16 URR of OPCs

	Methylation (%) in OPC with intact E2	Methylation (%) in OPC with disrupted E2	2 tailed P-value (Mann Whitney U-test)
E2 Binding Site 1	**8.2**	**12.6**	**0.061**
CpG 7452	8.2	12.9	0.064
CpG 7458	8.0	12.2	0.093
E2 Binding Site 3	**13.4**	**20.7**	**0.157**
CpG 37	13.3	19.7	0.223
CpG 43	13.4	21.5	0.135
E2 Binding Site 4	**13.1**	**20.0**	**0.231**
CpG 52	13.5	22.8	0.093
CpG 58	12.6	17.3	0.338
Overall Inhibitory (E2BS3 and 4)	**13.3**	**20.4**	**0.189**

At the promoter inhibiting E2BSs, overall methylation when E2BS3 and E2BS4 mean methylation were combined, showed higher mean methylation in the E2 disrupted group (20.4%) when compared with the E2 intact group (13.3%). This was however not statistically significant on Mann-Whitney U test (p=0.189) (Table 5.15). When assessed separately, there was still no observed difference with E2BS3 having 20.7% and 13.4% in the 'E2 disrupted' and 'E2 intact' groups, respectively (p=0.157). Similarly, at the E2BS4, intermediate methylation of 20.0% was found in the 'E2 disrupted' group in contrast to low methylation of

13.1% in the 'E2 intact' group but not statistically significant on Mann-Whitney U test (p=0.231). There was still no difference in mean methylation by E2 status when assessed at individual CpG sites (Table 5.15).

5.6.5 Correlation of methylation levels within E2 binding sites:

Spearman's Correlation coefficient showed a strong positive association between the mean methylation at CpG sites 7452 and 7458 (R=0.919 and p<0.05) (Figure 5.17)

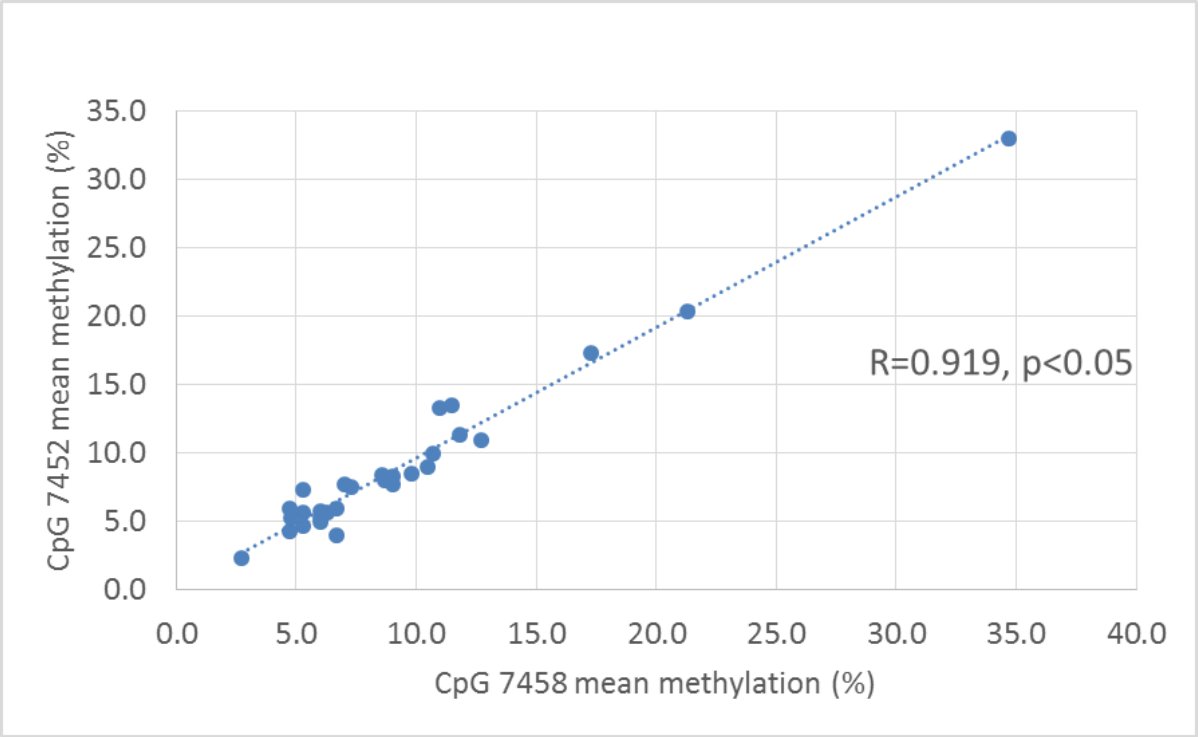


Figure 5.17: Correlation of mean methylation at CpG sites 7452 and 7458 in E2BS1

There was a strong correlation in the mean methylation obtained at CpG sites within the inhibitory E2BS3 (CpG 37 and 43) with a R=0.908 and p<0.05 (Figure 5.18).

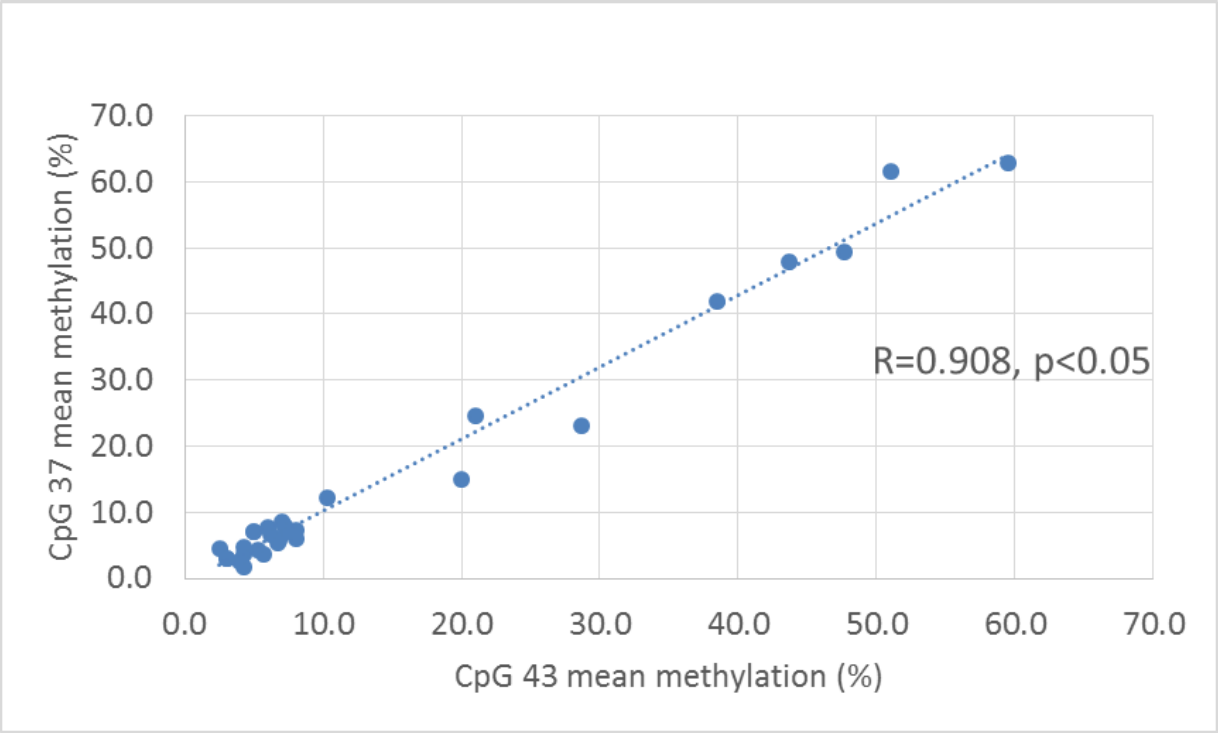


Figure 5.18: Correlation in mean methylation at the E2BS3 CpG sites 37, 43.

Mean methylation at the CpG sites in E2BS4 (CpG 52 and 56) also showed strong positive correlation on using Spearman's correlation coefficient ($R=0.849$, $p<0.05$) (Figure 5.19)

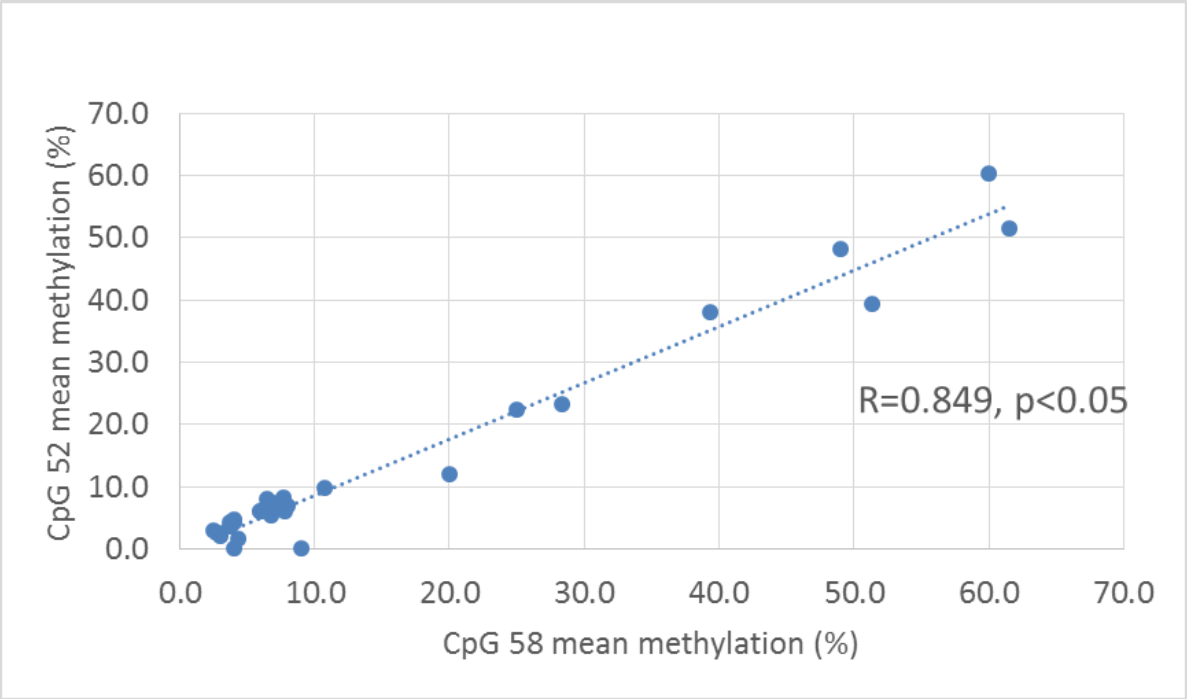


Figure 5.19: Positive correlation of CpG sites 52, 58 at the E2BS4.

5.6.6 Overall survival in OPC by E2BS1 methylation and E2 status:

The impact of E2BS1 methylation on overall survival was assessed using Kaplan-Meier analysis. There were 29 samples that had both methylation data and survival data. Of the 29 samples, 27 samples had low methylation (6 E2 gene disrupted and 21 E2 gene intact). There were 2 patient samples that had intermediate methylation (one each of disrupted and intact E2 gene). There were 4 events (deaths) in this cohort and in all 4 patient samples, there was low methylation. The 2 patients with intermediate methylation were still alive. There was no demonstrable difference in overall survival between those patients with low mean methylation at E2BS1 and those with intermediate methylation on log-rank test (p=0.560) (Figure 5.20).

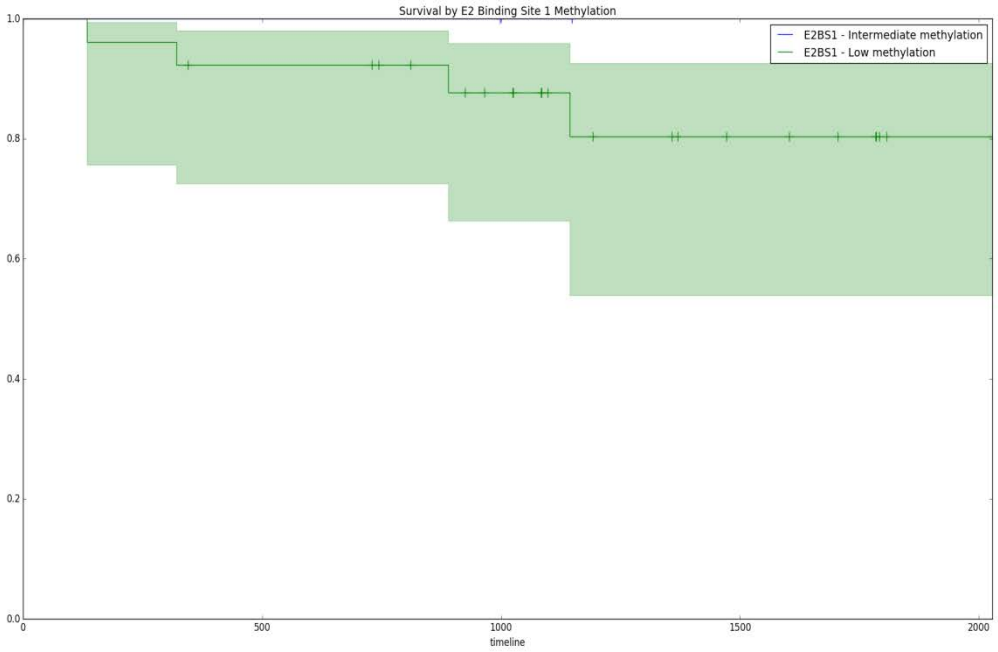


Figure 5.20: E2BS1 methylation categories and survival. *No signficant impact of E2BS1 methylation observed on survival.*

5.6.7 E2BS3 and 4 intermediate methylation with disrupted E2 and survival:

Of the 29 patients, low methylation was present at the E2BS3 and E2BS4 in 22 patients (5 disrupted, 17 intact). There were 7 patients with intermediate methylation (2 disrupted, 5 had the viral E2 gene intact). There was no statistically significant difference in overall survival between the E2 intact and E2 disrupted groups on log rank test (p=0.731).

There was no difference in overall survival by E2 status in the inhibitory E2 binding sites 3 and 4 with low methylation on log rank test (p=0.158) (Figure 5.21).

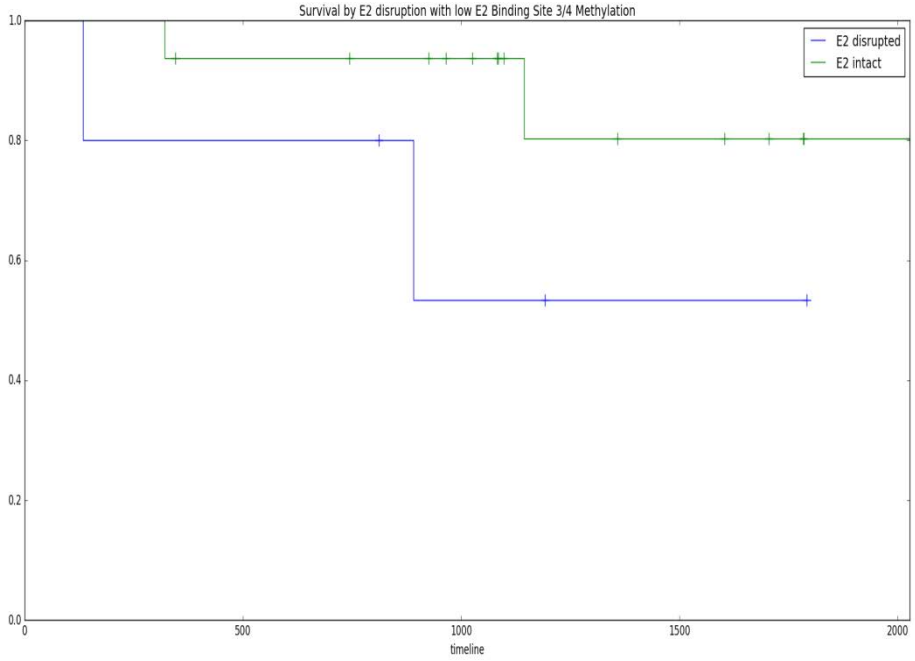


Figure 5.21: Overall survival in OPC patients with E2BS3 and 4 low methylation and disrupted E2. *No significant impact on survival shown in OPC cases with low methylation at E2BS3 and E2BS4 and disrupted E2.*

5.6.8 Investigating HPV URR disruptions in select OPC samples:

10 Oropharyngeal cancer DNA samples failed to amplify prior to pyrosequencing. These samples were therefore excluded from the analysis on methylation profiles in the HPV 16 URR. Further exploratory tests were performed to examine why amplification was unsuccessful in these samples.

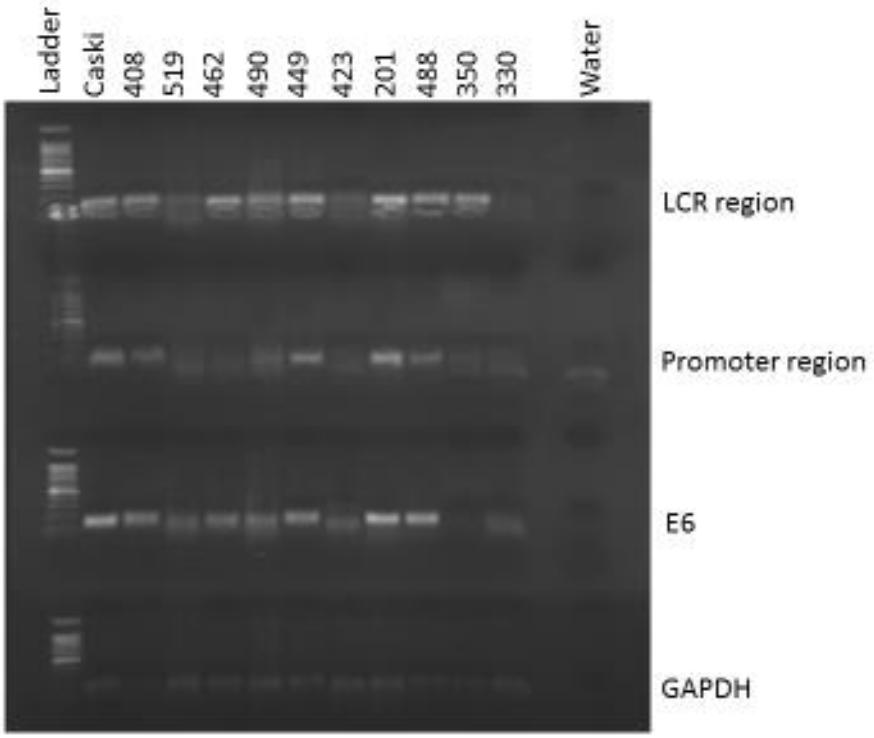


Figure 5.22: Investigation of HPV URR disruptions by PCR

HPV 16 URR integrity was assessed in 10 OPC samples by PCR reactions set up using 10µL of PCR mastermix, 3.5µL of water, 1.5µL of primers and 5µL of DNA. PCR was performed running denaturation step at 95°C for 15minutes, followed by 50 cycles at 95°C for 20seconds, 54°C for 40seconds, 72°C for 50seconds. Final step of 72°C for 1 minute and 4°C holding temperature. The PCR reactions included a positive control (Caski), a negative control (DNAse – free sterile water), a house keeping gene (GAPDH) and a viral positive control (HPV 16 E6 gene). The reactions were separated for both the Long Control Region (LCR) and the Promoter region of the viral HPV 16 URR.

The initial observation was that 9/10 samples also had HPV 16 E2 gene disruption. From current literature, it is now known that HPV integrates in various regions of the HPV genome and it was possible that HPV integrations had occurred in tandem (alongside the E2 gene disruption observed) at the viral upstream regulatory region. This was investigated using primers specific to the HPV 16 URR using PCR performed in duplicates on DNA samples.

The first PCR result is displayed in Figure 5.22 above.

Remarkably, there were disruptions (absence of PCR bands) observed in the HPV 16 URR in the following OPCs:

LCR region: OPC 519, 423, 330

Promoter region: OPC 519, 423, 330, 462, 490, 350

The OPC samples that did have bands present on PCR may have failed bisulfite conversion or simply contain too low DNA concentrations to amplify on pyrosequencing. OPC 350 was found to also be disrupted at E6 while OPC 408 did not give amplification product for GAPDH and hence was excluded.

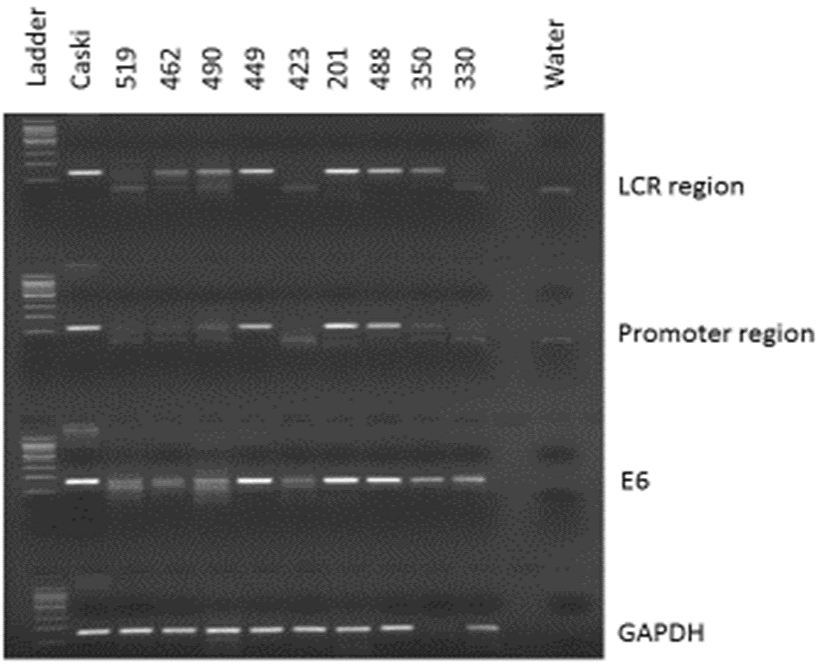


Figure 5.23: Duplicate PCR investigating HPV URR disruptions

HPV 16 URR integrity was re-assessed in 9 OPC samples by duplicate PCR reactions set up using 10µL of PCR mastermix, 3.5µL of water, 1.5µL of primers and 5µL of DNA. PCR was performed running denaturation step at 95°C for 15minutes, followed by 50 cycles at 95°C for 20seconds, 54°C for 40seconds, 72°C for 50seconds. Final step of 72°C for 1 minute and 4°C holding temperature. The PCR reactions included a positive control (Caski), a negative control (DNAse – free sterile water), a house keeping gene (GAPDH) and a viral positive control (HPV 16 E6 gene). The reactions were separated for both the Long Control Region (LCR) and the Promoter region of the viral HPV 16 URR.

The disruptions were confirmed in duplicate PCR (Figure 5.23).

In both the LCR and promoter regions, the following OPC had absent bands on PCR: OPC 519, 423, 330. There was faint amplification in OPC 462, 490, 350 in the promoter region of the HPV 16 URR (Figure 5.23). It can be deduced that the low levels of DNA concentrations were below pyrosequencing detection thresholds. Some samples did however amplify well on

duplicate PCR including OPC 449, 201, 488. The house-keeping gene (GAPDH) was also absent in OPC 408 and 350 and as such, both samples were excluded from further analysis. By the time of this experiment, the DNA stores were depleted and no further experiment was possible.

Overall, Pyrosequencing results of the E2 binding sites in the HPV 16 URR showed the following:

1. Overall hypomethylation in the HPV 16 URR
2. Hypomethylation at the promoter stimulating E2 binding site 1
3. Moderate methylation at the inhibitory E2 binding sites 3 and 4
4. There are no statistical differences at E2 binding sites for methylation between intact and integrated HPV
5. There was no statistically significant correlation between mean methylation at E2 binding sites and overall survival in the small number of samples assessed.

5.6.9 HPV 16 E2, E6, E7 RNA expression profiles in Oropharyngeal cancer using reverse transcription real-time quantitative PCR (qPCR):

40 OPC samples were further investigated for RNA expression of HPV 16 E2, E6 and E7 genes. These results were normalised to β-actin with VU147 (an HPV positive cell line) as the endogenous control. The average fold changes were obtained using the standard $2^{-\Delta\Delta C_T}$ calculations and then log transformed.

3 OPC samples did not express the control (β-actin) and were hence excluded from the analysis. The excluded samples included a sample with disrupted E2 gene (OPC 346) while the other two samples had intact HPV 16 E2 gene (OPC 525, 201). In addition, there were no available samples for further analyses in 2 patients (OPC 70 and 414).

Three Isoforms of the HPV 16 E2 RNA were investigated for any clinical correlation with duplicate real-time PCR. These isoforms were 'E2', 'E2 880^2709', and 'E2 880^2582'. These E2 isoforms have been previously investiged by Soeda et al. (2016) who showed that translation of the E2 protein was mainly by the 880^2709 RNA species with only very low levels of 880^2582 mRNA species.

All samples expressed at least one isoform of E2 RNA except a single sample with disrupted E2 gene – OPC 423 (Table 5.16). The E2 expression fold change relative to reference VU147 sample was calculated for each triplicate and the average E2 fold changes with their corresponding standard deviations (Table 5.16).

Table 5.16: E2 RNA expression in oropharyngeal cancer by qPCR *Average fold change for E2 RNA expression assessed by qPCR in triplicate reactions. Samples highlighted in pink have E2 gene disruption while samples highlighted in white have an intact E2 gene. Reference sample VU147 is highlighted in green.*

Tumour ID	ΔΔCт	E2 Fold change	ΔΔCт2	E2 Fold change_2	ΔΔCт3	E2 Fold change_3	Average E2 fold change	Standard deviation E2
VU147	0	1	0	1	0	1	1	0
185	-2.4	5.3	-0.7	1.6	5.7	0.0	2.3	2.7
330	-2.1	4.2	-1.9	3.6	-2.0	3.9	3.9	0.3
519	-0.4	1.3	4.5	0.0	-4.1	17.2	6.2	9.5
57	0.3	0.8	-5.1	34.1	-0.2	1.2	12.0	19.1
449	4.3	0.0	5.6	0.0	-5.3	39.6	13.2	22.9
462	0.2	0.9	7.0	0.0	-6.2	71.5	24.1	41.0
490	8.3	0.0	0.1	0.9	-7.1	136.1	45.7	78.3
55	-8.1	283.9	0.1	0.9	-4.0	16.2	100.3	159.1
142	-5.7	53.8	-7.9	237.0	-6.8	112.9	134.6	93.5
224	-8.7	408.6	-6.6	97.7	-2.9	7.3	171.2	210.5
321	-7.5	177.1	-7.7	214.0	-8.3	322.4	237.8	75.5
408	-9.7	816.5	-5.6	49.6	1.9	0.3	288.8	457.7
44	-6.7	101.0	-9.2	601.7	-7.9	246.5	316.4	257.6
145	-10.6	1502.2	-6.2	74.5	-2.1	4.2	527.0	845.3
14	-11.6	3057.8	-4.5	22.3	-8.0	260.9	1113.7	1687.9
451	-5.9	58.6	-5.0	32.8	-11.8	3631.4	1240.9	2070.2
42	-9.8	870.6	-11.3	2545.7	-8.6	379.3	1265.2	1135.8
48	-4.6	23.7	-4.2	18.9	-12.2	4810.1	1617.6	2764.8
69	-12.7	6580.4	-10.0	1021.7	-1.6	3.1	2535.1	3540.2
19	-5.1	33.1	-5.0	32.3	-13.1	8530.9	2865.4	4906.4
488	-0.1	1.0	-1.3	2.5	-14.0	16599.6	5534.4	9582.8
214	-5.4	41.1	-14.2	18732.5	-9.8	877.9	6550.5	10558.2
22	-13.6	12108.4	-15.0	32168.0	-8.0	255.1	14843.8	16131.3
39	-15.5	45091.5	-14.6	24100.7	-8.1	274.3	23155.5	22423.6
450	-6.6	94.1	-16.5	93824.3	0.2	0.9	31306.4	54142.1
418	-7.7	211.4	-16.6	96595.4	0.6	0.7	32269.2	55708.3
311	-16.4	85467.0	-12.4	5376.2	-14.5	22939.3	37927.5	42096.5
423	6.1	0.0	-16.9	122981.8	7.4	0.0	40993.9	71003.5
372	-11.2	2300.2	-17.3	158252.6	-1.8	3.4	53518.7	90709.5
537	-7.3	162.6	-18.0	269411.8	-1.7	3.3	89859.3	155497.1
400	-8.2	294.0	-18.2	292755.8	-2.1	4.3	97684.7	168936.6
491	-7.5	179.2	-18.6	398886.2	-0.5	1.5	133022.3	230244.9
350	-6.5	87.6	-7.0	131.7	-20.4	1430917.7	477045.6	826077.4
141	-11.4	2785.4	-22.2	4759072.7	-6.0	65.0	1587307.7	2746829.4
320	-15.4	41963.9	-27.3	167762170.8	-5.9	61.6	55934732.1	96845405.0

There was an inverse relationship observed between E2 gene integrity and E2 RNA expression. There was lower E2 RNA expression in those oropharyngeal cancer with a disrupted E2 gene compared to those with an intact E2 gene (mixed infection) (Table 5.17, Figure 5.24). This is consistent with published literature (Reuschenbach et al., 2015). E2 RNA expression is not compatible with E2 gene disruption so it is likely these samples had both intact and disrupted forms (mixed) that allowed the limited expression of some E2 RNA. The limitation of the E2 disruption technique used has already been discussed.

Table 5.17: E2 RNA average fold change in comparison to VU147 *Average fold change for E2 RNA expression assessed by qPCR in triplicate reactions. Samples highlighted in pink have E2 gene disruption while samples highlighted in white have an intact E2 gene. Reference sample VU147 is highlighted in green.*

Tumour ID	Average E2 fold change	Standard deviation E2	Log10(fold change) E2	E2 status
VU147	1	0	0	Intact
185	2.3	2.7	0.4	Intact
330	3.9	0.3	0.6	Disrupted
519	6.2	9.5	0.8	Disrupted
57	12.0	19.1	1.1	Intact
449	13.2	22.9	1.1	Disrupted
462	24.1	41.0	1.4	Disrupted
490	45.7	78.3	1.7	Disrupted
55	100.3	159.1	2.0	Intact
142	134.6	93.5	2.1	Intact
224	171.2	210.5	2.2	Intact
321	237.8	75.5	2.4	Disrupted
408	288.8	457.7	2.5	Disrupted
44	316.4	257.6	2.5	Disrupted
145	527.0	845.3	2.7	Intact
14	1113.7	1687.9	3.0	Disrupted
451	1240.9	2070.2	3.1	Intact
42	1265.2	1135.8	3.1	Intact
48	1617.6	2764.8	3.2	Intact
69	2535.1	3540.2	3.4	Disrupted
19	2865.4	4906.4	3.5	Intact
488	5534.4	9582.8	3.7	Disrupted
214	6550.5	10558.2	3.8	Disrupted
22	14843.8	16131.3	4.2	Disrupted
39	23155.5	22423.6	4.4	Intact
450	31306.4	54142.1	4.5	Intact
418	32269.2	55708.3	4.5	Intact
311	37927.5	42096.5	4.6	Intact
423	40993.9	71003.5	4.6	Disrupted
372	53518.7	90709.5	4.7	Intact
537	89859.3	155497.1	5.0	Intact
400	97684.7	168936.6	5.0	Intact
491	133022.3	230244.9	5.1	Intact
350	477045.6	826077.4	5.7	Disrupted
141	1587307.7	2746829.4	6.2	Intact
320	55934732.1	96845405.0	7.7	Intact

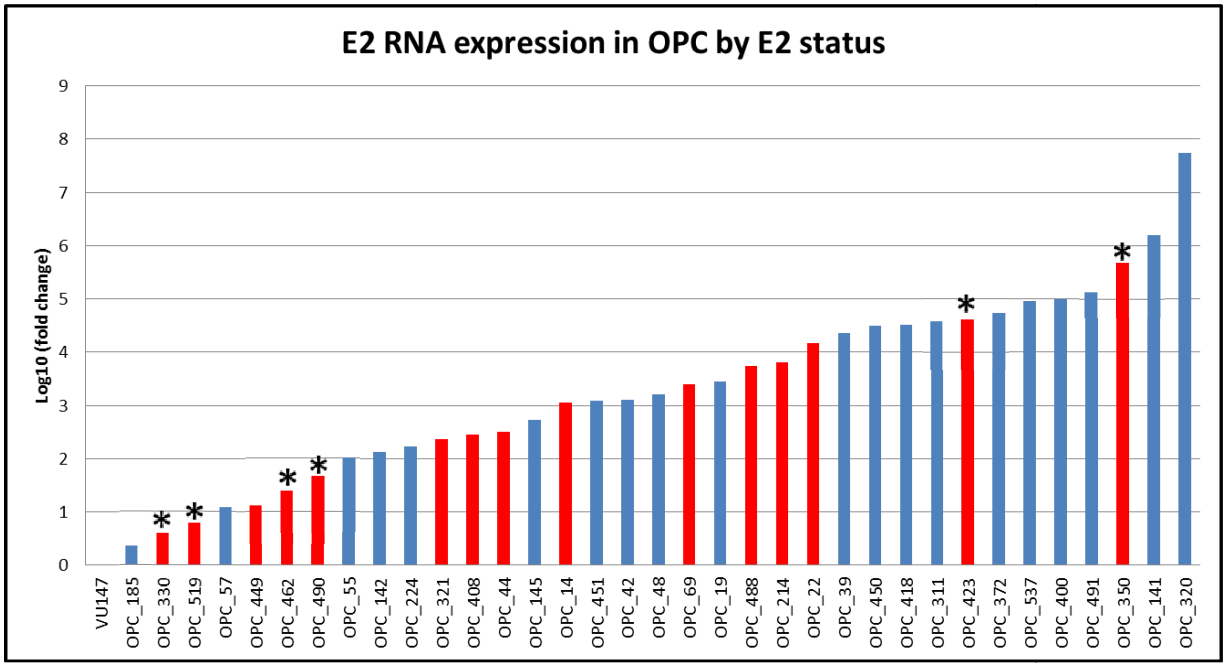


Figure 5.24: HPV 16 E2 RNA expression in OPCs *Average fold change for E2 RNA expression was assessed by real-time PCR in triplicate reactions and calculated in relation to the VU147 reference sample. OPC samples with E2 disruption are shown in red while OPC samples with an intact E2 gene are shown in blue. Samples with additional disruptions in the URR (330, 519, 462, 490, 423 and 350) are highlighted with an asterix **

A good proportion of the oropharyngeal cancer with high expression levels of E2 RNA also had an intact E2 gene as shown on the right side of Figure 4.25. Of the samples with additional URR disruption, four of them have low E2 expression (330, 519, 462 and 490) and two of them have the highest E2 expression amongst samples with disrupted E2 gene (423, 350). These samples had both E2 disruption and URR disruption and are highlighted with an asterix * in Figure 5.24

5.6.10 E6 expression and E2 gene disruption:

There were high levels of E6 expression in the oropharyngeal cancer cohort with E2 gene disruption when compared with those cancer samples with an intact E2 gene (Table 5.18). This is consistent with current literature on HPV carcinogenesis. The samples with E2 gene disruption are depicted in red, while the samples with an intact E2 gene are not coloured in Table 5.18. The positive control is VU147 shown in green. The standard deviations for each average fold change in E6 expression alongside the log transformation are shown. There were three samples without E6 expression data depicted as ‘not available’ n/a.

Table 5.18: E6 RNA expression in oropharyngeal cancer by qPCR. *Average fold change for E6 RNA expression assessed by qPCR in triplicate reactions. Samples highlighted in pink have E2 gene disruption while samples highlighted in white have an intact E2 gene.*

Tumour ID	ΔΔCT	E6 Fold change	ΔΔCT2	E6 Fold change	Average E6 fold change	Standard deviation E6	Log10(fold change) E6
330	3.2	0.1	4.4	0.0	0.1	0.0	-1.1
VU147	0	1	0	1	1	0	0
224	-4.1	17.3	-0.8	1.8	9.5	11.0	1.0
145	-4.4	20.8	-3.0	8.2	14.5	8.9	1.2
57	-5.9	61.0	0.6	0.6	30.8	42.7	1.5
450	-6.1	67.8	-4.6	24.9	46.4	30.3	1.7
400	-6.8	109.9	-5.1	34.8	72.3	53.1	1.9
451	-7.1	141.6	-3.4	10.7	76.2	92.6	1.9
537	-6.9	119.5	-6.0	66.0	92.7	37.8	2.0
142	-7.5	179.2	-3.4	10.7	94.9	119.1	2.0
48	-6.8	107.8	-6.9	120.6	114.2	9.1	2.1
418	-7.5	187.2	-6.0	63.4	125.3	87.5	2.1
321	-7.9	246.3	-7.9	245.1	245.7	0.8	2.4
42	-9.0	505.6	-6.6	99.8	302.7	287.0	2.5
19	-9.4	668.9	-5.4	42.9	355.9	442.6	2.6
185	-9.0	513.3	-7.7	204.7	359.0	218.2	2.6
141	-5.5	44.2	-9.5	738.6	391.4	491.0	2.6
55	-9.7	816.7	-6.2	73.0	444.9	525.9	2.6
491	-9.7	817.6	-8.7	418.0	617.8	282.6	2.8
449	-9.8	883.6	-8.9	478.0	680.8	286.8	2.8
39	-10.6	1597.4	-2.9	7.3	802.3	1124.4	2.9
320	-10.9	1913.8	-10.3	1258.3	1586.0	463.5	3.2
44	-11.6	3186.6	-10.0	1052.1	2119.4	1509.3	3.3
14	-12.6	6146.6	-9.2	604.5	3375.6	3918.9	3.5
488	-12.6	6155.8	-11.8	3566.3	4861.0	1831.0	3.7
22	-13.2	9239.1	-10.7	1672.9	5456.0	5350.1	3.7
350	-10.2	1152.2	-13.8	14317.5	7734.9	9309.2	3.9
490	2.8	0.1	-14.2	18799.3	9399.7	13293.0	4.0
311	-13.6	12741.8	-13.7	13444.5	13093.1	496.9	4.1
214	-12.7	6569.9	-14.7	26197.5	16383.7	13878.8	4.2
462	-21.1	2324712.1	5.1	0.0	1162356.0	1643819.6	6.1
372	-16.5	89687.6	-23.1	8842933.9	4466310.8	6189479.8	6.6
519	-25.5	46178322.6	5.7	0.0	23089161.3	32653005.0	7.4
69	-12.5	5931.1	n/a	n/a	n/a	n/a	n/a
408	-8.2	299.9	n/a	n/a	n/a	n/a	n/a
423	9.6	0.0	n/a	n/a	n/a	n/a	n/a

The high expression of E6 RNA in those samples with disrupted E2 gene is presented in Figure 5.25. This is a significant finding that confirms the oropharyngeal cancer in this cohort are transcriptionally active with dysregulation of the E6 oncogene. The samples with additional URR disruption are depicted by asterix and show either the highest E6 expression (four samples: 350, 490, 462 and 519) or lack of E6 expression (two samples: 330 and 423).

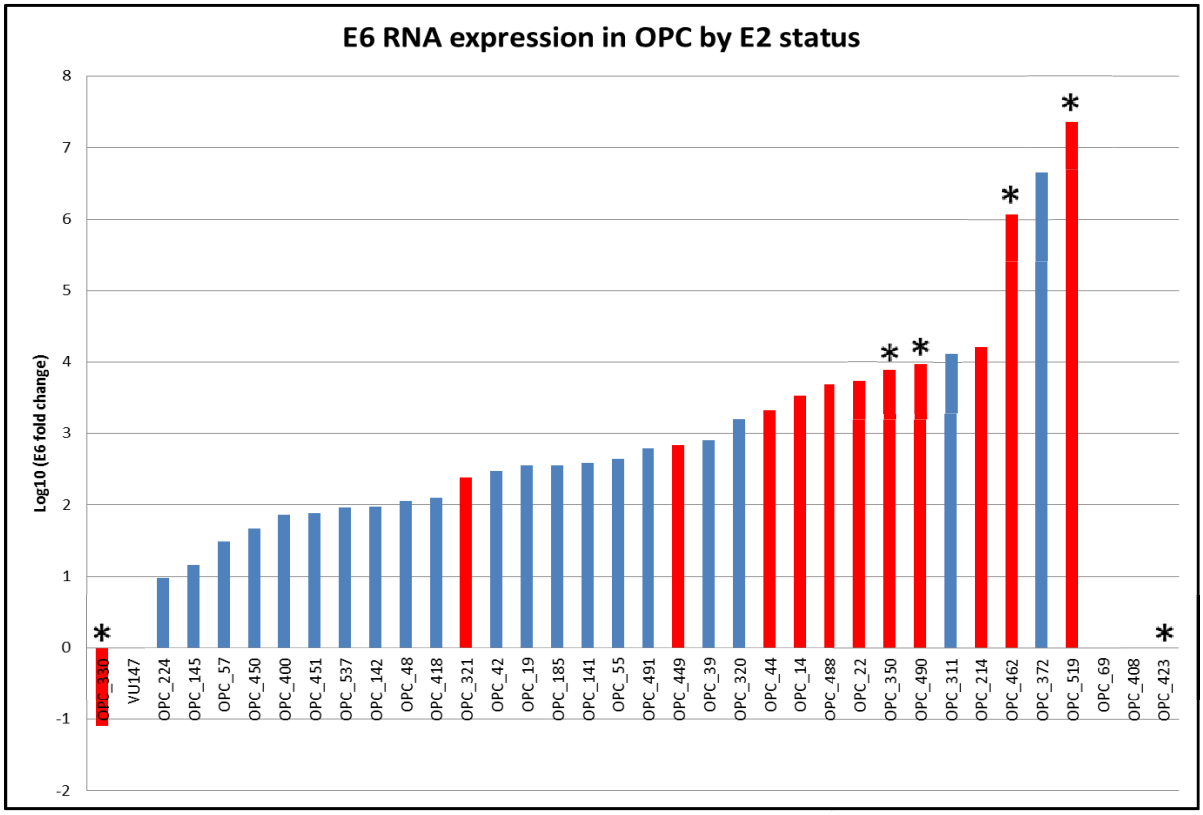


Figure 5.25: HPV 16 E6 RNA expression in OPC by q PCR *Average fold change for E6 RNA expression was assessed by real-time PCR in triplicate reactions and calculated in relation to the VU147 reference sample. OPC samples with E2 disruption are shown in red while OPC samples with an intact E2 gene are shown in blue. Samples with additional disruptions in the URR (330, 519, 462, 490, 423 and 350) are highlighted with an asterix **

5.6.11 E7 expression and E2 gene disruption:

Similar to E6 expression, there was a high expression of E7 RNA in those oropharyngeal cancer samples that also had a disrupted E2 gene (see Table 5.19 and Figure 5.26). The average of the log10(fold change) in the disrupted group was 4.7 compared to 2.4 obtained in the E2 intact group.

Table 5.19: E7 RNA expression in Oropharyngeal cancer by qPCR. *Average fold change for E7 RNA expression assessed by qPCR in triplicate reactions. Samples highlighted in pink have E2 gene disruption while samples highlighted in white have an intact E2 gene.*

Tumour ID	ΔΔCт	E7 Fold change	ΔΔCт2	E7 Fold change	Average E7 fold change	Standard deviation E7	Log10(fold change) E7
VU147	0	1	0	1	1	0	0
450	-0.9	1.9	2.6	0.2	1.0	1.2	0.0
145	-3.7	12.7	-3.1	8.3	10.5	3.1	1.0
224	-3.7	12.9	-4.1	16.8	14.8	2.8	1.2
400	-6.7	104.8	-4.5	22.3	63.5	58.4	1.8
451	-5.4	43.2	-6.9	122.2	82.7	55.8	1.9
142	-7.8	228.9	-6.0	62.9	145.9	117.4	2.2
537	-8.1	274.0	-4.5	22.7	148.4	177.7	2.2
48	-7.4	165.4	-7.1	134.0	149.7	22.2	2.2
418	-8.3	316.9	-6.1	67.4	192.1	176.5	2.3
185	-8.0	248.1	-7.5	175.3	211.7	51.5	2.3
55	-9.2	569.9	-7.7	213.7	391.8	251.9	2.6
449	-8.0	264.1	-9.6	761.5	512.8	351.7	2.7
19	-9.2	578.9	-9.6	803.0	690.9	158.5	2.8
321	-9.5	737.0	-9.7	849.9	793.5	79.8	2.9
491	-10.4	1382.2	-8.0	264.4	823.3	790.4	2.9
42	-11.4	2736.6	-6.6	99.7	1418.2	1864.6	3.2
44	-10.9	1919.0	-10.7	1686.8	1802.9	164.2	3.3
14	-11.7	3257.8	-8.9	471.6	1864.7	1970.1	3.3
39	-12.1	4425.9	-8.8	452.2	2439.0	2809.8	3.4
141	-12.1	4436.6	-8.9	487.5	2462.1	2792.5	3.4
320	-12.3	4994.0	-9.9	973.7	2983.8	2842.8	3.5
311	-11.6	3183.1	-12.1	4424.8	3804.0	878.0	3.6
22	-13.3	9871.0	-9.0	519.6	5195.3	6612.5	3.7
488	-12.7	6499.8	-12.9	7840.4	7170.1	947.9	3.9
350	-14.1	17976.5	-12.5	5786.3	11881.4	8619.8	4.1
214	-11.9	3792.3	-15.2	36653.3	20222.8	23236.2	4.3
372	-15.5	46361.1	-17.5	179578.9	112970.0	94199.2	5.1
490	-24.2	19677081.1	1.1	0.5	9838540.8	13913797.2	7.0
462	-27.7	215232424.0	-14.4	21352.8	107626888.4	152177207.9	8.0
330	-2.2	4.5	-30.1	1149663212.9	574831608.7	812934650.7	8.8
57	-4.2	19.0	n/a	n/a	n/a	n/a	n/a
69	-11.8	3584.0	n/a	n/a	n/a	n/a	n/a
408	-8.3	305.5	n/a	n/a	n/a	n/a	n/a
423	1.8	0.3	n/a	n/a	n/a	n/a	n/a
519	-9.6	767.3	n/a	n/a	n/a	n/a	n/a

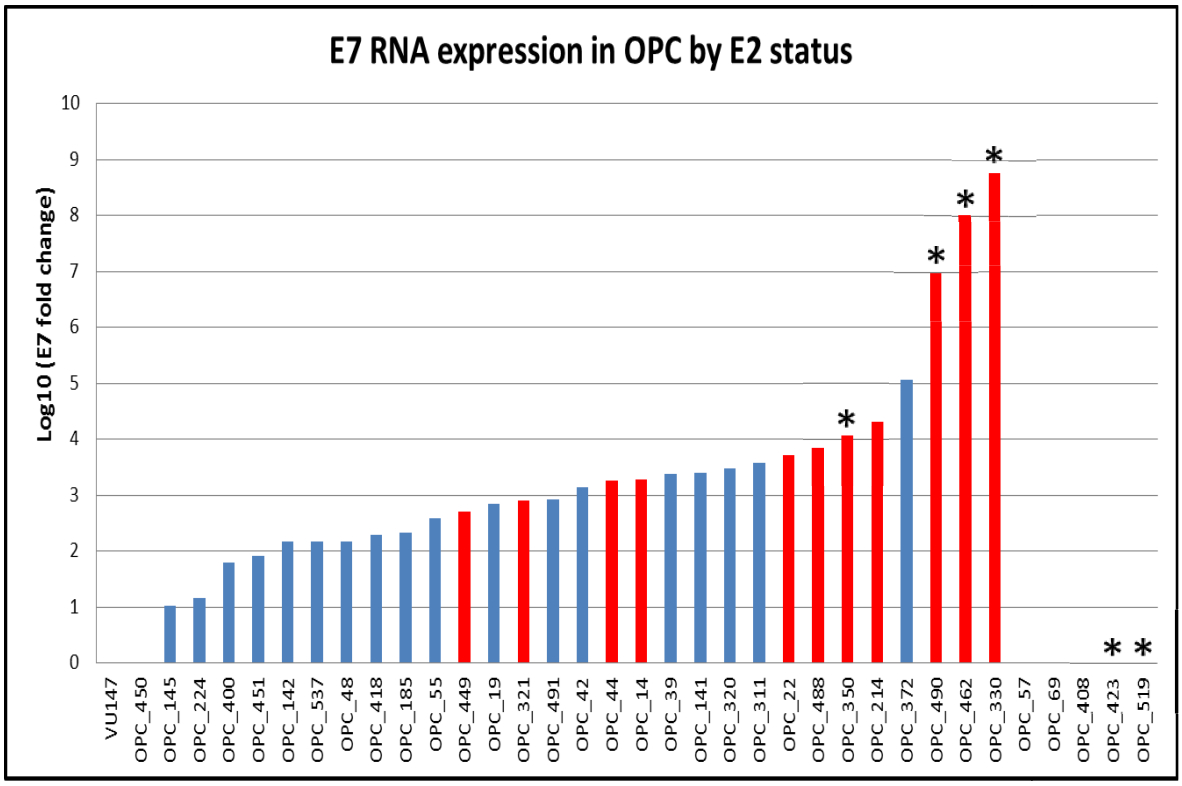


Figure 5.26: HPV 16 E7 RNA expression in OPC by q PCR *Average fold change for E7 RNA expression was assessed by real-time PCR in triplicate reactions and calculated in relation to the VU147 reference sample. OPC samples with E2 disruption are shown in red while OPC samples with an intact E2 gene are shown in blue. Samples with additional disruptions in the URR (330, 519, 462, 490, 423 and 350) are highlighted with an asterix **

The E2 disrupted cancer samples are depicted in red showing a higher proportion of samples in the E7 high expression profile in comparison to the intermediate to low expression levels demonstrated in those cancer samples with E2 gene intact depicted in blue (Figure 5.26). Similar to E6 expression, the samples with the highest E7 expression were those with additional URR disruption (350, 490, 462 and 330), the two other from this group did not express E7 (423 and 519) (marked by asterix).

Table 5.20: Summary of E2, E6, E7 expression data in OPC in relation to HPV 16 E2 gene integrity and URR methylation levels. *Comparison of average fold change for E2, E6 and E7 RNA expression assessed by qPCR in triplicate reactions, stratified by URR disruptions. Samples highlighted in red have E2 gene disruption while samples highlighted in black have an intact E2 gene.*

Tumour ID	E2 status	Average E2 fold change	Average E6 fold change	Average E7 fold change	E2BS3and4 Methylation	Mean Methylation
VU147	Intact	1	1	1	n/a	n/a
350	Disrupted	477045.6	7734.9	11881.4	URR disruption	n/a
330	Disrupted	3.9	0.1	574831608.7	URR disruption	n/a
462	Disrupted	24.1	1162356.0	107626888.4	URR disruption	n/a
408	Disrupted	288.8	n/a	n/a	n/a	n/a
488	Disrupted	5534.4	4861.0	7170.1	n/a	n/a
490	Disrupted	45.7	9399.7	9838540.8	URR disruption	n/a
423	Disrupted	40993.9	n/a	n/a	URR disruption	n/a
519	Disrupted	6.2	23089161.3	n/a	URR disruption	n/a
449	Disrupted	13.2	680.8	512.8	n/a	n/a
372	Intact	53518.7	4466310.8	112970.0	Intermediate	60.7
44	Disrupted	316.4	2119.4	1802.9	Intermediate	56.4
400	Intact	97684.7	72.3	63.5	Intermediate	48.6
69	Disrupted	2535.1	n/a	n/a	Intermediate	45.6
141	Intact	1587307.7	391.4	2462.1	Intermediate	39.5
491	Intact	133022.3	617.8	823.3	Intermediate	25.9
321	Disrupted	237.8	245.7	793.5	Low	16.8
57	Intact	12.0	30.8	n/a	Low	10.8
418	Intact	32269.2	125.3	192.1	Low	8
214	Disrupted	6550.5	16383.7	20222.8	Low	7.7
22	Disrupted	14843.8	5456.0	5195.3	Low	7.3
42	Intact	1265.2	302.7	1418.2	Low	7.2
311	Intact	37927.5	13093.1	3804.0	Low	7.2
450	Intact	31306.4	46.4	1.0	Low	7.2
48	Intact	1617.6	114.2	149.7	Low	6.6
39	Intact	23155.5	802.3	2439.0	Low	6
451	Intact	1240.9	76.2	82.7	Low	6
55	Intact	100.3	444.9	391.8	Low	5.3
19	Intact	2865.4	355.9	690.9	Low	4.5
142	Intact	134.6	94.9	145.9	Low	4.4
537	Intact	89859.3	92.7	148.4	Low	4.1
185	Intact	2.3	359.0	211.7	Low	3.1
145	Intact	527.0	14.5	10.5	Low	3.1
224	Intact	171.2	9.5	14.8	Low	3
320	Intact	55934732.1	1586.0	2983.8	Low	2.9
14	Disrupted	1113.7	3375.6	1864.7	Low	2.5

5.6.12 E2, E6 and E7 gene expression and E2 gene integrity:

Expression levels of E2 RNA were further correlated with E6 and E7 RNA expression in this cohort. There was low E2 RNA expression in those oropharyngeal cancer samples with a disrupted E2 gene which was associated with high expression levels of E6 and E7 RNA (presented as average fold change in Table 5.20 and as log10(fold change) averages in Table 5.21.

Table 5.21: Correlation of E2, E6, E7 expression in OPC by E2 gene status

Comparison of average log10(fold change) of E2, E6 and E7 RNA expression by E2 gene status. Samples highlighted in red have E2 gene disruption while samples highlighted in black have an intact E2 gene.

Tumour ID	E2 status	Log10(fold change) E2	Log10(fold change) E6	Log10(fold change) E7
VU147	Intact	0	0	0
14	Disrupted	3.0	3.5	3.3
22	Disrupted	4.2	3.7	3.7
44	Disrupted	2.5	3.3	3.3
214	Disrupted	3.8	4.2	4.3
321	Disrupted	2.4	2.4	2.9
330	Disrupted	0.6	-1.1	8.8
350	Disrupted	5.7	3.9	4.1
449	Disrupted	1.1	2.8	2.7
462	Disrupted	1.4	6.1	8.0
488	Disrupted	3.7	3.7	3.9
490	Disrupted	1.7	4.0	7.0
519	Disrupted	0.8	7.4	n/a
19	Intact	3.5	2.6	2.8
39	Intact	4.4	2.9	3.4
42	Intact	3.1	2.5	3.2
48	Intact	3.2	2.1	2.2
55	Intact	2.0	2.6	2.6
141	Intact	6.2	2.6	3.4
142	Intact	2.1	2.0	2.2
145	Intact	2.7	1.2	1.0
185	Intact	0.4	2.6	2.3
224	Intact	2.2	1.0	1.2
311	Intact	4.6	4.1	3.6
320	Intact	7.7	3.2	3.5
372	Intact	4.7	6.6	5.1
400	Intact	5.0	1.9	1.8
418	Intact	4.5	2.1	2.3
450	Intact	4.5	1.7	0.0
451	Intact	3.1	1.9	1.9
491	Intact	5.1	2.8	2.9
537	Intact	5.0	2.0	2.2
57	Intact	1.1	1.5	n/a
69	Disrupted	3.4	n/a	n/a
408	Disrupted	2.5	n/a	n/a
423	Disrupted	4.6	n/a	n/a

5.6.13 Correlation of E6 and E7 expression:

There was a strong positive correlation between E6 and E7 expression levels when the log10 (average fold change) values were correlated as shown in Figure 5.27. This was mainly driven by E6 and E7 expression levels showed a moderate positive correlation when the E2 gene was intact but no correlation when the E2 gene was disrupted (Figure 5.28).

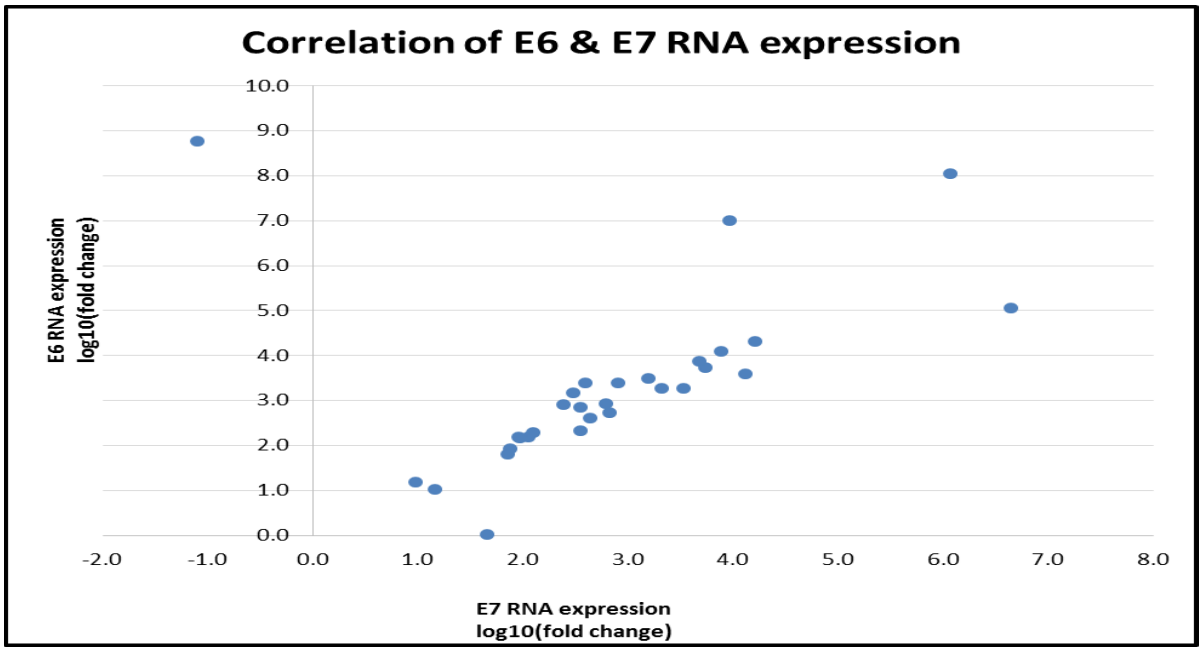


Figure 5.27: E6 and E7 expression levels show strong positive correlation

A *B*

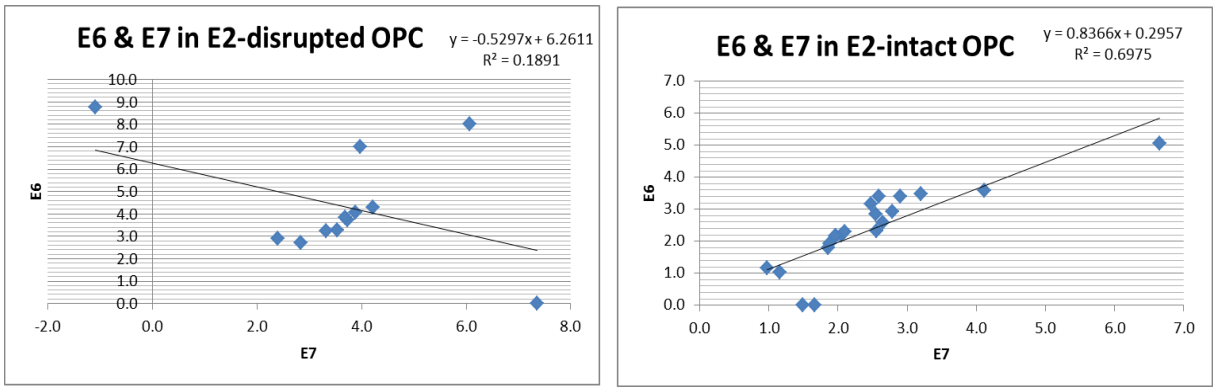


Figure 5.28 (A and B): E6 and E7 RNA expression by E2 gene status. *No correlation in E6 and E7 expression shown in E2-disrupted OPC (A) although a downward trend and moderate positive correlation between E6 and E7 expression in E2-intact OPC (B) was observed with (R^2=0.6975).*

5.6.14 Correlation of E2 expression with E6 and E7 expression:

There was no significant correlation shown between E2 expression and either E6 or E7 expression using the transformed fold change data (Table 5.21). The correlation between E2 RNA expression was further divided by E2 gene status i.e. disrupted or intact. In the E2-disrupted group, E2 expression did not correlate with E6 (R^2=0.0057) or E7 (R^2=0.0361) RNA expression (Figure 5.29 A and B).

Similarly, in the E2-intact group, E2 expression had no correlation with E6 expression (R^2=0.1207), and with E7 expression (R^2=0.2068) (Figure 5.29 C and D).

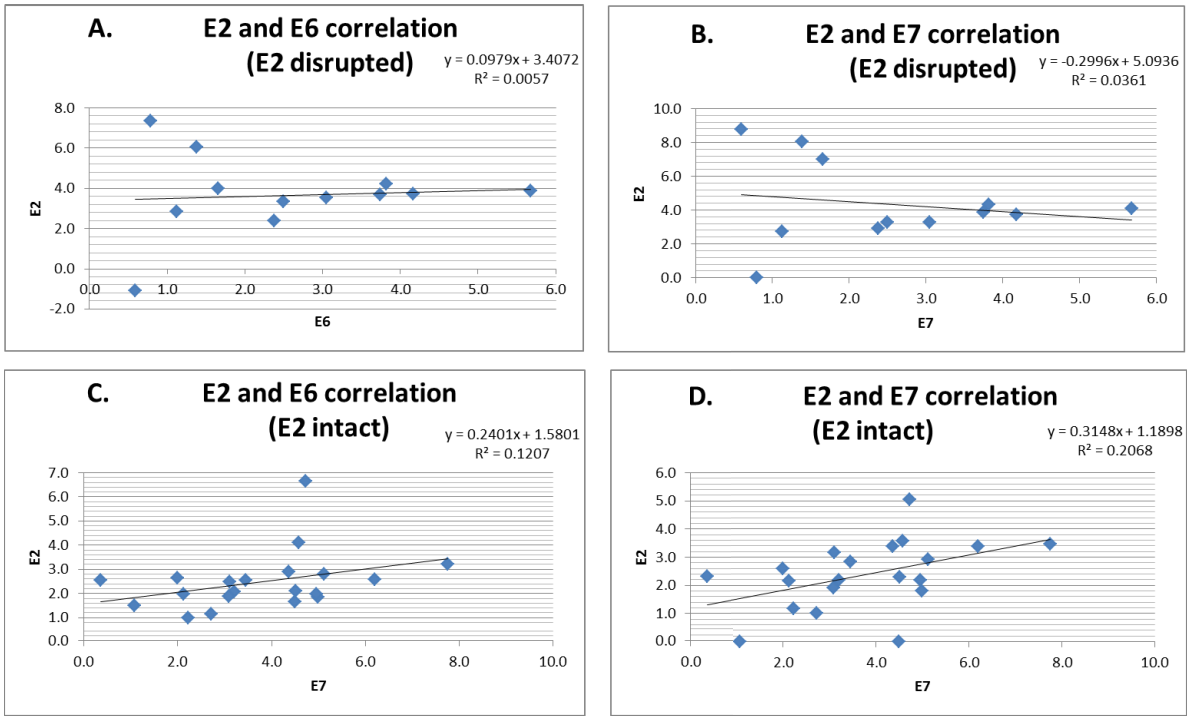


Figure 5.29: Correlation of E2 expression with E6 and E7 expression, stratified by viral E2 gene status. *No correlation of E2 expression with E6 in E2-disrupted samples (A), with E7 in E2-disrupted samples (B), with E6 in E2-intact samples (C), and with E7 in E2-intact samples (D).*

5.6.15 Methylation levels and E2, E6 and E7 RNA expression:

Methylation data was sub-divided into low methylation (<20%), intermediate methylation (≥ 20 – 80%) and high methylation (>80%) as described in the methodology. There was no sample in the high methylation group in this cohort. The analysis on methylation and E2 gene disruption has been presented in section 5.27. Methylation levels at the E2BS3 and E2BS4 inhibitory sites were correlated with E2, E6 and E7 RNA expression in 26 OPC samples with methylation data. Intermediate mean methylation was observed in six OPC samples and low mean methylation in twenty OPC samples (Table 5.22). The average of mean methylation at E2BS3 and 4 was 6.2% in the low methylation group and 46.1% in the intermediate methylation group for all samples regardless of E2 gene integrity (Table 5.22).

Table 5.22: Methylation at E2BS3 and E2BS4, E2 integrity and survival

Tumour ID	E2 status	E2BS3and4 Methylation	Mean Methylation	Survival (days)
VU147	Intact	n/a	n/a	n/a
350	Disrupted	URR disruption	n/a	1111
330	Disrupted	URR disruption	n/a	1069
462	Disrupted	URR disruption	n/a	1068
408	Disrupted	n/a	n/a	1032
488	Disrupted	n/a	n/a	723
490	Disrupted	URR disruption	n/a	714
423	Disrupted	URR disruption	n/a	688
519	Disrupted	URR disruption	n/a	657
449	Disrupted	n/a	n/a	261
372	Intact	Intermediate	60.7	718
44	Disrupted	Intermediate	56.4	1781
400	Intact	Intermediate	48.6	984
69	Disrupted	Intermediate	45.6	1132
141	Intact	Intermediate	39.5	1349
491	Intact	Intermediate	25.9	1008
321	Disrupted	Low	16.8	799
57	Intact	Low	10.8	1069
418	Intact	Low	8	952
214	Disrupted	Low	7.7	876
22	Disrupted	Low	7.3	1767
42	Intact	Low	7.2	1127
311	Intact	Low	7.2	1082
450	Intact	Low	7.2	913
48	Intact	Low	6.6	1758
39	Intact	Low	6	1998
451	Intact	Low	6	340
55	Intact	Low	5.3	317
19	Intact	Low	4.5	1760
142	Intact	Low	4.4	1679
537	Intact	Low	4.1	734
185	Intact	Low	3.1	1579
145	Intact	Low	3.1	1338
224	Intact	Low	3	1011
320	Intact	Low	2.9	1067
14	Disrupted	Low	2.5	130

The log10 (average fold change) for each early gene (E2, E6 and E7) were calculated as shown in the respective Table 5.18, Table 5.19 and Table 5.20. The means of these individual values were calculated for the URR status (URR disruption, promoter intermediate methylation and promoter hypomethylation) and compared to the E2-intact and E2-disrupted groups (Table 5.23).

Table 5.23: E2, E6 and E7 expression levels in relation to methylation, E2 status and URR disruption

Mechanism	E6	E7	E2
Intact (22)	2.3	2.3	4.0
E2 disruption (15)	3.5	3.3	2.5
E2 disruption + URR disruption (6)	3.9	5.5	1.6
E2BS3 and 4 Intermediate Methylation (6)	2.7	3.1	4.9
E2BS3 and 4 Hypomethylation (20)	2.4	2.5	3.2

The data represent the mean of the log10(average fold expression) of each of E2, E6, E7.

Oropharyngeal cancer samples with intermediate methylation at the inhibitory E2BS3 and 4 had higher E2 RNA expression (4.9) than those cancer samples with low methylation (3.2). These intermediate methylation samples also had higher mean of log10(fold change) levels in both E6 (2.7) and E7 (3.1) RNA expression in comparison to those samples with low

methylation at 2.4 and 2.5, respectively (see Table 5.23). This observation was however not clinically significant.

In OPC with intact E2 gene, there were higher levels of E2 RNA expression (4.0) when compared to those cancer samples with a disrupted E2 gene (2.5). The expression levels of E6 (3.5) and E7 (3.3) in the cancer samples with disrupted E2 gene were higher than the levels of E6 (2.3) and E7 (2.3) in the samples with an intact E2 gene. Interestingly, although the cancer samples that had both E2 gene disruption and URR disruption (N – 6) had lower E2 expression levels (1.6) when compared to the E2 intact group, there had significantly higher expression levels of both E6 (3.9) and E7 (5.5). This may be a significant carcinogenesis pathway as this group with multiple disruptions had levels of E6 and E7 expression higher than if all they had was E2 disruption (Table 5.23).

5.6.16 Impact of HPV 16 URR disruption on E2, E6 and E7 gene expression:

The URR disruptions have only been observed in samples with E2 disruption and may indicate large rearrangements of viral genome upon integration. This could lead to a loss of control mechanisms between the promoter, E2, E6 and E7 expression. Indeed, the high positive correlation between E6 and E7 expression observed in general in our cohort is not present in this subgroup. These samples are characterized by either very high expression of both E6 and E7 (OPC 350, 490, 462) or a complete loss of one (E7 in 519 and E6 in 330) or both of their expressions (423). This might or might not be accompanied by low E2 expression (see Table 5.24).

Table 5.24: E2, E6 and E7 expression in samples with URR disruption

	E6	E7	E2
350	3.9	4.1	5.7
490	4	7	1.7
462	6	8	1.4
519	7.3	0	0.8
423	0	0	4.6
330	-1.1	8.8	0.6

5.6.17 Summary:

The overview of all results obtained in achieving the stated objectives of this HPV 16 genetics and epigenetics research are demonstrated in the flowchart in Figure 5.30.

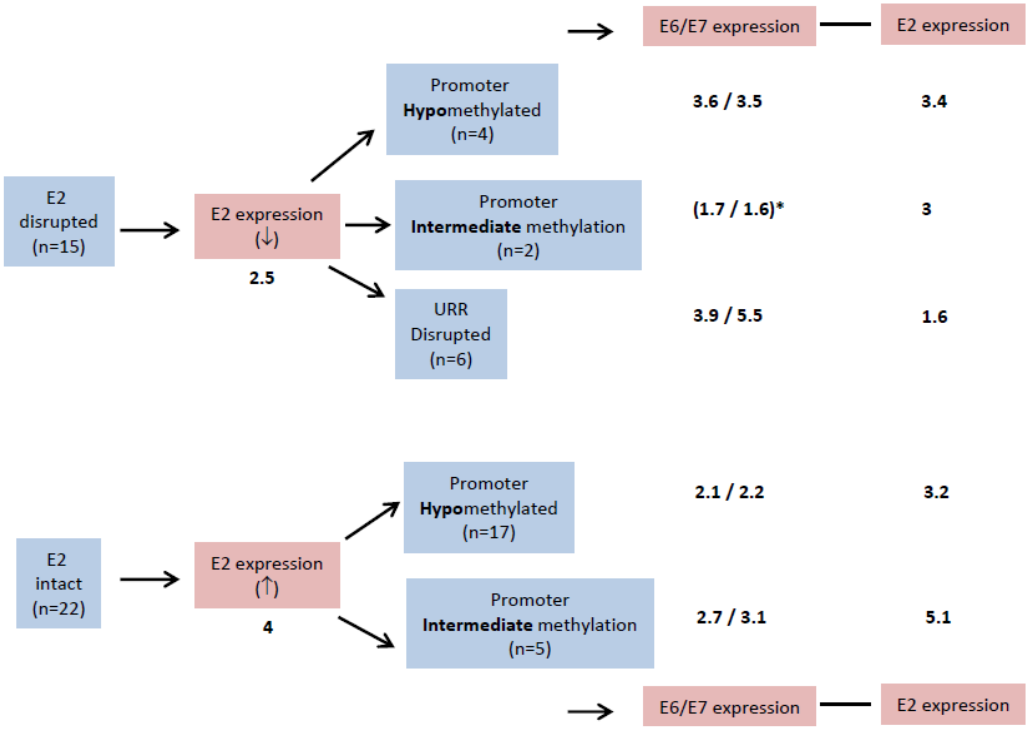


Figure 5.30: Comparison between E2, E6 and E7 expression with E2 gene status, URR status and promoter methylation. *This comparative data is descriptive and preliminary due to the small sizes of the subgroups. Although E2 expression is generally lower in the E2 disrupted group as compared to the intact group (2.5 vs. 4) it appears to be mostly due to the values in the E2 + URR disrupted samples (1.6). These anti-correlate with high expression levels of E6 and E7 (3.9 and 5.5). The other disrupted samples, regardless of methylation, have similar E2 expression (3.4 and 3) to the intact group with hypomethylated promoter (3.2). The highest E2 expression (5.1) is observed in the E2-intact group with intermediate promoter methylation. E6 and E7 expression levels are generally higher in the E2-disrupted group as compared to the E2-intact group. (*The result for the E2-disrupted with intermediate promoter methylation – only two samples with values 3.3 and 0.0 for both E6 and E7). Within the E2-intact group, intermediate promoter methylation group is characterized by slightly higher E6/E7 expression and, as mentioned before, the highest E2 expression.*

5.7 Discussion:

Human papillomavirus-mediated carcinogenesis as it relates to a subset of Oropharyngeal cancer, remains poorly understood. Viral acquisition (especially of HPV 16 genotype), integration into host tissues, replication and ultimately genomic dysregulation of host defence mechanisms, are the focus of current scientific scrutiny worldwide. The molecular processes that underpin the development of cervical cancer from persistent infection with high-risk HPV are the current template from which our understanding of HPV-related interactions are built at this point in time. The tendency to extrapolate both laboratory and clinical facts that have been established since zur Hausen (2009) conclusively demonstrated high-risk HPV as an aetiology for cervical cancer is understandable. However, there are some key differences emerging between HPV in cervical cancer and HPV in oropharyngeal cancer. There exists limited data at present addressing pertinent questions relating to the HPV-mediated mechanisms of oropharyngeal carcinogenesis including both genomic and epigenetic events.

There is currently an agreement that in order to control the host's cell cycle mechanisms the expression of viral oncogenes E6 and E7 has to be maintained at high levels. How this is achieved is still under debate, especially in oropharyngeal carcinogenesis. This study contributes in some degree to the on-going discourse on HPV 16 – mediated oropharyngeal cancer by investigating the role of disruption of the viral regulatory E2 gene and the importance of aberrant methylation in the homeostasis of viral E6 and E7 proto-oncogenes. In highlighting what we already know from cervical cancer literature, the viral regulatory E2 gene is invariably disrupted in approximately 70% of cervical cancers (Chaturvedi et al., 2013). The reverse appears to be the case in HPV-oropharyngeal cancer as the limited literature indicates viral intergration occurs in 30 – 40% of tumours with the majority of tumours being episomally driven. Recent studies also confirm that viral integration into the

host DNA is accompanied by the presence of multiple disruptions in the HPV 16 genome itself including E1, E2, E4, E5, and L1 (Wilson et al., 2013; Gao et al., 2013). The data obtained in the presented study, although preliminary due to the small group size, led to interesting observations discusses below.

Physical status of the HPV 16 regulatory E2 gene in oropharyngeal cancer:

This study evaluated the integrity of the HPV 16 regulatory E2 gene (intact or disrupted) in HPV positive-oropharyngeal cancer using type-specific PCR with overlapping primers to the HPV 16 E2 open reading frame. 16/40 (40%) of the HPV positive-oropharyngeal cancer samples had HPV 16 E2 gene disruption on PCR. This finding is consistent with the 16 – 43% HPV integration in oropharyngeal cancer published in PCR-based studies that included more than 10 cancer samples (Vojtechova et al., 2015; Reuschenbach et al., 2015; Olthof et al., 2014; Barbieri et al., 2014; Park et al., 2011). This further adds weight to the observation that HPV-oropharyngeal cancer is mainly an episomally-driven tumour in contrast to cervical cancer with significantly higher viral integration events. The HPV 16 E2 gene is therefore intact but dysfunctional in maintaining viral homeostasis of E6 and E7 oncogenes. The hypothesis in this study was that methylation is responsible for the URR dysfunction and as such differential methylation profiles should be observable at the palindromic E2 binding sites in the HPV upstream regulatory region (URR), based on the status of the viral E2 gene.

The main limitation of the technique used to assess viral physical status in this study i.e. overlapping primers aimed at the HPV 16 E2 ORF is that viral regulation is poly-cistronic and although disruptions occur predominantly in the viral E2 gene, integration is seen in multiple sites in the genome as well (Wilson et al. 2013; Gao et al., 2013). Although the absence of a band on electrophoresis is indicative of disruption in at least the viral E2 gene, the presence of

a band cannot conclusively exclude the presence of viral integrations in other genomic sites or mixed infections.

The scientific inquiry into the mechanisms that lead to over-expression of E6 and E7 viral genes and ultimately cancer in the two independent groups (E2 intact and E2 disrupted) of oropharyngeal cancers is valid and requires ascertainment. The presence of HPV in OPC is a good prognostic marker as these tumours (in comparison to their HPV negative counterparts), tend to respond better to chemo radiotherapy or surgical treatments. These viral genomic differences could therefore be additional biomarkers that provide further prognostic parameters in addition to the current prognostic marker that is the presence of HPV in oropharyngeal cancer.

Differential methylation at the E2 binding sites of the HPV 16 upstream regulatory region:

In this study, methylation profiles were further determined at each transcriptionally-relevant E2 binding site (E2BS1, E2BS3 and E2BS4) in the HPV 16 URR, using pyrosequencing of bisulfite-modified DNA from HPV positive OPCs. The first key finding from the pyrosequencing experiments in the HPV 16 URR was the overall hypomethylation state of the HPV 16 URR. The methylation levels in all of the E2 binding sites were generally less than or around 20% regardless of the viral E2 status (intact or disrupted). This in itself is a potentially useful biomarker in HPV oropharyngeal cancer as it has been previously reported by Park et al. (2011) and Balderas-Loaeza et al. (2007). Both studies indicate the HPV 16 genome is hypomethylated in cases of advanced HPV positive head and neck cancer with further preferential hypomethylated noted in the long control region (LCR) of the viral genome.

There is however a contrasting information as to methylation at specific E2BSs. A recent paper by Barbieri et al (2014) on 60 HPV positive oropharyngeal cancer using bisulfite-

modified DNA did however show higher methylation frequency in the E2BS1 when compared to E2BS3 and 4. This is in contrast to the findings in both our study and the study by Reuschenbach et al. (2015) where higher methylation in the E2BS3 and 4 was observed when compared with the E2BS1.

The marginal differential methylation observed in our study between the promoter-stimulating E2BS1 and the inhibitory E2BS3 and E2BS4, is of uncertain clinical significance, mainly due to the small group sizes. However, methylation profiles published in recent studies in head and neck literature points to the existence of a certain dichotomy in the E2 binding sites with correlation to cancer survival. Reuschenbach et al. (2015) were able to stratify a cohort of 57 HPV positive oropharyngeal cancer into three subgroups based on their differential methylation at the E2BS3 and E2BS4. High methylation frequencies at these inhibitory E2 binding sites (E2BS3 and E2BS4) were shown to be associated with a worse overall survival (Reuschenbach et al., 2015).

It is of importance that 10 HPV positive oropharyngeal cancer in this cohort (25%) failed PCR amplification prior to pyrosequencing largely because of the concomitant presence of disruptions in the HPV 16 URR itself that houses the E2 binding sites. These samples were excluded from the methylation results with an impact on the overall study sample size. However, at the same time they emphasized the role of URR disruption in OPC carcinogenesis as discussed later.

Impact of viral E2 gene status on expression of E2, E6 and E7 mRNA using qPCR:

This study sought to confirm the proportion of the HPV oropharyngeal cancer samples that were transcriptionally active with expression of E2, E6 and E7 mRNA using reverse-transcription real-time PCR. It is well-recognised that the mere presence of HPV DNA in cancer samples is not confirmation of cancer causation. In addition to HPV DNA positivity by PCR and p16 immunohistochemistry, high expression levels of E6 and E7 is required to demonstrate tumours are transcriptionally active. All the HPV OPC samples used in this study expressed E2 mRNA or its isoforms ('E2 880^2709' and 'E2 880^2582') although at highly variable levels. In fact, integration/E2 disruption accompanied with no E2 expression is very rare (Reuschenbach et al., 2015) pointing towards the co-presence of both intact/episomal and disrupted viral genomes. Similarly the levels of E6 and E7 expression were variable but the expression was not detected in three and five samples for E6 and E7, respectively.

The viral gene expression was then compared to the methylation levels at the E2BS3 and E2BS4 sites. Unfortunately due to the small group sizes the statistical analysis was not possible and no firm conclusions could be made.

E2 RNA expression:

E2 expression was higher when the E2 gene was intact when compared to its expression levels when the E2 gene was disrupted. This is a recognised pattern of E2 expression dependent on E2 gene integrity (Reuschenbach et al., 2015) and our data reaffirms this mechanism in HPV carcinogenesis. In addition, E2 expression was the lowest when both E2 and the HPV 16 URR were disrupted in the same sample. This is a unique finding that merits further investigation.

In terms of E2 expression and methylation, this study showed that intermediate methylation was linked to higher E2 expression levels when compared with samples that had low methylation and lower E2 expression. It is therefore possible that methylation affects E2 expression or vice versa.

E6 and E7 RNA expression:

E6 RNA expression was lower when E2 gene was intact. This is in keeping with the understood mechanisms of HPV auto-regulation described by Reuschenbach et al. (2015). Interestingly, E6 expression was highest when both the viral E2 gene and the URR were disrupted. E6 expression was similar in both low and intermediate methylation at the inhibitory E2BS3 and E2BS4 but highest in URR disruption.

Similarly, E7 RNA expression was lower when the E2 gene was intact compared to when the E2 gene was disrupted. E7 expression was also highest when both the regulatory E2 gene and the URR were disrupted in the same cancer sample.

In terms of methylation, E7 expression was higher in the presence of intermediate methylation when compared with samples that had low methylation. E7 expression was however highest when URR disruption was also present. This suggests that URR disruption tends to accentuate the dysregulation of E6 and E7 with their further over-expression. However, there was no statistically demonstrable significance of this observation given the small sample size so these conclusions still require further investigation.

In addition, E6 and E7 RNA expression levels showed a good positive correlation especially in the E2-intact cancers. E6 and E7 expression levels were generally higher in the E2-disrupted group as compared to the E2-intact group. Within the E2-intact group, intermediate methylation at E2BS3 and 4 was characterized by slightly higher E6/E7 expression and as

mentioned before the highest E2 expression. It is possible the decrease of promoter inhibition leads to increased levels of all E6, E7 and E2 genes, therefore the auto-regulatory element of viral transcription could still be maintained in these samples. Therefore, in OPC high E6/E7 levels result primarily from the disruption of E2 gene, especially when accompanied by URR disruption. While, when the E2 gene remains intact, increased methylation at HPV 16 promoter inhibitory E2 binding sites could lead to increased E6 and E7 expression despite high E2 expression levels.

HPV 16 URR disruptions:

This study highlights the possibility that URR disruption could be another mechanism of HPV carcinogenesis. Viral integration has been shown to occur elsewhere in the HPV genome aside just the E2 gene based on whole genome sequencing (Wilson et al., 2013). Specifically, the URR disruptions were found in the presence of E2 disruption but intact E6, E7 and L1 genomic domains. In our cohort, the URR disruptions have only been observed in samples with E2 disruption and may indicate large rearrangements of viral genome upon integration. This could lead to a loss of control mechanisms between the promoter, E2, E6 and E7 expression. Indeed, the high positive correlation between E6 and E7 expression observed in general in our cohort is not present in this subgroup. Specifically, two samples had a very high expression of one of the oncogenes with the other one remaining silent.
This data underline the importance of the URR in OPC carcinogenesis and the fact that it can be affected either by physical disruption or by methylation.

Viral E2 status and E2 binding sites' methylation profiles correlation with clinical factors and overall survival in oropharyngeal cancer:

HPV 16 E2 physical status (intact or disrupted), the methylation profiles in each E2 binding site, and the RNA expression data were correlated with epidemiologic and clinical outcomes such as demographic variables, risk factors' data, Oropharyngeal cancer recurrence and overall survival from HPV positive Oropharyngeal cancer. The key findings include:

a. There was no statistically significant difference between the tonsil and tongue base anatomic sites when considering E2 gene disruption.

b. The observed gender difference with females having higher proportions of OPC with HPV E2 gene disruption is however not statistically significant given the small numbers of females in the analysis. This merits further investigation.

c. On average, patients who had tumours with intact HPV were approximately 6 years younger than those patients found to have disrupted HPV. Increasing age is therefore an independent univariate predictor for HPV 16 E2 gene disruption in OPC.

With increasing mean age, there appears to be a tendency towards increasing HPV 16 E2 gene disruption in OPC, with this age effect more marked amongst females.

d. There was no statistically significant different between tumour size categories and HPV E2 gene disruption

e. In this cohort, smoking had a strong assocation with OPC (70% smokers) consistent with current literature but smoking did not show a statistically significant impact on viral E2 status in OPC when stratified into current smoker, ex-smoker or non-smoker.

f. The Kaplan-Meier probabilities show a trend of worse survival at 1, 3 and 5 years following diagnosis for oropharyngeal cancer in the cohort with disrupted HPV E2 gene but this trend was not statistically significant. Significantly, Oropharyngeal cancer recurrence was associated with an increased likelihood of mortality on Kaplan-Meier survival analyses.

Due to the highly decreased group sizes, it was not possible to correlate epidemiological and clinical data with the URR methylation and viral gene expression status.

In summary, this study observes the following potential mechanisms of HPV carcinogenesis in OPC:

1. E2 disruption
2. URR methylation – data inconclusive in this cohort however other studies support this view
3. URR disruption – more likely to be connected with E2 disruption and multiple sites of viral integration. This is on a background of intact E6, E7, and L1 regions.

PhD CHAPTER 6:

CONCLUSIONS AND FUTURE DIRECTIONS

CONCLUSIONS:

This thesis on the epidemiology and epigenetics of oral HPV infection and oropharyngeal cancer has been executed utilising a systematic review and meta-analysis, a prospective cohort study, a retrospective case series as well as a molecular study focused on HPV 16 genome and epigenome. These interwoven strands of research have provided some new insights into our understanding of the natural history of oral HPV infection and highlighted observations on potential mechanisms of HPV carcinogenesis in the oropharynx. The findings have corroborated the current limited literature on oral HPV natural history and provided a window into future directions for epigenetic research into HPV-associated oropharyngeal cancer.

6.1 Systematic review and meta-analysis:

The systematic review and meta-analysis has provided a comprehensive analysis of the literature on the epidemiology of oral human papillomavirus infection in healthy adults. It corroborates the most recent systematic review on oral HPV infection in healthy populations published by Kreimer et al. (2010) in demonstrating a low prevalence of oral HPV infection, but with a seven fold population size.

Kreimer et al. (2010) published a pooled prevalence of 4.5% (95% C.I: 3.9% – 5.1%) for any oral HPV, 3.5% (95% C.I: 3.0% - 4.1%) for oral high-risk HPV, and 1.3% (95% C.I: 1.0% - 1.7%) for oral HPV 16 (Kreimer et al., 2010). In comparison, this thesis reiterates pooled prevalence in healthy individuals worldwide for any oral HPV of 4.9% (95% CI: 3.7% - 6.3%); oral high-risk HPV of 1.8% (95% CI: 1.2% - 2.6%); oral HPV 16 of 0.7% (95% CI: 0.4% - 1.0%). This thesis adds to the literature because it excludes population groups that are considered ‘high-risk’ such as sex workers and attendants at sexually transmitted diseases’

clinics, which had not been excluded in Kreimer's study. The study population size pooled from worldwide papers in this thesis is approximately 7 times more than the one included in Kreimer's review (28,516 participants compared to 4,070). The combination of a higher powered study with exclusion of 'high-risk' populations, provides a more robust evaluation of oral HPV prevalence in healthy adults provided by this thesis. This is demonstrated by the lower pooled prevalence for oral high-risk HPV in this thesis.

The meta-analysis in this thesis also did not identify any differences in the pooled prevalence of oral HPV by gender. Males and females had similar prevalence rates amongst healthy adults which corroborates the reported prevalence rates by sex in the Kreimer et al. (2010) systematic review. The implications of this in clinical practice and the development of oropharyngeal cancer requires further investigation because men are more likely to present with HPV oropharyngeal cancer despite similar prevalence rates of oral infection when compared with females. The factors responsible for the susceptibility of males to HPV oropharyngeal cancer is unclear and as such future research is required into exposure to risk factors, hormonal differences, dietary factors, or genetic predispositions.

Additionally, this thesis identified and reported separately from healthy individuals, oral HPV infection rates specifically in populations at-risk of higher oral HPV prevalence including individuals with immunodeficiency states such as infection with HIV, sex workers, and women with cervical HPV infection. These findings are corroborated by current literature (Matsushita et al., 2011; Beachler et al., 2012).

Our meta-analysis observed geographic variation in oral HPV prevalence. The interpretation of this finding should be done cautiously given the heterogeneity inherent in the pooled studies with the variability in HPV testing across world regions. The majority of the included

studies had some PCR-based testing algorithms but the techniques are not standardised and as such lack uniformity. Some of the HPV testing techniques are more sensitive than others and this may exaggerate the variations observed in world regions. Notwithstanding this limitation, there are socio-cultural complexities in each world region that may either predispose or preclude high oral HPV prevalence. In Asia for example, there were only two studies included with high oral HPV rates shown in Pakistani men. Gichki et al., (2012) reported the highest prevalence for oral HPV of 24.5% (95% C.I: 18.9 – 31.0%) in 200 Pakistani men. There are also limited studies from Africa and as such the reported continental rates may be sub-optimal.

Studies have shown that oral HPV infection is sexually transmitted. This means prevalence could be in part dependent on the sexual practices inherent in each world region. Recent publications show the rising incidence of HPV-associated oropharyngeal cancer to be in the Western world. However, HPV may not necessarily be a driver for oropharyngeal cancer in regions such as Africa or Asia due to potentially different sexual practices and attitudes. More qualitative and quantitative research is needed to establish true geographic differences in both oral HPV prevalence in healthy adults, regional sexual practices and ultimately HPV oropharyngeal cancer development.

This meta-analysis also demonstrates variation in the reported oral HPV prevalence rates depending on the sample taken for analysis. HPV detection was highest in studies using

scrapings / brushings / swabs at 5.9%; followed by studies using oral rinse / gargle / saliva at 3.5%; and the lowest in tonsils at 1.0%. This is useful information that can guide the development of population screening strategies for oral HPV. It was surprising to observe low HPV infection worldwide in tonsil specimens given that Begum et al. (2005) reported HPV

localisation in the reticulated epithelium of tonsil crypts. Our finding of 0% in FFPE and 1.0% oral HPV prevalence in normal tonsils in a UK population is corroborated by Palmer et al. (2014) in the single largest study on FFPE specimens and non-malignant tonsils in the UK. Palmer et al. (2014) found 0% in 3,377 FFPE and 0% in 511 tonsils. Demonstrating how oral HPV infection progresses from acquisition at sexual exposure to eventual HPV-oropharyngeal cancer thus remains a subject of future research in a longitudinal study which could last years.

6.2 Prospective cohort study:

An important research question assessed in Chapter 3 of this thesis is the potential impact (if any) of tonsillectomy on the natural history of oral HPV infection. This cohort sub-study is part of a larger multicentre prospective UK study evaluating the prevalence of oral HPV infection in a tonsillectomy population. This is the first UK study to provide crucial data on the hypothesis that surgical removal of tonsils results in a reduction in oral HPV prevalence or acquisition.

Tonsillectomy did show a trend towards abrogation of oral HPV infection with HPV positivity in the pre- and post-tonsillectomy oral rinse samples being 11% and 2% respectively. However, this observation was not statistically significant due to the acquisition of new oral HPV infections in 9% of participants post-tonsillectomy who were previously negative before surgery. Therefore, tonsillectomy in itself did not appear to prevent oral HPV infection but could potentially be useful in reducing oral HPV burden. A limitation of this cohort study was the lack of control of confounding factors on oral HPV acquisition such as sexual activity, change in sexual relationships or behaviours between the pre- and post-tonsillectomy samples. This could have been accounted for by sexual history questionnaires which were administered prior to the study recruitment but not at the post-operative sample

collections. The second limitation was the lack of a control group to adequately assess if tonsillectomy was indeed an intervention that reduced oral HPV infection or if this effect was due to chance and the natural cycle in immune-competent individuals. The Finnish family study has reported that oral HPV infection is cleared within 18 – 24 months in healthy individuals. A control group that were age and sex-matched who did not undergo tonsillectomy would have been useful in interpreting the data if they had oral rinse samples taken at two separate time points similar to the tonsillectomy group.

The findings from this cohort study need validation in a future prospective study that is sufficiently powered to establish an association between tonsillectomy and reduced oral HPV infection.

6.3 Retrospective UK population study:

This strand of the thesis triangulated data on oropharyngeal cancer registered on the UK Hospital Episodes Statistics (HES) database and UK population-level data obtained from the Office of National Statistics (ONS) to describe the epidemiology, incidence and determinants of survival in patients diagnosed with oropharyngeal cancer. Kaplan-Meier survival and Cox-regression analyses were performed with determinants of survival from oropharyngeal cancer shown to include age, gender, socioeconomic status, deprivation levels, and presence of co-morbidities and smoking. These survival determinants are consistent with published literature (Granata et al., 2012). HPV positivity is known to be a significant prognostic classifier in oropharyngeal cancer (Ang et al., 2010) and the clinical significance in squamous carcinomas of the head and neck in the UK is also documented (Heath et al., 2012) but this could not be studied conclusively in this thesis because of limited data available on HPV status on the HES

database. In future, HPV status will become more readily available as cancer networks across the UK now include this information in their data submissions.

6.4 Genomics and epigenetics study:

There is a need for additional diagnostic and prognostic biomarkers for OPC alongside HPV positivity. Disruption of the HPV regulatory E2 gene and methylation of the E2 binding sites in cervical cancer have been identified as potential prognostic mechanisms driven by HPV in cervical cancers. In contrast, there is currently limited data available on the genomic changes and epigenetic modifications that occur as a prelude to HPV-mediated oropharyngeal cancer. This section of the thesis investigated the frequency of occurrence of viral E2 gene disruption and the role of aberrant methylation in HPV mediated OPC, as well as and the impact of these on clinical outcomes. Firstly, this thesis demonstrates 40% viral E2 gene disruption in a cohort of HPV positive oropharyngeal cancer. This is similar to reported rates of viral E2 gene disruption in the limited literature available on oropharyngeal cancer but in contrast to the higher rates reported in cervical cancers. The limitation in this study was inherent in the technique utilised for determining E2 gene disruption and by extension HPV integration. The methodology is limited to physical disruption within the E2 gene sequence which does not exclude the presence of viral integrations occurring elsewhere in the HPV genome that could also affect viral transcriptional regulation or the presence of mixed infections. In an attempt to address the limitation, the study was extended by the analysis of the E6, E7, L1 and URR regions of the HPV genome which resulted in interesting observations.

This study also showed that E2 gene disruption (as evidenced by an absence of a band on agarose gel electrophoresis) was associated with oropharyngeal cancer recurrence in this cohort. This finding requires further validation in a larger study.

The key finding of the epigenetic analysis was the general hypomethylation at the HPV 16 URR. Further analysis of the methylation profiles at the E2 binding sites in the HPV 16 URR, revealed no statistically significant differences between E2-intact and E2-disrupted HPV OPC. There were some observations made in this study that merit further scrutiny. Firstly, in the E2-disrupted oropharyngeal cancer, there were lower E2 mRNA expression levels associated with higher E6 and E7 expression levels. Secondly, disruptions in the HPV URR were observed especially among the E2-disrupted oropharyngeal cancer cohort. This observation could potentially be an additional mechanism of carcinogenesis but it requires further evaluation.

HPV 16 DNA methylation has been shown to be a prognostic biomarker in cervical cancer. In Oropharyngeal cancer, hypermethylation of the inhibitory E2 binding sites in the HPV URR has been demonstrated by Reuschenbach et al. (2015) to confer worse survival. This requires further investigation in oropharyngeal cohorts and could lead to the development of novel clinical strategies and applications.

6.5 Future perspectives on oral HPV:

Epidemiologic and laboratory evidence have now conclusively shown that high-risk HPV genotypes are causative for cervical cancer, a fraction of cancers of the vagina, vulva, anus, penis and more recently a subset of oropharyngeal cancer. Persistent infection with high-risk HPV genotypes in the transitional epithelium of the cervix is demonstrably the precursor to cervical carcinogenesis (Munoz et al., 2006). It is unclear if this confirmed pathway of

persistent infection with high-risk HPV prior to ano-genital cancer development occurs similarly in cancers of the oropharynx. Such an assertion at this stage represents a huge leap in the dark in terms of HPV oropharyngeal cancer, for a number of reasons that require scientific scrutiny. Firstly, tonsil tissue and tongue base tissue are lymphoid tissue and are pre-eminently different from the cervical tissue from whence cervical cancer arises – namely the squamo-columnar junction. Secondly, there exists pre-malignant states in cervical cancer (cervical intraepithelial neoplasia I, II, III) which allow disease progression to be monitored and an early diagnosis made from biopsies. This is not the case in the oropharynx as there is no recognised 'premalignant state' for oropharyngeal cancer. Thirdly, the temporal trend that underlines progression from HPV infection in the cervix (or its detection) to cervical cancer development ranges from 10 years to 30 years (depending on the individual patient factors, exposure to other risk factors and the HPV genotype) with ≤50% of cervical intraepithelial neoplasia III (CIN3) transforming to invasive cervical cancer (McCredie et al., 2008). The time from the acquisition of HPV infection in the oral cavity / oropharynx, latency or persistent infection and ultimately cancer development is currently unknown as are all the suggested stages of progression or reinfection. In immune-competent individuals, the limited data indicates that HPV infection in the oral cavity is a dynamic process in which the person(s) affected are able to clear the infection within a period of 12 – 24 months. There is also a bi-modal peak described for oral HPV infection by (Gillison et al., 2012) with the first peak in individuals aged 20 – 29 years and the second peak in adults 50 – 54years. There are currently two schools of thought on the relatedness of these two peaks of oral HPV infection: either they are independent peaks and the second peak represents a re-infection or they are related peaks with the second peak simply being a reactivation of latent oral HPV infection after 20 – 30 years. These are largely unexplored areas in our understanding of oral HPV

carcinogenesis that will require prospective longitudinal studies to definitely answer - albeit predictably time and money-consuming.

In addition to persistent HPV infection in the aetiology of cervical cancer, there are other recognised risk factors that include multiparity, multiple sexual partners, use of hormonal contraceptives, smoking, immunosuppression and co-infection with other viruses such as chlamydia trachomatis and herpes simplex (Munoz et al., 2006). In Oropharyngeal cancer, the risk factors primarily include the synergistic effect of tobacco smoking and excessive alcohol consumption alongside genetic predisposition and HPV involvement. Research into risk factors for OPC that can further stratify into prognostic groups is required including epidemiologic, epigenetic, hormonal and dietary factors. Viral factors including viral load, viral integration, and the HPV genotype involved are all areas with limited data currently as regards oropharyngeal cancer development. Host factors such as susceptibility to genomic instability, exposure to risk factors and immune states can also contribute to the severity and progression of oropharyngeal cancer, and merit further scientific scrutiny.

REFERENCES:

Amin, M. B. Edge, S., Greene, F., Byrd, D. R., Brookland, R. K., Washington, M. K., … Meyer, L. R. (2017). Oropharyngeal cancer staging. *AJCC Cancer staging manual*. Springer International Publishing.

Ang, K. K., Harris, J., Wheeler, R., Weber, R., Rosenthal, D. I., Nguyen-Tan, P. F., . . . Gillison, M. L. (2010). Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med, 363*(1), 24-35. doi:10.1056/NEJMoa0912217

Antonsson, A., Cornford, M., Perry, S., Davis, M., Dunne, M. P., & Whiteman, D. C. (2014). Prevalence and risk factors for oral HPV infection in young Australians. *PLoS One, 9*(3), e91761. doi:10.1371/journal.pone.0091761

Araujo, M. V. D., Pinheiro, H. H. C., Pinheiro, J. D. V., Quaresma, J. A. S., Fuzii, H. T., & Medeiros, R. C. (2014). Prevalence of human papillomavirus (HPV) in Belem, Para State, Brazil, in the oral cavity of individuals without clinically diagnosable injuries. *Cad Saude Publica, 30*(5), 1115-1119. doi:10.1590/0102-311x00138513

Arbabi-Kalati, F., Nosratzehi, T., Bameri, Z., & Rigi, F. M. (2014). Detection of Salivary Human Papilloma Viruses 16 and 18 (HPV) in Smoker Men in an Iranian Population by PCR: A Pilot Study. *Int J High Risk Behav Addict, 3*(3), e18053. doi:10.5812/ijhrba.18053

Arias-Pulido, H., Peyton, C. L., Joste, N. E., Vargas, H. & Wheeler, C. M. (2006). Human papillomavirus type 16 integration in cervical carcinoma in situ and in invasive cervical cancer. *J Clin Microbiol, 44*, 1755-1762.

Barbieri, D., Nebiaj, A., Strammiello, R., Agosti, R., Sciascia, S., Gallinella, G., . . . Venturoli, S. (2014). Detection of HPV16 African variants and quantitative analysis of viral DNA methylation in oropharyngeal squamous cell carcinomas. *J Clin Virol, 60*(3), 243-249. doi:10.1016/j.jcv.2014.04.014

Barbosa, M. S., & Schlegel, R. (1989). The E6 and E7 genes of HPV-18 are sufficient for inducing two-stage in vitro transformation of human keratinocytes. *Oncogene, 4* (12):1529-32.

Baseman, J. G., & Koutsky, L. A. (2005). The epidemiology of human papillomavirus infections. *J Clinical Virol, 32*, 16-24.

Beachler, D. C., Weber, K. M., Margolick, J. B., Strickler, H. D., Cranston, R. D., Burk, R. D., . . . D'Souza, G. (2012). Risk Factors for Oral HPV Infection among a High Prevalence Population of HIV-Positive and At-Risk HIV-Negative Adults. *Cancer Epidem Biomar, 21*(1), 122-133. doi:10.1158/1055-9965.Epi-11-0734

Begum, S., Cao, D., Gillison, M., Zahurak, M., & Westra, W. H. (2005). Tissue distribution of human papillomavirus 16 DNA integration in patients with tonsillar carcinoma. *Clin Cancer Res, 11*(16), 5694-5699. doi:10.1158/1078-0432.CCR-05-0587

Bekker, J. B., Evans, M. F., Threlkeld, K. J., Rajendran, V., Adamson, C. S., & Cooper, K. (2012). Screening for HPV in Clinically Benign Tonsillectomy Specimens. *Modern Pathol, 25*, 305a-305a.

Bennett, K.L, Lee, W., Lamarre, E., Zhang, X., Seth, R., Scharpf, J., Hunt, J., & Eng, C. (2010). HPV status-independent association of alcohol and tobacco exposure or prior radiation therapy with promoter methylation of FUSSEL18, EBF3, IRX1, and SEPT9, but not SLC5A8, in head and neck squamous cell carcinomas. *Gene Chromosome Canc, 49* (4):319-26. doi: 10.1002/gcc.20742.

Berezutskaya E., Yu B., & Morozov A. (1997). Differential regulation of the pocket domains of the retinoblastoma family proteins by the HPV 16 E7 oncoprotein. *Cell Growth Differ, 8*:1277 – 86

Bernard, H. U. (2002). Gene expression of genital human papillomaviruses and considerations on potential antiviral approaches. *Antivir Ther, 7,* 219-37.

Blitzer, G. C., Smith, M. A., Harris, S. L., & Kimple, R. J. (2014). Review of the clinical and biologic aspects of human papillomavirus-positive squamous cell carcinomas of the head and neck. *Int J Radiat Oncol Biol Phys, 88*(4), 761-770. doi:10.1016/j.ijrobp.2013.08.029

Bodily, J., & Laimins, L. A. (2011). Persistence of human papillomavirus infection: keys to malignant progression. *Trends Microbiol, 19*(1), 33-39.

Bouda, M., Gorgoulis, V. G., Kastrinakis, N. G., Giannoudis, A., Tsoli, E., Danassi-Afentaki, D., . . . Herrington, C. S. (2000). " High risk" HPV types are frequently detected in potentially malignant and malignant oral lesions, but not in normal oral mucosa. *Modern Pathol, 13*(6), 644.

Boyer, S. N, Wazer, D. E, & Band, V. (1996). E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. *Cancer Res, 56* (20):4620-4.

Brendle, S. A., Bywaters, S. M. & Christensen, N. D. (2014). Pathogenesis of infection by human papillomavirus. *Curr Probl Dermatol, 45*, 47-57.

Blausen.Com Staff (2014). "Medical gallery of Blausen Medical 2014". *WikiJournal of Medicine* **1** (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436.

Cavenaghi, V. B., Ghosn, E. J. E., Cruz, N., Rossi, L. M., da Silva, L., Costa, H. O., & Villa, L. L. (2013). Determination of HPV prevalence in oral/oropharyngeal mucosa samples in a rural district of Sao Paulo. *Braz J Otorhinolar, 79*(5), 599-602. doi:10.5935/1808-8694.20130107

Chaiwongkot, A., Vinokurova, S., Pientong, C., Ekalaksananan, T., Kongyingyoes, B., Kleebkaow, P., … Von Knebel Doeberitz, M. (2013). Differential methylation of E2 binding sites in episomal and integrated HPV 16 genomes in preinvasive and invasive cervical lesions. *Int J Cancer, 132,* 2087-94.

Chatterjee, R., Mukhopadhyay, D., Murmu, N., & Mitra, P. K. (1998). Correlation between human papillomavirus DNA detection in maternal cervical smears and buccal swabs of infants. *Indian J Exp Biol, 36*(2), 199-202.

Chaturvedi, A. K., Anderson, W. F., Lortet-Tieulent, J., Curado, M. P., Ferlay, J., Franceschi, S., . . . Gillison, M. L. (2013). Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. *J Clin Oncol, 31*(36), 4550-4559. doi:10.1200/JCO.2013.50.3870

Chaturvedi, A. K., Engels, E. A., Anderson, W. F., & Gillison, M. L. (2008). Incidence trends for human papillomavirus-related and -unrelated oral squamous cell carcinomas in the United States. *J Clin Oncol, 26*(4), 612-619. doi:10.1200/JCO.2007.14.1713

Chau, N. G., Perez-Ordonez, B., Zhang, K., Pham, N. A., Ho, J., Zhang, T., . . . Siu, L. L. (2011). The association between EGFR variant III, HPV, p16, c-MET, EGFR gene copy number and response to EGFR inhibitors in patients with recurrent or metastatic squamous cell carcinoma of the head and neck. *Head Neck Oncol, 3,* 11. doi:10.1186/1758-3284-3-11

Chesson, H. W., Ekwueme, D. U., Saraiya, M., Watson, M., Lowy, D. R. & Markowitz, L. E. (2012). Estimates of the annual direct medical costs of the prevention and treatment of

disease associated with human papillomavirus in the United States. *Vaccine, 30*, 6016-9.

Chung C. H., & Gillison M.L. (2009). Human papillomavirus in head and neck cancer: its role in pathogenesis and clinical implications. *Clin Cancer Res, 22*, 6758 – 6762

Chung, C. H., Bagheri, A., & D'Souza, G. (2014). Epidemiology of oral human papillomavirus infection. *Oral Oncol, 50*(5), 364-369. doi:10.1016/j.oraloncology.2013.09.003

Colacino, J. A, Dolinoy, D. C, Duffy, S. A, Sartor, M. A, Chepeha, D. B, … Rozek, L. S. (2013). Comprehensive analysis of DNA methylation in head and neck squamous cell carcinoma indicates differences by survival and clinicopathologic characteristics. *PLoS One, 8* (1):e54742. doi: 10.1371/journal.pone.0054742.

Combes J.D., Chen A.A., & Franceschi S. (2014). Prevalence of human papillomavirus in cancer of the oropharynx by gender. *Cancer Epidem Biomar, 23*: 2954–2958. doi: 10.1158/1055-9965.EPI-14-0580.

Cook, R. L., Thompson, E. L., Kelso, N. E., Friary, J., Hosford, J., Barkley, P., . . . Giuliano, A. R. (2014). Sexual behaviors and other risk factors for oral human papillomavirus infections in young women. *Sex Transm Dis, 41*(8), 486-492. doi:10.1097/OLQ.0000000000000159

Cruz, I. B., Snijders, P. J., Steenbergen, R. D., Meijer, C. J., Snow, G. B., Walboomers, J. M., & van der Waal, I. (1996). Age-dependence of human papillomavirus DNA presence in oral squamous cell carcinomas. *Eur J Cancer B Oral Oncol, 32B*(1), 55-62.

Cuschieri, K. S., Cubie, H. A., Whitley, M. W., Gilkison, G., Arends, M. J., Graham, C., & McGoogan, E. (2005). Persistent high risk HPV infection associated with development of cervical neoplasia in a prospective population study. *J Clin Pathol, 58*(9), 946-950. doi:10.1136/jcp.2004.022863

Deng, Z., Hasegawa, M., Kiyuna, A., Matayoshi, S., Uehara, T., Agena, S., … Suzuki, M. (2013). Viral load, physical status, and E6/E7 m RNA expression of human papillomavirus in head and neck squamous cell carcinoma. *Head Neck, 35* (6): 800-8.

D'Souza, G., Agrawal, Y., Halpern, J., Bodison, S., & Gillison, M. L. (2009). Oral Sexual Behaviors Associated with Prevalent Oral Human Papillomavirus Infection. *J Infect Dis, 199*(9), 1263-1269. doi:10.1086/597755

D'Souza, G., Cullen, K., Bowie, J., Thorpe, R., & Fakhry, C. (2014a). Differences in oral sexual behaviors by gender, age, and race explain observed differences in prevalence of oral human papillomavirus infection. *PLoS One, 9*(1), e86023. doi:10.1371/journal.pone.0086023

D'Souza, G., Fakhry, C., Sugar, E. A., Seaberg, E. C., Weber, K., Minkoff, H. L., . . . Gillison, M. L. (2007a). Six-month natural history of oral versus cervical human papillomavirus infection. *Int J Cancer, 121*(1), 143-150. doi:10.1002/ijc.22667

D'Souza, G., Gross, N. D., Pai, S. I., Haddad, R., Anderson, K. S., Rajan, S., . . . Posner, M. R. (2014b). Oral human papillomavirus (HPV) infection in HPV-positive patients with oropharyngeal cancer and their partners. *J Clin Oncol, 32*(23), 2408-2415. doi:10.1200/JCO.2014.55.1341

D'Souza, G., Kreimer, A. R., Viscidi, R., Pawlita, M., Fakhry, C., Koch, W. M., . . . Gillison, M. L. (2007b). Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med, 356*(19), 1944-1956. doi:10.1056/NEJMoa065497

Davidson, C. L., Richter, K. L., Van der Linde, M., Coetsee, J., & Boy, S. C. (2014). Prevalence of oral and oropharyngeal human papillomavirus in a sample of South African men: a pilot study. *S Afr Med J, 104*(5), 358-361. doi:10.7196/samj.7542

do Sacramento, P. R., Babeto, E., Colombo, J., Cabral Ruback, M. J., Bonilha, J. L., Fernandes, A. M., . . . Rahal, P. (2006). The prevalence of human papillomavirus in the oropharynx in healthy individuals in a Brazilian population. *J Med Virol, 78*(5), 614-618. doi:10.1002/jmv.20583

Doeberitz, M. & Vinokurova, S. (2009). Host factors in HPV-related carcinogenesis: cellular mechanisms controlling HPV infections. *Arch Med Res, 40,* 435-42.

Doorbar, J. (2006). Model systems of human papillomavirus-associated disease. *J Pathol, 238,* 166-79.

Doorbar, J., Quint, W., Banks, L., Bravo, I.G, Stoler, M., Broker, T.R, & Stanley Ma. (2012). The biology and life-cycle of human papillomaviruses. *Vaccine, 30* Suppl 5: F55-70.

Du, J., Nordfors, C., Ährlund-Richter, A., Sobkowiak, M., Romanitan, M., Näsman, A., . . . Dalianis, T. (2012). Prevalence of oral human papillomavirus infection among youth, Sweden. *Emerg Infect Dis, 18*(9), 1468.

Durzynska, J., Pacholska-Bogalska, J., Kaczmarek, M., Hanc, T., Durda, M., Skrzypczak, M., & Gozdzicka-Jozefiak, A. (2011). HPV genotypes in the oral cavity/oropharynx of children and adolescents: cross-sectional survey in Poland. *Eur J Pediatr, 170*(6), 757-761. doi:10.1007/s00431-010-1345-x

Dyson, N., Howley, P.M., Münger, K., & Harlow, E. (1989). The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science, 243* (4893):934-7.

Edelstein, Z. R., Schwartz, S. M., Hawes, S., Hughes, J. P., Feng, Q., Stern, M. E., . . . Koutsky, L. A. (2012). Rates and determinants of oral human papillomavirus infection in young men. *Sex Transm Dis, 39*(11), 860-867. doi:10.1097/OLQ.0b013e318269d098

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Brit Med J, 315*(7109), 629-634.

Eike, A., Buchwald, C., Rolighed, J., & Lindeberg, H. (1995). Human Papillomavirus (Hpv) Is Rarely Present in Normal Oral and Nasal-Mucosa. *Clin Otolaryngol, 20*(2), 171-173. doi:DOI 10.1111/j.1365-2273.1995.tb00038.x

Ernster, J. A., Sciotto, C. G., O'Brien, M. M., Robinson, L. J., & Willson, T. (2009). Prevalence of oncogenic human papillomavirus 16 and 18 in the palatine tonsils of the general adult population. *Arch Otolaryngol, 135*(6), 554-557.

Esquenazi, D., Bussoloti, I., Carvalho, M. D. D., & de Barros, F. S. (2010). The frequency of human papillomavirus findings in normal oral mucosa of healthy people by PCR. *Braz J Otorhinolar, 76*(1), 78-84.

Fakhr, M. G., Hagh, M. F., Shanehbandi, D., & Baradaran, B. (2013). DNA methylation pattern as important epigenetic criterion in cancer. *Genet Res Int*, Vol 2013; pg 1 – 9.

Fakhry, C., D'souza, G., Sugar, E., Weber, K., Goshu, E., Minkoff, H., . . . Gillison, M. (2006). Relationship between prevalent oral and cervical human papillomavirus infections in human immunodeficiency virus-positive and -negative women. *J Clin Microbiol, 44*(12), 4479-4485. doi:10.1128/Jcm.01321-06

Fakhry, C., Westra, W. H., Li, S., Cmelak, A., Ridge, J. A., Pinto, H., . . . Gillison, M. L. (2008). Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst, 100*(4), 261-269. doi:10.1093/jnci/djn011

Fakhry, C., Andersen, K.K., Christensen, J., Agrawal, N., & Eisele, D.W. (2015). The impact of tonsillectomy upon the risk of oropharyngeal carcinoma diagnosis and prognosis in

the Danish cancer registry. *Cancer Prev Res (Phila), 8* (7):583-9. doi: 10.1158/1940-6207.CAPR-15-0101.

Feeney, K. M., & Parish, J. L. (2009). Targeting mitotic chromosomes: a conserved mechanism to ensure viral genome persistence. *Proc Biol Sci, 276*(1662), 1535-1544. doi:10.1098/rspb.2008.1642

Fernandes, J. V., Araújo, J. d., & Fernandes, T. (2013). Biology and natural history of human papillomavirus infection. *Open Access J Clin Trials, 5*, 1-12.

Flake, C., Arafa, J., Hall, A., Ence, E., Howard, K., & Kingsley, K. (2012). Screening and detection of human papillomavirus (HPV) high-risk strains HPV16 and HPV18 in saliva samples from subjects under 18 years old in Nevada: a pilot study. *BMC Oral Health, 12*, 43. doi:10.1186/1472-6831-12-43

Forslund, O., Johansson, H., Madsen, K. G., & Kofoed, K. (2013). The nasal mucosa contains a large spectrum of human papillomavirus types from the betapapillomavirus and gammapapillomavirus genera. *J Infect Dis, 208*(8), 1335-1341. doi:10.1093/infdis/jit326

Fu, T. C., Hughes, J. P., Feng, Q. H., Hulbert, A., Hawes, S. E., Xi, L. F., . . . Winer, R. L. (2015). Epidemiology of human papillomavirus detected in the oral cavity and fingernails of mid-adult women. *Sex Trans Dis, 42*(12), 677-685. doi:10.1097/Olq.0000000000000362

Furniss, C.S, Marsit, C.J, Houseman, E.A, Eddy, K., Kelsey, K.T. (2008). Line region hypomethylation is associated with lifestyle and differs by human papillomavirus status in head and neck squamous cell carcinomas. *Cancer Epidem Biomar, 17 (*4):966-71. doi: 10.1158/1055-9965.EPI-07-2775.

Ganti, K., Broniarczyk, J., Manoubi, W., Massimi, P., Mittal, S., Pim, D., … Banks L. (2015). The human papillomavirus E6 PDZ binding motif: from life cycle to malignancy. *Viruses, 7* (7):3530-51.

Gao, G., Johnson, S. H., Kasperbauer, J. L., Eckloff, B. W., Tombers, N. M., Vasmatzis, G. & Smith, D. I. (2014). Mate pair sequencing of oropharyngeal squamous cell carcinomas reveals that HPV integration occurs much less frequently than in cervical cancer. *J Clin Virol, 59*, 195-200.

Garnett T. O., Filippova M., & Duerksen-Hughes P.J. (2006). Accelerated degradation of FADD and procaspase 8 in cells expressing human papillomavirus 16 E6 impairs TRAIL-mediated apoptosis. *Cell Death Differ, 13*: 1915 – 26

Garolla A, Pizzol D, Vasoin F, Barzon L, Bertoldo A, & Foresta C. (2014). Counseling reduces HPV persistence in coinfected couples. *J Sex Med, (1):127*-35. doi: 10.1111/jsm.12358.

Ghittoni, R., Accardi, R., Hasan, U., Gheit, T., Sylla, B. & Tommasino, M. (2010). The biological properties of E6 and E7 oncoproteins from human papillomaviruses. *Virus Genes*, 40, 1-13.

Gichki, A. S., Buajeeb, W., Doungudomdacha, S., & Khovidhunkit, S.-O. P. (2012). Detection of human papillomavirus in normal oral cavity in a group of Pakistani subjects using real-time PCR. *Asian Pac J Cancer P, 13*(5), 2299-2304.

Gillison, M. L., Broutian, T., Pickard, R. K., Tong, Z.-y., Xiao, W., Kahle, L., . . . Chaturvedi, A. K. (2012). Prevalence of oral HPV infection in the United States, 2009-2010. *JAMA Otolaryngol, 307*(7), 693-703.

Gillison, M. L., Koch, W. M., Capone, R. B., Spafford, M., Westra, W. H., Wu, L., . . . Symer, D. E. (2000). Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer I, 92*(9), 709-720.

Goncalves, A. K., Giraldo, P., Barros-Mazon, S., Gondo, M. L., Amaral, R. L., & Jacyntho, C. (2006). Secretory immunoglobulin A in saliva of women with oral and genital HPV infection. *Eur J Obstet Gynecol Reprod Biol, 124*(2), 227-231. doi:10.1016/j.ejogrb.2005.06.028

Gonzalez-Losa, M. D., Barrera, E. S., Herrera-Pech, V., Conde-Ferraez, L., Puerto-Solis, M., & Ayora-Talavera, G. (2015). Epidemiology of oral HPV in the oral mucosa in women without signs of oral disease from Yucatan, Mexico. *Braz J Microbiol, 46*(1), 301-306. doi:10.1590/S1517-838246120130976

Gonzalez-Losa Mdel, R., Manzano-Cabrera, L., Rueda-Gordillo, F., Hernandez-Solis, S. E., & Puerto-Solis, L. (2008). Low prevalence of high risk human papillomavirus in normal oral mucosa by hybrid capture 2. *Braz J Microbiol, 39*(1), 32-34. doi:10.1590/S1517-83822008000100008

Gonzalez, J. V., Gutierrez, R. A., Keszler, A., Colacino Mdel, C., Alonio, L. V., Teyssie, A. R., & Picconi, M. A. (2007). Human papillomavirus in oral lesions. *Medicina (B Aires), 67*(4), 363-368.

Granata, R., Miceli, R., Orlandi, E., Perrone, F., Cortelazzi, B., Franceschini, M., . . . Licitra, L. (2012). Tumor stage, human papillomavirus and smoking status affect the survival of patients with oropharyngeal cancer: an Italian validation study. *Ann Oncol, 23*(7), 1832-1837. doi:10.1093/annonc/mdr544

Gubanova E, Brown B, Ivanov Sv, Helleday T, Mills Gb, Yarbrough Wg, Issaeva N. (2012). Downregulation of SMG-1 in HPV-positive head and neck squamous cell carcinoma due to promoter hypermethylation correlates with improved survival. *Clin Cancer Res, 18* (5):1257-67. doi: 10.1158/1078-0432.CCR-11-2058.

Haedicke J. & Iftner T. (2013). Human papillomaviruses and cancer. *Radiother Oncol Vol 108,* (3): 397 – 402

Hafkamp Hc, Speel Ej, Haesevoets A, Bot Fj, Dinjens Wn, Ramaekers Fc, Hopman Ah, Manni Jj. (2003). A subset of head and neck squamous cell carcinomas exhibits integration of HPV 16/18 DNA and overexpression of p16INK4A and p53 in the absence of mutations in p53 exons 5-8. *Int J Cancer 107* (3):394-400**.**

Hang, D., Liu, F. F., Liu, M. F., He, Z. H., Sun, M., Liu, Y., . . . Ke, Y. (2014). Oral Human Papillomavirus Infection and Its Risk Factors among 5,410 Healthy Adults in China, 2009-2011. *Cancer Epidem Biomar, 23*(10), 2101-2110. doi:10.1158/1055-9965.Epi-14-0084.

Harbour, J.W., Luo, R.X., Dei Santi, A., Postigo, A.A., & Dean, D.C. (1999). Cdk phosphorylation triggers sequential intramolecular interactions that progressively block Rb functions as cells move through G1. *Cell; 98:*859 – 69.

Harris, S. L., Thorne, L. B., Seaman, W. T., Neil Hayes, D., Couch, M. E., & Kimple, R. J. (2011). Association of p16INK4a overexpression with improved outcomes in young patients with squamous cell cancers of the oral tongue. *Head & Neck, 33*(11), 1622-1627.

Hawley-Nelson P, Vousden Kh, Hubbert Nl, Lowy Dr, & Schiller Jt. (1989). HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *EMBO Journal, 8* (12):3905-10.

Heath, S., Willis, V., Allan, K., Purdie, K., Harwood, C., Shields, P., Simcock, R., Williams, T. & Gilbert, D. C. (2012). Clinically significant human papilloma virus in squamous cell carcinoma of the head and neck in UK practice. *Clin Oncol (Royal College of Radiologists), 24*, e18-23.

Herrero, R., Castellsague, X., Pawlita, M., Lissowska, J., Kee, F., Balaram, P., . . . Group, I. M. O. C. S. (2003). Human papillomavirus and oral cancer: the International Agency for Research on Cancer multicenter study. *J Natl Cancer Inst, 95*(23), 1772-1783.

Herrero, R., Quint, W., Hildesheim, A., Gonzalez, P., Struijk, L., Katki, H. A., … Kreimer, A. R. (2013). Reduced prevalence of oral human papillomavirus (HPV) 4 years after bivalent HPV vaccination in a randomized clinical trial in Costa Rica. *PLoS One; 8* (7): e68329.

HES Database. Available at: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics Accessed July 2016.

Hobbs, C. G., Sterne, J. A., Bailey, M., Heyderman, R. S., Birchall, M. A., & Thomas, S. J. (2006). Human papillomavirus and head and neck cancer: a systematic review and meta-analysis. *Clin Otolaryngol, 31*(4), 259-266. doi:10.1111/j.1749-4486.2006.01246.x

Homer, J. & Rees, G. (2012). Pharynx: oropharynx. Chapter 31, Stell and Maran's Textbook of Head and Neck Surgery and Oncology, 5th Edition; pgs: 612 – 613. Hodder and Stoughton Ltd publishers.

Huh K, Zhou X, Hayakawa H, Cho Jy, Libermann Ta, Jin J, Harper Jw, Munger K. (2007). Human papillomavirus type 16 E7 oncoprotein associates with the cullin 2 ubiquitin ligase complex, which contributes to degradation of the retinoblastoma tumor suppressor. *J Virol 81* (18):9737-47.

Huibregtse, J.M, Scheffner, M., & Howley, P.M. (1991). A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. *EMBO Journal, 10:* 4129–35.

Jacquin, E., Baraquin, A., Ramanah, R., Carcopino, X., Morel, A., Valmary-Degano, S., Bravo, I. G., De Sanjose, S., Riethmuller, D., Mougin, C. & Pretet, J. L. (2013). Methylation of Human Papillomavirus Type 16 CpG Sites at E2-Binding Site 1 (E2BS1), E2BS2, and the Sp1-Binding Site in Cervical Cancer Samples as Determined by High-Resolution Melting Analysis-PCR. *J Clin Microbiol, 51,* 3207-3215.

Javadi, P., Sharma, A., Zahnd, W. E., & Jenkins, W. D. (2017). Evolving disparities in the epidemiology of oral cavity and oropharyngeal cancers. *Cancer Cause Control, 28*(6), 635-645. doi:10.1007/s10552-017-0889-8

Jenison, S. A., Yu, X.-p., Valentine, J. M., Koutsky, L. A., Christiansen, A. E., Beckmann, A. M., & Galloway, D. A. (1990). Evidence of prevalent genital-type human papillomavirus infections in adults and children. *J Infect Diseases, 162*(1), 60-69.

Johansson, C. & Schwartz, S. (2013). Regulation of human papillomavirus gene expression by splicing and polyadenylation. *Nat Rev Microbiol, 11,* 239-251.

Johnson, A. M., Mercer, C. H., Erens, B., Copas, A. J., McManus, S., Wellings, K., . . . Field, J. (2001). Sexual behaviour in Britain: partnerships, practices, and HIV risk behaviours. *Lancet, 358*(9296), 1835-1842. doi:10.1016/S0140-6736(01)06883-0

Jones, D.L., Thompson, D.A., & Munger, K. (1997). Destabilization of the RB tumour suppressor protein and stabilization of p53 contribute to HPV type 16 E7-induced apoptosis. *Virology, 239:* 97 – 107.

Jones, D.L., Alani, R.M., & Munger, K. (1997b). The human papillomavirus E7 oncoprotein can uncouple cellular differentiation and proliferation in human keratinocytes by abrogating p21Cip1-mediated inhibition of cdk2. *Gene Dev; 11:* 2101–11.

Kalantari, M., Karlsen, F., Kristensen, G., Holm, R., Hagmar, B. & Johansson, B. (1998). Disruption of the E1 and E2 reading frames of HPV 16 in cervical carcinoma is associated with poor prognosis. *Int J Gynecol Pathol, 17*, 146-153.

Kato, A., Hulse, K. E., Tan, B. K., & Schleimer, R. P. (2013). B-lymphocyte lineage cells and the respiratory system. *J Allergy Clin Immun, 131*(4), 933-957 doi:10.1016/j.jaci.2013.02.023

Katzenellenbogen, R.A., Vliet-Gregg, P., Xu, M., & Galloway, D.A. (2009). NFX1–123 increases hTERT expression and telomerase activity posttranscriptionally in human papillomavirus type 16 E6 keratinocytes. *J Virol; 83:* 6446 – 56.

Kero, K., Rautava, J., Syrjanen, K., Grenman, S., & Syrjanen, S. (2012). Oral mucosa as a reservoir of human papillomavirus: point prevalence, genotype distribution, and incident infections among males in a 7-year prospective study. *Eur Urol, 62*(6), 1063-1070. doi:10.1016/j.eururo.2012.06.045

Kero, K., Rautava, J., Syrjänen, K., Willberg, J., Grenman, S., & Syrjänen, S. (2014). Smoking increases oral HPV persistence among men: 7-year follow-up study. *Eur J Clin Microbiol, 33*(1), 123-133.

Khleif S.N., Degregori J., Yee C. L, Otterson, G.A., Kaye, F.J., Nevins, J.R., & Howley, P.M. (1996). Inhibition of cyclin D-CDK4/CDK6 activity is associated with an E2F-mediated induction of cyclin kinase inhibitor activity. *Proc Nat Acad Sci USA; 93*:4350–4.

Khoury, J.D, Tannir, N.M, Williams, M.D, Chen, Y., Yao, H., Zhang, J., … Su, X. (2013). Landscape of DNA virus associations across human malignant cancers: analysis of 3,775 cases using RNA-Seq. *J Virol 87* (16):8916-26

Kim, J. J., Kobus, K. E., Diaz, M., O'Shea, M., Van Minh, H., & Goldie, S. J. (2008). Exploring the cost-effectiveness of HPV vaccination in Vietnam: insights for evidence-based cervical cancer prevention policy. *Vaccine, 26*(32), 4015-4024. doi:10.1016/j.vaccine.2008.05.038S0264-410X(08)00601-4 [pii]

Kim, S. H., Koo, B. S., Kang, S., Park, K., Kim, H., Lee, K. R., . . . Cho, N. H. (2007). HPV integration begins in the tonsillar crypt and leads to the alteration of p16, EGFR and c-myc during tumor formation. *Int J Cancer, 120*(7), 1418-1425. doi:10.1002/ijc.22464

Klingenberg, B., Hafkamp, H. C., Haesevoets, A., Manni, J. J., Slootweg, P. J., Weissenborn, S. J., . . . Speel, E. J. M. (2010). p16INK4A overexpression is frequently detected in tumour-free tonsil tissue without association with HPV. *Histopathology, 56*(7), 957-967.

Knight, G. L., Needham, L., Ward, D., & Roberts, S. (2016). Pilot study investigating the prevalence of oral Human Papilloma Viral (HPV) infection in young adults. *Public Health, 132*, 105-107. doi:10.1016/j.puhe.2015.12.006

Koppikar, P., deVilliers, E. M., & Mulherkar, R. (2005). Identification of human papillomaviruses in tumors of the oral cavity in an Indian community. *Int J Cancer, 113*(6), 946-950. doi:10.1002/ijc.20664

Koskinen, W.J., Chen, R.W., Leivo, I., Mäkitie, A., Bäck, L., Kontio, R., … Aaltonen, L.M. (2003). Prevalence and physical status of human papillomavirus in squamous cell carcinomas of the head and neck. *Int J Cancer 10*;107(3):401-6.

Koskimaa, H. M., Waterboer, T., Pawlita, M., Grenman, S., Syrjanen, K., & Syrjanen, S. (2012). Human Papillomavirus Genotypes Present in the Oral Mucosa of Newborns and their Concordance with Maternal Cervical Human Papillomavirus Genotypes. *J Pediatr, 160*(5), 837-843. doi:10.1016/j.jpeds.2011.10.027

Kostareli. E., Holzinger, D., Bogatyrova, O., Hielscher, T., Wichmann, G., Keck, M., … Hess, J. (2013). HPV-related methylation signature predicts survival in oropharyngeal squamous cell carcinomas. *J Clin Invest 123* (6):2488-501. doi: 10.1172/JCI67010.

Kravchenko, S. (2011). Risk factors for transmission and persistence of HPV infection in children's oropharynx in the first 6 months of life. *Eur J Med Res, 16*, 53-54.

Kreimer, A. R., Bhatia, R. K., Messeguer, A. L., Gonzalez, P., Herrero, R., & Giuliano, A. R. (2010). Oral human papillomavirus in healthy individuals: a systematic review of the literature. *Sex Transm Dis, 37*(6), 386-391. doi:10.1097/OLQ.0b013e3181c94a3b

Kreimer, A. R., Pierce Campbell, C. M., Lin, H. Y., Fulp, W., Papenfuss, M. R., Abrahamsen, M., . . . Giuliano, A. R. (2013). Incidence and clearance of oral human papillomavirus infection in men: the HIM cohort study. *Lancet, 382*(9895), 877-887. doi:10.1016/S0140-6736(13)60809-0

Kreimer, A. R., Villa, A., Nyitray, A. G., Abrahamsen, M., Papenfuss, M., Smith, D., . . . Giuliano, A. R. (2011). The epidemiology of oral HPV infection among a multinational sample of healthy men. *Cancer Epidem Biomar, 20*(1), 172-182. doi:10.1158/1055-9965.Epi-10-0682

Kubba, T. (2008). Human papillomavirus vaccination in the United Kingdom: what about boys? *Reprod Health matter, 16*(32), 97-103.

Kuhs, K. A. L., Gonzalez, P., Struijk, L., Castro, F., Hildesheim, A., van Doorn, L. J., . . . Grp, C. R. V. T. (2013). Prevalence of and Risk Factors for Oral Human Papillomavirus Among Young Women in Costa Rica. *J Infect Dis, 208*(10), 1643-1652. doi:10.1093/infdis/jit369

Kurose, K., Terai, M., Soedarsono, N., Rabello, D., Nakajima, Y., Burk, R. D., & Takagi, M. (2004). Low prevalence of HPV infection and its natural history in normal oral mucosa among volunteers on Miyako Island, Japan. *Oral Surg Oral Med O, 98*(1), 91-96. doi:10.1016/j.tripleo.2003.12.029

Lace, M.J., Anson, J.R., Klussmann, J.P., Wang, D.H., Smith, E.M., Haugen, T.H. & Turek L.P. (2011). Human papillomavirus type 16 (HPV-16) genomes integrated in head and neck cancers and in HPV-16-immortalized human keratinocyte clones express chimeric virus-cell mRNAs similar to those found in cervical cancers. *J Virol, 85*(4):1645-54. doi: 10.1128/JVI.02093-10.

Lawton, G., Thomas, S., Schonrock, J., Monsour, F., & Frazer, I. (1992). Human papillomaviruses in normal oral mucosa: a comparison of methods for sample collection. *J Oral Pathol Med, 21*(6):265-9.

Lechner M, Fenton T, West J, Wilson G, Feber A, Henderson S, … Beck S. (2013). Identification and functional validation of HPV-mediated hypermethylation in head and neck squamous cell carcinoma. *Genome Med, 5*(2):15. doi: 10.1186/gm419. eCollection 2013.

Licitra, L., Perrone, F., Bossi, P., Suardi, S., Mariani, L., Artusi, R., . . . Pilotti, S. (2006). High-risk human papillomavirus affects prognosis in patients with surgically treated oropharyngeal squamous cell carcinoma. *J Clin Oncol, 24*(36), 5630-5636. doi:10.1200/JCO.2005.04.6136

Lindel, K., Beer, K. T., Laissue, J., Greiner, R. H., & Aebersold, D. M. (2001). Human papillomavirus positive squamous cell carcinoma of the oropharynx: a radiosensitive subgroup of head and neck carcinoma. *Cancer, 92*(4), 805-813.

Longworth, M. S., & Laimins, L. A. (2004). Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiol Mol Biol R, 68*(2), 362-372.

Louie, K. S., Mehanna, H., & Sasieni, P. (2015). Trends in head and neck cancers in England from 1995 to 2011 and projections up to 2025. *Oral Oncol, 51*(4), 341-348. doi:10.10

Louvanto, K., Rautava, J., Willberg, J., Wideman, L., Syrjanen, K., Grenman, S. & Syrjanen, S. (2013). Genotype-specific incidence and clearance of human papillomavirus in oral mucosa of women: a six-year follow-up study. *PLoS One, 8*(1): e53413. doi: 10.1371/journal.pone.005341316/j.oraloncology.2015.01.002

Machado, A. P., de Almeida, F. G., Bonin, C. M., Prata, T. T. M., Avilla, L. S., Padovani, C. T. J., . . . Tozetti, I. A. (2014). Presence of highly oncogenic human papillomavirus in the oral mucosa of asymptomatic men. *Braz J Infect Dis, 18*(3), 266-270. doi:10.1016/j.bjid.2013.07.013

Mammas, I. N., Sourvinos, G., Giamarelou P., Michael C., & Spandidos, D. A. (2012). Human papillomavirus in the oral cavity of children and mode of delivery: a retrospective study. *Int J STD AIDS 23*(3):185-8.

Mammas, I. N., Sourvinos, G., Michael, C., & Spandidos, D. A. (2006). Human papilloma virus in hyperplastic tonsillar and adenoid tissues in children. *Pediatr Infect Dis J, 25*(12), 1158-1162.

Mantovani, F. & Banks, L. (2001). The human papillomavirus E6 protein and its contribution to malignant progression. *Oncogene, 20*(54):7874-87.

Mao, E. J. (1995). Prevalence of human papillomavirus 16 and nucleolar organizer region counts in oral exfoliated cells from normal and malignant epithelia. *Oral Surg Oral Med O, 80*(3), 320-329.

Marais, D. J., Sampson, C., Jeftha, A., Dhaya, D., Passmore, J. A. S., Denny, L., . . . Williamson, A. L. (2006). More men than women make mucosal IgA antibodies to Human papillomavirus type 16 (HPV-16) and HPV-18: a study of oral HPV and oral HPV antibodies in a normal healthy population. *BMC Infect Dis, 6*. doi:Artn

Martinelli, M., Zappa A., Bianchi S., Frati E., Colzani D., Amendola A., & Tanzi E. (2012). Human papillomavirus (HPV) infection and genotype frequency in the oral mucosa of newborns in Milan, Italy. *Clin Microbiol Infect, 18*(6):E197-9. doi: 10.1111/j.1469-0691.2012.038399510.1186/1471-2334-6-95

Martinez, G. G., & Troconis, J. N. (2014). Natural history of the infection for human papillomavirus: an actualization. *Invest Clin, 55*(1), 82-91.

Matsushita, K., Sasagawa, T., Miyashita, M., Ishizaki, A., Morishita, A., Hosaka, N., . . . Ichimura, H. (2011). Oral and cervical human papillomavirus infection among female sex workers in Japan. *Jpn J Infect Dis, 64*(1), 34-39.

McClory, S., Hughes, T., Freud, A. G., Briercheck, E. L., Martin, C., Trimboli, A. J., . . . Caligiuri, M. A. (2012). Evidence for a stepwise program of extrathymic T cell development within the human tonsil. *J Clin Invest, 122*(4), 1403-1415. doi:10.117

McCord, C., Xu, J., Xu, W., Qiu, X., Mccomb R. J., Perez-Ordonez, B. et al. (2013). Association of high-risk human papillomavirus infection with oral epithelial dysplasia. *Or Surg Or Med Or Pa, 115* (4): 541 – 9.2/JCI46125.

McCredie, M. R., Sharples, K. J., Paul, C., Baranyai, J., Medley, G., Jones, R. W., & Skegg, D. C. (2008). Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study. *Lancet Oncol, 9*(5), 425-434. doi:10.1016/S1470-2045(08)70103-7

Mclaughlin-Drubin, M. E., Meyers, J. & Munger, K. (2012). Cancer associated human papillomaviruses. *Curr Opin Virol, 2*, 459-66.

Mehanna, H., Beech, T., Nicholson, T., El-Hariry, I., McConkey, C., Paleri, V., & Roberts, S. (2013). Prevalence of human papillomavirus in oropharyngeal and nonoropharyngeal head and neck cancer - systematic review and meta-analysis of trends by time and region. *Head Neck, 35*(5), 747-755. doi:10.1002/hed.22015

Mehanna, H., Jones, T. M., Gregoire, V., & Ang, K. K. (2010). Oropharyngeal carcinoma related to human papillomavirus. *BMJ, 340*, c1439. doi:10.1136/bmj.c1439

Mehanna, H., Wong, W. L., McConkey, C. C., Rahman, J. K., Robinson, M., Hartley, A. G., . . . Group, P.-N. T. M. (2016). PET-CT Surveillance versus Neck Dissection in Advanced Head and Neck Cancer. *N Engl J Med, 374*(15), 1444-1454. doi:10.1056/NEJMoa1514493

Mellin, H., Dahlgren, L., Munck-Wikland, E., Lindholm, J., Rabbani, H., Kalantari, M., & Dalianis, T. (2002). Human papillomavirus type 16 is episomal and a high viral load may be correlated to better prognosis in tonsillar cancer. *Int J Cancer, 102*(2), 152-158.

Mercer, C. H., Tanton, C., Prah, P., Erens, B., Sonnenberg, P., Clifton, S., . . . Johnson, A. M. (2013). Changes in sexual attitudes and lifestyles in Britain through the life course and over time: findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). *Lancet, 382*(9907), 1781-1794. doi:10.1016/S0140-6736(13)62035-8

Migaldi, M., Pecorari, M., Forbicini, G., Nanni, N., Grottola, A., Grandi, T., . . . Sgambato, A. (2012). Low prevalence of human papillomavirus infection in the healthy oral mucosa of a Northern Italian population. *J Oral Pathol Med, 41*(1), 16-20. doi:10.1111/j.1600-0714.2011.01062.x

Modrow, S., Falke, D., Truyen, U. & Schätzl, H., (2013). In: Molecular Virology. Heidelberg: Springer Akademischer Verlag, pp. 679-680.

Mooren, J. J., Kremer, B., Claessen, S. M., Voogd, A. C., Bot, F. J., Klussmann, P. J., … Speel E. J. (2013). Chromosome stability in tonsillar squamous cell carcinoma is associated with HPV16 integration and indicates a favorable prognosis. *Int J Cancer, 132*(8):1781-9. doi: 10.1002/ijc.27846.

Munoz, N., Castellsague, X., de Gonzalez, A. B., & Gissmann, L. (2006). Chapter 1: HPV in the etiology of human cancer. *Vaccine, 24* Suppl 3, S3/1-10. doi:10.1016/j.vaccine.2006.05.115

National Cancer Intelligence Network (2010). Profile of Head and Neck Cancers in England: Incidence, Mortality and Survival. Vol 1: 1 – 72

Nguyen, D.X., Westbrook, T.F. & McCance, D.J., (2002). Human papillomavirus type 16 E7 maintains elevated levels of the cdc25A tyrosine phosphatase during deregulation of cell cycle arrest. *J Virol, 76*: 619 – 32.

Nordfors, C., Grun, N., Haeggblom, L., Tertipis, N., Sivars, L., Mattebo, M., . . . Dalianis, T. (2013). Oral human papillomavirus prevalence in high school students of one municipality in Sweden. *Scand J Infect Dis, 45*(11), 878-881. doi:10.3109/00365548.2013.821626.

Olthof, N. C., Huebbers, C. U., Kolligs, J., Henfling, M., Ramaekers, F. C. S., Cornet, I., … Speel, E. J. M. (2015). Viral load, gene expression and mapping of viral integration sites in HPV16-associated HNSCC cell lines. *Int J Cancer, 136,* E207-E218.

Ostwald, C., Muller, P., Barten, M., Rutsatz, K., Sonnenbung, M., Mildelangosch, K., & Loning, T. (1994). Human papillomavirus DNA in oral squamous-cell carcinomas and normal mucosa. *J Oral Pathol Med, 23*(5), 220-225. doi:DOI 10.1111/j.1600-0714.1994.tb01117.x

Pakarian, F., Kaye, J., Cason, J., Kell, B., Jewers, R., Derias, N. W., . . . Best, J. M. (1994). Cancer associated human papillomaviruses: perinatal transmission and persistence. *Brit J Obstet Gyne, 101*(6), 514-517.

Palefsky, J. M. (2010). Human papillomavirus-related disease in men: not just a women's issue (vol 46, pg S12, 2010). *J Adolescent Health, 46*(6), 614-614. doi:10.1016/j.jadohealth.2010.04.001

Paleri, V., Urbano, T. G., Mehanna, H., Repanos, C., Lancaster, J., Roques, T., . . . Sen, M. (2016). Management of neck metastases in head and neck cancer: United Kingdom national multidisciplinary guidelines. *J Laryngol Otol, 130*(S2), S161-S169. doi:10.1017/S002221511600058X

Palmer, E., Newcombe, R. G., Green, A. C., Kelly, C., Noel, G. O., Hall, G., …. Powell., N. G. (2014). Human papillomavirus infection is rare in non-malignant tonsil tissue in the UK: implications for tonsil cancer precursor lesions. *Int J Cancer; 135* (10): 2437 – 43.

Park, I. S., Chang, X., Loyo, M., Wu, G., Chuang, A., Kim, M. S., . . . Pai, S. I. (2011). Characterization of the methylation patterns in human papillomavirus type 16 viral DNA in head and neck cancers. *Cancer Prev Res (Phila), 4*(2), 207-217. doi:10.1158/1940-6207.CAPR-10-0147

Pedersen, H., & Hariri, J. (2012). Oropharyngeal HPV in healthy individuals assessed by self-sampling. *APMIS, 120*, 31-31.

Petitti, D. B. (2001). Approaches to heterogeneity in meta-analysis. *Stat Med, 20*(23), 3625-3633. doi:DOI 10.1002/sim.1091.

Phelps, W. C. & Howley, P. M. 1987. Transcriptional trans-activation by the human papillomavirus type 16 E2 gene product. *J Virol, 61,* 1630-8.

Pickard, R. K. L., Xiao, W. H., Broutian, T. R., He, X., & Gillison, M. L. (2012). The Prevalence and incidence of oral human papillomavirus infection among young men and women, aged 18-30 years. *Sex Transm Dis, 39*(7), 559-566. doi:10.1097/OLQ.0b013e31824f1c65

Pinheiro Rdos, S., de Franca, T. R., Ferreira Dde, C., Ribeiro, C. M., Leao, J. C., & Castro, G. F. (2011). Human papillomavirus in the oral cavity of children. *J Oral Pathol Med, 40*(2), 121-126. doi:10.1111/j.1600-0714.2010.00954.x

Poage, G.M., Houseman, E.A., Christensen, B.C., Butler, R.A., Avissar-Whiting, M., McClean, M.D., … Kelsey K.T. (2011). Global hypomethylation identifies Loci targeted for hypermethylation in head and neck cancer. *Clin Cancer Res, 17*(11):3579-89. doi: 10.1158/1078-0432.CCR-11-0044.

Office Of National Statistics (2016). Population Estimates For UK, England And Wales, Scotland And Northern Ireland: Mid 2016 https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2016

Puranen, M., Yliskoski, M., Saarikoski, S., Syrjänen, K., & Syrjänen, S. (1996). Vertical transmission of human papillomavirus from infected mothers to their newborn babies and persistence of the virus in childhood. *Am J Obstet Gynecol, 174*(2), 694-699.

Rautava, J., Willberg, J., Louvanto, K., Wideman, L., Syrjanen, K., Grenman, S., & Syrjanen, S. (2012). Prevalence, genotype distribution and persistence of human papillomavirus in oral mucosa of women: a six-year follow-up study. *PLoS One, 7*(8), e42171. doi:10.1371/journal.pone.0042171

Raybould, R., Fiander, A., Wilkinson, G. W. G. & Hibbitts, S. (2014). HPV integration detection in CaSki and SiHa using detection of integrated papillomavirus sequences and restriction-site PCR. *J Virol Methods, 206*, 51-54.

Rayur.com (2012). Picture of Oropharynx – image thanks to uc.edu. Available at: http://www.rayur.com/oropharynx-inflammation-of-the-tonsils.html

Reimers, N., Kasper, H. U., Weissenborn, S. J., Stutzer, H., Preuss, S. F., Hoffmann, T. K., . . . Klussmann, J. P. (2007). Combined analysis of HPV-DNA, p16 and EGFR expression to predict prognosis in oropharyngeal cancer. *Int J Cancer, 120*(8), 1731-1738. doi:10.1002/ijc.22355

Remmink, A. J., Walboomers, J. M. M., Helmerhorst, T. J. M., Voorhorst, F. J., Rozendaal, L., Risse, E. K. J., . . . Kenemans, P. (1995). The Presence of Persistent High-Risk Hpv Genotypes in Dysplastic Cervical Lesions Is Associated with Progressive Disease - Natural-History up to 36 Months. *Int J Cancer, 61*(3), 306-311. doi:DOI 10.1002/ijc.2910610305

Ren, S., Gaykalova, D., Wang, J., Guo, T., Danilova, L., Favorov, A., … Califano Ja (2018). Discovery and development of differentially methylated regions in human papillomavirus related oropharyngeal squamous cell carcinoma. *Int J Cancer.* doi: 10.1002/ijc.31778. [Epub ahead of print]

Reuschenbach, M., Huebbers, C. U., Prigge, E. S., Bermejo, J. L., Kalteis, M. S., Preuss, S. F., … Von Knebel Doeberitz, M. (2015). Methylation status of HPV16 E2-binding sites classifies subtypes of HPV-associated oropharyngeal cancer. *Cancer, 121,* 1966-76.

Ribeiro, K. M. X., Alvez, J. M., Pignatari, S. S., & Weckx, L. L. M. (2006). Detection of human papilloma virus in the tonsils of children undergoing tonsillectomy. *Braz J Infect Dis, 10*(3), 165-168.

Richards, K.l, Zhang, B., Baggerly, K.A., Colella, S., Lang, J.C., Schuller, D.E., & Krahe, R. (2009). Genome-wide hypomethylation in head and neck cancer is more pronounced in HPV-negative tumors and is associated with genomic instability. *PLoS One 4*(3):e4941. doi: 10.1371/journal.pone.0004941.

Rintala, M., Grénman, S., Puranen, M., & Syrjänen, S. (2006). Natural history of oral papillomavirus infections in spouses: a prospective Finnish HPV family study. *J Clin Virol, 35*(1), 89-94.

Rischin, D., Young, R. J., Fisher, R., Fox, S. B., Le, Q.-T., Peters, L. J., . . . Kenny, L. M. (2010). Prognostic significance of p16INK4A and human papillomavirus in patients with oropharyngeal cancer treated on TROG 02.02 phase III trial. *J Clin Oncol, 28*(27), 4142-4148.

Roberts, S. (2015). HPV lifecycle picture taken from Papillomaviruses. *eLS article*. Available from: https://doi.org/10.1002/9780470015902.a0000422.pub3.

Roberts, S., Kingsbury, S. R., Stoeber, K., Knight, G. L., Gallimore, P. H., & Williams, G. H. (2008). Identification of an arginine-rich motif in human papillomavirus type 1 E1;E4 protein necessary for E4-mediated inhibition of cellular DNA synthesis in vitro and in cells. *J Virol, 82*(18), 9056-9064. doi:10.1128/JVI.01080-08.

Romanczuk, H., Thierry, F. & Howley, P. M. (1990). Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters. *J Virol, 64*, 2849-59.

Rombaldi, R. L., Serafini, E. P., Mandelli, J., Zimmermann, E., & Losquiavo, K. P. (2009). Perinatal transmission of human papilomavirus DNA. *Virol J, 6*, 83. doi:10.1186/1743-422X-6-83

Rosen, B. J., Walter, L., Gilman, R. H., Cabrerra, L., Gravitt, P. E., & Marks, M. A. (2016). Prevalence and correlates of oral human papillomavirus infection among healthy males and females in Lima, Peru. *Sex Transm Infect, 92*(2), 149-154. doi:10.1136/sextrans-2014-051942

Rosenquist, S. E. (2012). Is oral sex really a dangerous carcinogen? Let's take a closer look. *J Sex Med, 9*(9), 2224-2232.

Rubin, S.M., Gall, A.L., Zheng, N. & Pavletich N.P. (2005). Structure of the Rb C-terminal domain bound to E2F1–1: a mechanism for phosphorylation-induced E2F release. *Cell, 123*:1093 – 106.

Ruiz, G., Ojeda, P., & Di Giampietro, L. (2009). Determinación del VPH en cavidad oral por técnica del hisopo. *Rev Venez Oncol, 21*(2).

SahebJamee, M., Boorghani, M., Ghaffari, S. R., AtarbashiMoghadam, F., & Keyhani, A. (2009). Human papillomavirus in saliva of patients with oral squamous cell carcinoma. *Med Oral Pathol Oral Cir Bucal, 14*(10), e525-528.

Sanders, A. E., Slade, G. D., & Patton, L. L. (2012). National prevalence of oral HPV infection and related risk factors in the U.S. adult population. *Oral Dis, 18*(5), 430-441. doi:10.1111/j.1601-0825.2011.01892.x

Sartor M.A., Dolinoy D.C., Jones T.R., Colacino J.A., Prince M.E., Carey T.E. & Rozek L.S. (2011). Genome-wide methylation and expression differences in HPV(+) and HPV(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. *Epigenetics, 6*(6):777-87.

Satterwhite, C. L., Torrone, E., Meites, E., Dunne, E. F., Mahajan, R., Ocfemia, M. C. B., . . . Weinstock, H. (2013). Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008. *Sex Transm Dis, 40*(3), 187-193. doi:10.1097/OLQ.0b013e318286bb53

Schache, A. G., Liloglou, T., Risk, J. M., Filia, A., Jones, T. M., Sheard, J., . . . Shaw, R. J. (2011). Evaluation of human papilloma virus diagnostic testing in oropharyngeal squamous cell carcinoma: sensitivity, specificity, and prognostic discrimination. *Clin Cancer Res, 17*(19), 6262-6271. doi:10.1158/1078-0432.CCR-11-0388

Schache, A. G., Powell, N. G., Cuschieri, K. S., Robinson, M., Leary, S., Mehanna, H., … Jones, T. M. (2006). HPV-related oropharyngeal cancer in the United Kingdom: an evolution in understanding of disease etiology. *Cancer Res, 76*: 6598 – 6606.

Scheffner, M., Werness, B.A, Huibregtse, J.M, Levine, A.J., & Howley P.M. (1990). The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell, 63*: 1129 – 36.

Scheffner M., Huibregtse J.M., Vierstra R.D. & Howley P.M. (1993). The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell, 75*:495–505.

Schettino, M. T., Ammaturo, F. P., Grimaldi, E., Legnante, A., Marcello, A., Donnarumma, G., . . . Torella, M. (2014). Persistent papillomavirus type-31 and type-45 infections predict the progression to squamous intraepithelial lesion. *Taiwan J Obstet Gyne, 53*(4), 494-497. doi:10.1016/j.tjog.2014.06.001

Schiffman, M., Bauer, H., Lorincz, A., Manos, M., Byrne, J., Glass, A., . . . Howley, P. (1991). Comparison of Southern blot hybridization and polymerase chain reaction

methods for the detection of human papillomavirus DNA. *J Clin Microbiol, 29*(3), 573-577.
Schiffman, M., Castle, P. E., Jeronimo, J., Rodriguez, A. C. & Wacholder, S. 2007. Human papillomavirus and cervical cancer. *Lancet, 370,* 890-907.
Schlecht, N. F., Rojas, M., Lorde-Rollins, E., Nucci-Sack, A., Strickler, H. D., Burk, R. D., & Diaz, A. (2013). Burden of cervical, anal, and oral HPV in an inner-city pre-vaccine adolescent population. *J Urban Health 90*(1), 141-146. doi:10.1007/s11524-012-9756-9
Schnelle, C., Whiteman, D. C., Porceddu, S. V., Panizza, B. J., & Antonsson, A. (2017). Past sexual behaviors and risks of oropharyngeal squamous cell carcinoma: a case-case comparison. *Int J Cancer, 140*(5), 1027-1034. doi:10.1002/ijc.30519
Seaman, W. T., Andrews, E., Couch, M., Kojic, E. M., Cu-Uvin, S., Palefsky, J., . . . Webster-Cyriaque, J. (2010). Detection and quantitation of HPV in genital and oral tissues and fluids by real time PCR. *Virol J, 7*, 194. doi:10.1186/1743-422X-7-194
Seifi, S., Kermani, I. A., Dolatkhah, R., Kermani, A. A., Sakhinia, E., Asgarzadeh, M., . . . Kermani, T. A. (2013). Prevalence of Oral Human Papilloma Virus in Healthy Individuals in East Azerbaijan Province of Iran. *Iran J Public Health, 42*(1), 79-85.
Senba, M. & Mori, N. (2012). Mechanisms of virus immune evasion lead to development from chronic inflammation to cancer formation associated with human papillomavirus infection. *Oncology Reviews.* Available at www.oncologyreviews.org
Shah, N. G., Trivedi, T. I., Tankshali, R. A., Goswami, J. V., Jetly, D. H., Shukla, S. N., . . . Verma, R. J. (2009). Prognostic significance of molecular markers in oral squamous cell carcinoma: a multivariate analysis. *Head Neck, 31*(12), 1544-1556. doi:10.1002/hed.21126
Shaw, R. & Robinson, M. (2011). The increasing clinical relevance of human papillomavirus type 16 (HPV-16) infection in oropharyngeal cancer. *Brit J Oral Max Surg, 49,* 423-9.

Shindoh, M., Chiba, I., Yasuda, M., Saito, T., Funaoka, K., Kohgo, T., Amemiya, A., Sawada, Y. & Fujinaga, K. (1995) Detection of human papillomavirus DNA sequences in oral squamous cell carcinomas and their relation to p53 and proliferating cell nuclear antigen expression. *Cancer, 76,* 1513–1521
Sikka, S., & Sikka, P. (2014). Association of Human Papilloma Virus 16 Infection and p53 Polymorphism among Tobacco using Oral Leukoplakia Patients: A Clinicopathologic and Genotypic Study. *Int J Prev Med, 5*(4), 430-438.
Singhi, A. D., & Westra, W. H. (2010). Comparison of human papillomavirus in situ hybridization and p16 immunohistochemistry in the detection of human papillomavirus-associated head and neck cancer based on a prospective clinical experience. *Cancer, 116*(9), 2166-2173. doi:10.1002/cncr.25033
Sisk, J., Schweinfurth, J. M., Wang, X. T., & Chong, K. (2006). Presence of human papillomavirus DNA in tonsillectomy specimens. *Laryngoscope, 116*(8), 1372-1374.
Slansky, J.E. & Farnham, P.J. (1996). Introduction to the E2F family: protein structure and gene regulation. *Curr Top Microbiol, 208*:1-30.
Smith, D. F., Maleki, Z., Coughlan, D., Gooi, Z., Akpeng, B., Ogawa, T., . . . Pai, S. I. (2014). Human papillomavirus status of head and neck cancer as determined in cytologic specimens using the hybrid-capture 2 assay. *Oral Oncol, 50*(6), 600-604. doi:10.1016/j.oraloncology.2014.02.011
Smith, E. M., Ritchie, J. M., Summersgill, K. F., Klussmann, J. P., Lee, J. H., Wang, D., . . . Turek, L. P. (2004). Age, sexual behavior and human papillomavirus infection in oral

cavity and oropharyngeal cancers. *Int J Cancer, 108*(5), 766-772. doi:10.1002/ijc.11633

Smith, E. M., Swarnavel, S., Ritchie, J. M., Wang, D., Haugen, T. H., & Turek, L. P. (2007). Prevalence of human papillomavirus in the oral cavity/oropharynx in a large population of children and adolescents. *Pediatr Infect Dis J, 26*(9), 836-840.

Smotkin, D. & Wettstein, F.O. (1986). Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proc Natl Acad Sci U S A, 83*(13):4680-4.

Soeda, E., Ferran, M. C., Baker, C. C., Mcbride, A. A. (2006). Repression of HPV 16 early region transcription by the E2 protein. *Virology 351,* 29 – 41.

Stanley, M. A. (2012). Epithelial Cell Responses to Infection with Human Papillomavirus. *Clin Microbiol Rev, 5*(2), pp. 215-220.

Strati, K. & Lambert, P.F. (2007). Role of Rb-dependent and Rb-independent functions of papillomavirus E7 oncogene in head and neck cancer. *Cancer Res, 67*(24):11585-93.

Sturgis, E. M. & Ang, K. K. (2011). The epidemic of HPV-associated oropharyngeal cancer is here: is it time to change our treatment paradigms? *J Nat Comp Cancer Net,* 9, 665-673.

Summersgill, K. F., Smith, E. M., Levy, B. T., Allen, J. M., Haugen, T. H., & Turek, L. P. (2001). Human papillomavirus in the oral cavities of children and adolescents. *Oral Surg, Oral Med, O, 91*(1), 62-69.

Suwalska, A., Owczarek, W., & Fiedor, P. (2014). [Clinical usefulness of diagnostic methods for human papilloma virus dependent lesions]. *Pol Merkur Lekarski, 36*(212), 129-132.

Suzich, J. A., Ghim, S. J., Palmerhill, F. J., White, W. I., Tamura, J. K., Bell, J. A., . . . Schlegel, R. (1995). Systemic Immunization with Papillomavirus L1 Protein Completely Prevents the Development of Viral Mucosal Papillomas. *P Nat Acad Sci USA, 92*(25), 11553-11557. doi:DOI 10.1073/pnas.92.25.11553

Syrjanen, S. (2003). Human papillomavirus infections and oral tumors. *Med Microbiol Immun, 192*(3), 123-128. doi:10.1007/s00430-002-0173-7

Syrjanen, S. (2010). Current concepts on human papillomavirus infections in children. *APMIS, 118*(6-7), 494-509. doi:10.1111/j.1600-0463.2010.02620.x

Syrjanen, S., Lodi, G., von Bultzingslowen, I., Aliko, A., Arduino, P., Campisi, G., . . . Jontell, M. (2011). Human papillomaviruses in oral carcinoma and oral potentially malignant disorders: a systematic review. *Oral Dis, 17 Suppl 1*, 58-72. doi:10.1111/j.1601-0825.2011.01792.x

Syrjanen, S., & Puranen, M. (2000). Human papillomavirus infections in children: the potential role of maternal transmission. *Crit Rev Oral Biol M, 11*(2), 259-274.

Syrjanen, S., Termine, N., Capra, G., Paderni, C., Panzarella, V., & Campisi, G. (2012). Oral HPV infection: current strategies for prevention and therapy. *Curr Pharm Design, 18*(34), 5452-5469.

Tachezy, R., Klozar, J., Rubenstein, L., Smith, E., Salakova, M., Smahelova, J., . . . Hamsikova, E. (2009). Demographic and risk factors in patients with head and neck tumors. *J Med Virol, 81*(5), 878-887. doi:10.1002/jmv.21470

Thain A., Jenkins O., Clarke A.R., & Gaston K. (1996). CpG methylation directly inhibits binding of the human papillomavirus type 16 E2 protein to specific DNA sequences. *J Virol., 70*(10):7233-5.

Thierry, F. (2009). Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. *Virology, 384,* 375-9.

Thomas, M., Pim, D., & Banks, L. (1999). The role of the E6-p53 interaction in the molecular pathogenesis of HPV. *Oncogene, 18:* 7690 – 700.

Tristao, W., Ribeiro, R. M. P., de Oliveira, C. A., Betiol, J. C., & Bettini, J. D. R. (2012). Epidemiological study of HPV in oral mucosa through PCR. *Braz J Otorhinolar, 78*(4), 66-70.

Turner, D. O., Williams-Cocks, S. J., Bullen, R., Catmull, J., Falk, J., Martin, D., . . . Kingsley, K. (2011). High-risk human papillomavirus (HPV) screening and detection in healthy patient saliva samples: a pilot study. *BMC Oral Health, 11*, 28. doi:10.1186/1472-6831-11-28

Van Kempen, P. M., Noorlag, R., Braunius, W. W., Stegeman, I., Willems, S. M. & Grolman, W. (2014). Differences in methylation profiles between HPV-positive and HPV-negative oropharynx squamous cell carcinoma: a systematic review. *Epigenetics, 9,* 194-203.

Vega-Pena, A., Illades-Aguiar, B., Flores-Alfaro, E., Lopez-Bayghen, E., Leyva-Vazquez, M. A., Castaneda-Saucedo, E., & Alarcon-Romero, L. D. (2013). Risk of progression of early cervical lesions is associated with integration and persistence of HPV-16 and expression of E6, Ki-67, and telomerase. *J Cytol, 30*(4), 226-232. doi:10.4103/0970-9371.126644.

Villavicencio-Torres, W., Lacourt-Ventura, M. Y., Fonseca-Williams, S. C., Del Mar Gonzalez-Pons, M., Diaz-Algori, Y., Torres-Ramos, C., Bernabe-Dones, R. D. & Cruz-Correa, M. R. (2013) [abstract]. In: Proceedings of the 104th Annual Meeting of the American Association for Cancer Research; 2013 Apr 6-10; Washington, DC. Philadelphia (PA): AACR; Cancer Res 2013;73(8 Suppl):Abstract nr 4776. doi:10.1158/1538-7445.AM2013-4776

Vjekoslav T., (2016). Functional Roles of E6 and E7 Oncoproteins in HPV-Induced Malignancies at Diverse Anatomical Sites. *Cancers (Basel)., 8*(10): 95. doi: 10.3390/cancers8100095

Vojtechova, Z., Sabol, I., Salakova, M., Turek, L., Grega, M., Smahelova, J., Vencalek, O., Lukesova, E., Klozar, J. & Tachezy, R. (2016). Analysis of the integration of human papillomaviruses in head and neck tumours in relation to patients' prognosis. *Int J Cancer, 138,* 386-95.

Weinberger, P. M., Yu, Z., Haffty, B. G., Kowalski, D., Harigopal, M., Brandsma, J., . . . Psyrri, A. (2006). Molecular classification identifies a subset of human papillomavirus--associated oropharyngeal cancers with favorable prognosis. *J Clin Oncol, 24*(5), 736-747. doi:10.1200/JCO.2004.00.3335

Weiss D, Koopmann M, Rudack C. (2011). Prevalence and impact on clinicopathological characteristics of human papillomavirus-16 DNA in cervical lymph node metastases of head and neck squamous cell carcinoma. *Head Neck, 33*(6):856-62. doi: 10.1002/hed.21548.

Who International Agency For Research On Cancer (2007). IARC monographs on the evaluation of carcinogenic risks to humans. Vol 90, pg 222 – 234 Human papillomaviruses, Lyon France.

Wiest T., Schwarz E., Enders C., Flechtenmacher C., & Bosch F.X. (2002). Involvement of intact HPV16 E6/E7 gene expression in head and neck cancers with unaltered p53 status and perturbed pRb cell cycle control**.** *Oncogene, 21*(10):1510-7.

Wilson, G. A., Lechner, M., Koferle, A., Caren, H., Butcher, L. M., Feber, A., … Beck, S. (2013). Integrated virus-host methylome analysis in head and neck squamous cell carcinoma. *Epigenetics, 8,* 953-961.

Wimardhani, Y. S., Sasanti, H., Wardhany, II, Sarsito, A. S., Pradono, S. A., Subita, G. P., . . . Iamaroon, A. (2015). Saliva-Based Screening of High-Risk Human Papillomavirus Strains: Detection in Female Indonesian and Thai Dental Students. *Asian Pac J Cancer Prev, 16*(13), 5525-5529.

Woo, S. B., Cashman, E. C., & Lerman, M. A. 2013. Human papillomavirus-associated oral intraepithelial neoplasia. *Mod. Pathol; 26* (10): 1288 – 97.

Woodman, C. B., Collins, S. I., & Young, L. S. (2007). The natural history of cervical HPV infection: unresolved issues. *Nat Rev Cancer, 7*(1), 11-22. doi:10.1038/nrc2050.

Yim, E.-K., Park, J.-S., (2005). The Role of HPV E6 and E7 Oncoproteins in HPV associated Cervical Carcinogenesis. *Cancer Res Treat, 37*(6), pp. 319-324.

Zerfass-Thome K, Zwerschke W, Mannhardt B, Tindle R, Botz Jw, Jansen-Durr P. (1996). Inactivation of the cdk inhibitor p27KIP1 by the human papillomavirus type 16 E7 oncoprotein. *Oncogene, 13:* 2323 – 30.

Zhao, M., Rosenbaum, E., Carvalho, A. L., Koch, W., Jiang, W., Sidransky, D., & Califano, J. (2005). Feasibility of quantitative PCR-based saliva rinse screening of HPV for head and neck cancer. *Int J Cancer, 117*(4), 605-610. doi:10.1002/ijc.21216

Zur Hausen, H. (2009). Papillomaviruses in the causation of human cancers—a brief historical account. *Virology, 384*(2), 260-265.

Zur Hausen H. (2002). Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer,* 2: 342 – 50.

Zur Hausen H. (2000). Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis. *J Natl Cancer Inst, 92:* 690 – 8.