When one depends on the other: The effect of tobacco smoking on oral human papillomavirus infection in the presence of human immunodeficiency virus in a US population.

Babatunde Yusuf Alli

Master of Science, Dental Sciences

The Faculty of Dentistry

McGill University

Montreal, Quebec, Canada

June 2020

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

DEDICATION

To Oreofe, Omolara, Atanda, and little Adam. In that order.

ACKNOWLEDGEMENTS

All praise and thanks be to Almighty Allah, the Lord of the Worlds. I would like to open this work by acknowledging the recently retired student affairs coordinator at the Faculty of Dentistry, Ms. Maria Palumbo - a great mother to us students. She literally single-handedly brought me to McGill and to this wonderful city of Montreal. Her emails and timely reminders about funding opportunities and deadlines were crucial to the successful completion of my program. Indeed, she was instrumental to my securing of the McGill graduate mobility award to travel to Roswell Park Comprehensive Cancer Center, Buffalo, NY to work on this project.

Equally important is to thank Dr. Belinda Nicolau, my teacher and supervisor. She is a friend and a great mentor to me and without her I will be lost. I am glad that I was able to rely on her immense support throughout my master's study at the Faculty of Dentistry. She introduced me to the MuCOSAL study and the international team behind it. Without her, this thesis would not have come to fruition. More importantly, I thank her for recognizing and understanding my potentials from the start thereby entrusting me with great responsibilities and opportunities, and for continuing to show me the ropes of academic research.

I would also like to thank Dr. Nicolas Schlecht, my host at Roswell Park and co-supervisor. My one-month research stays in his Lab at Buffalo for this project, is one of the highlights of my master's studies; and my gratitude goes towards his research staff for making my US sojourn smoother. I thank him for his time, patience, and mentorship all through the course of this project. I am also indebted to Dr. Marie-Claude Rousseau, my co-supervisor. My multiple trips to meet her in Laval and the ensuing discussions at the outset of this project introduced me to a world of possibilities for the MuCOSAL study.

I am particularly grateful to Dr. Sreenath Madathil for being such an outstanding individual, friend, colleague, and teacher. Brainstorming and bouncing ideas off him was perhaps my most rewarding experience of the past 2-years. I thank him for being at the ready to listen to all my odd ideas and fine-tuning them, and in fact, helping me along the way to initiate and implement some of them – culminating in a big chunk of this thesis.

My gratitude also goes to all the members of the Faculty of Dentistry, McGill University for contributing to a lively, conducive, and stimulating environment for oral health research; and to my colleagues in the Belinda's Lab, in particular, Amarjot and Radhika for being my pillar of friendship and support during my first Montreal winter. Also, I would like to acknowledge the Réseau de recherche en Santé Buccodentaire et Osseuse for awarding me a fellowship funding for my master's and McGill University for the International Graduate Mobility Award.

I am eternally grateful to my parents Mr. Nurudeen and Mrs. Afusat Alli for bringing me into this world and for managing to spark and nurture the intellectual fire in me, despite my early life academic struggles and growing up in the tough neighbourhoods of Bariga, Lagos, Nigeria. Last but not the least, my special thanks are in order to my wife, the love of my life, for her enduring love and rock-solid support – you are the reason I go forth.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

HIV	Human Immunodeficiency Virus
HPV	Human Papillomavirus
PLWH	People Living With HIV
HNC	Head and Neck Cancer
IARC	International Agency for Research on Cancer
RERI	Relative Excess Risk due to Interaction

ABSTRACT

Background: Human Papillomavirus (HPV) infection is an established risk factor for a distinct subset of Head and Neck Cancers (HNCs). People living with HIV (PLWH) are more likely to smoke and harbor oral HPV infections, putting them at higher risk for HPV-related HNCs. Whereas previous studies have shown tobacco smoking and HIV as independent risk-factors for oral HPV, the joint effects (i.e., effect modification/interaction) of these two exposures on the risk of oral HPV is not yet elucidated. Understanding how the effect of tobacco smoking on oral HPV is affected in the presence of HIV infection will be important for public health purposes.

Aim: The overarching objective of this master's project is to evaluate the modification of the effect of tobacco smoking on oral HPV infection by HIV status.

Methods: Consecutive PLWH (n=169) and at-risk HIV-negative individuals (n=126) were recruited from two large health centers in the United States. Lifetime smoking history was collected using questionnaires. Participants provided oral rinse samples for HPV genotyping by Next-Generation Sequencing. We used multivariable logistic regression models with interaction terms for HIV to test for smoking effect on oral HPV and reported the Relative Excess Risk due to Interaction (RERI). Then, we reported this following the recommended guideline for reporting effect modification/interaction analyses (Manuscript I). Motivated by the need to meet this recommended guideline for full reporting of effect modification, we developed a novel and easy to use package – *interactionR* – for the R statistical programming environment (Manuscript II).

Results: HIV infection positively modified the association between smoking and detection of high-risk oral HPV with odds ratios (OR) for smoking of 3.46 [95% CI: 1.01, 11.94] and 1.59 [95% CI: 0.32, 8.73] among PLWH and HIV-negative individuals, respectively, and an observed

RERI of 3.34 [95% CI: -1.51, 8.18]. The RERI for HPV-16 was 1.79 [95% CI: -2.57, 6.16], and 2.78 for β1-HPV [95% CI: -0.08, 5.65].

Conclusion: The effect of tobacco smoking on oral HPV is more pronounced among PLWH, making this population a desirable target for HPV-related HNCs prevention.

RÉSUMÉ

Contexte : L'infection par le virus du papillome humain (VPH) est un facteur de risque établi pour un sous-ensemble distinct de cancers de la tête et du cou (CNH). Les personnes vivant avec le VIH (PVVIH) sont plus susceptibles de fumer et d'héberger des infections orales par le VPH, ce qui les expose à un risque accru de CNH liées au VPH. Alors que des études antérieures ont montré que le tabagisme et le VIH étaient des facteurs de risque indépendants pour l'infection buccale à VPH, les effets conjoints (c'est-à-dire la modification des effets/l'interaction) de ces deux expositions sur le risque de VPH buccal ne sont pas encore élucidés. Il sera important pour la santé publique de comprendre comment l'effet du tabagisme sur le VPH buccal est affecté en présence d'une infection par le VIH.

But : L'objectif principal de ce projet de maîtrise est d'évaluer la modification de l'effet du tabagisme sur l'infection buccale à VPH en fonction du statut VIH.

Méthodes : Des PVVIH (n=169) et des personnes séronégatives à risque (n=126) ont été recrutées consécutivement dans deux grands centres de santé aux États-Unis. Les antécédents de tabagisme au cours de la vie ont été recueillis à l'aide de questionnaires. Les participants ont fourni des échantillons de rinçage oral pour le génotypage du VPH par Séquençage ADN haut débit. Nous avons utilisé des modèles de régression logistique multi-variables avec des termes d'interaction pour le VIH afin de tester l'effet du tabagisme sur le VPH buccal et avons rapporté l'excès relatif de risque dû à l'interaction (RERI). Ensuite, nous l'avons rapporté en suivant la ligne directrice recommandée pour la déclaration des analyses de modification des effets/interactions (Manuscrit I). Motivés par le besoin de respecter la directive recommandée pour la déclaration complète de la modification des effets, nous avons mis au point un nouvel outil facile à utiliser - *interactionR* - pour l'environnement de programmation statistique R (Manuscrit II).

Résultats : L'infection par le VIH a modifié de manière positive l'association entre le tabagisme et la détection du VPH buccal à haut risque, avec des rapports de cotes (OR) pour le tabagisme de 3,46 [IC 95% : 1,01, 11,94] et 1,59 [IC 95% : 0,32, 8,73] chez les PVVIH et les personnes séronégatives, respectivement,

et un RERI observé de 3,34 [IC 95% : -1,51, 8,18]. Le RERI pour le HPV-16 était de 1,79 [IC 95% : -2,57, 6,16], et de 2,78 pour β1-HPV [IC 95% : -0,08, 5,65].

Conclusion : L'effet du tabagisme sur le VPH buccal est plus prononcé chez les PVVIH, ce qui fait de cette population une cible souhaitable pour la prévention des CNH liées au VPH.

PREFACE

This thesis is written in a manuscript-based format. Per McGill University's thesis guidelines, this should chronicle a single and logically coherent research program. The two manuscripts included in this thesis fulfill this requirement. The first manuscript assessed for the effect modification of tobacco smoking's relationship with oral HPV by HIV while the second introduced a new tool for better reporting of effect modification and interaction analyses to the epidemiological research community. This tool – an R package – was developed during the preparation of the first manuscript.

The second manuscript is a single-authored article by the candidate; however, the first manuscript is the effort of multiple authors and their contributions are detailed below.

CONTRIBUTION OF AUTHORS

Manuscript I:

HIV modifies the effect of tobacco smoking on oral human papillomavirus infection

Babatunde Y. Alli, Master's Candidate: Conceived the investigation, designed, and carried out statistical analysis and wrote the manuscript.

Robert D. Burk, Professor, Departments of Pediatrics (Genetics), Microbiology & Immunology, Obstetrics Gynecology & Women's Health, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY, USA: Designed and conducted laboratory assays for HPV detection, contributed to the study design, and reviewed and contributed to manuscript writing.

Mahnaz Fatahzadeh, Associate Professor, Department of Diagnostic Sciences, Rutgers School of Dental Medicine, Newark, NJ, USA: Recruited participants, reviewed and contributed to manuscript writing.

Julie Kazimiroff, Associate Professor, Department of Dentistry, Albert Einstein College of Medicine and Montefiore Medical Center, New York, USA: Recruited participants, reviewed and contributed to manuscript writing.

Robert M. Grossberg, Associate Professor, Department of Infectious Diseases, Albert Einstein College of Medicine and Montefiore Medical Center, New York, USA: Recruited participants, reviewed and contributed to manuscript writing.

Richard V. Smith, Associate Professor, Department of Otorhinolaryngology-Head and Neck Surgery, Albert Einstein College of Medicine and Montefiore Medical Center, New York, USA: Recruited participants, reviewed and contributed to manuscript writing.

Thomas J. Ow, Associate Professor, Department of Otorhinolaryngology-Head and Neck Surgery, Albert Einstein College of Medicine and Montefiore Medical Center, New York, USA: Reviewed and contributed to manuscript writing.

Mauricio Wiltz, Associate Professor, Department of Dentistry, Albert Einstein College of Medicine and Montefiore Medical Center, New York, USA: Recruited participants, reviewed and contributed to manuscript writing.

Jacqueline Polanco, Associate Professor, Department of Dentistry, Albert Einstein College of Medicine and Montefiore Medical Center, New York, USA: Recruited participants, reviewed and contributed to manuscript writing.

Marie-Claude Rousseau, Professor, Epidemiology and Biostatistics Unit, Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique (INRS), Laval, Quebec, Canada: Co-supervised the candidate, Principal Investigator, MuCOSAL Study (Canada site), contributed to statistical analysis, reviewed and contributed to manuscript writing.

Belinda Nicolau, Professor, Faculty of Dentistry, McGill University, Montreal, Canada: Supervised the candidate, Co-Principal Investigator, MuCOSAL Study (Canada site), reviewed and contributed to manuscript writing.

Nicolas F. Schlecht, Professor, Department of Cancer Prevention and Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA: Co-supervised the candidate, Principal Investigator, MuCOSAL Study, contributed to the design of the investigation and statistical analysis, and co-wrote the manuscript.

1 INTRODUCTION

Human Papillomavirus (HPV) infection is an aetiologic factor for a distinct subset of Head and Neck Cancers (HNCs) (1). It's responsible for approximately 25% of all HNCs worldwide (2), with an ever-increasing responsibility for a larger share of incident HNCs in North America (3, 4), including in particular cancers originating in the oropharynx, base of the tongue, and palatine and lingual tonsils.

People living with HIV (PLWH) have a higher prevalence, incidence, and persistence of oral HPV (Alpha (α) types) (5, 6). Similar to most cancers, they are also at higher risk of HPV-related HNCs (7). In addition, tobacco smoking – a strong and independent risk factor for HNCs that is highly prevalent among PLWH (8) – is also an important risk factor for oral HPV (9, 10). Whereas previous studies have reported differences in risk factors for oral HPV in PLWH compared with HIV-negative individuals (5, 11), the interaction of tobacco smoking and HIV on the risk of oral HPV has not been explored (5). Furthermore, most studies often only report interaction on a multiplicative rather than in an additive scale(12), which is surprising considering that the latter is more relevant to public health (12, 13). A possible explanation lies on the ready availability of estimation of multiplicative interaction as the exponentiated coefficient of interaction terms from the regression outputs of any standard statistical software. The fact that estimation of additive interaction requires further steps and calculation may deter reporting of estimates necessary for full evaluation of interaction by the reader.

This thesis investigates the modification of the effect of tobacco smoking on oral HPV outcomes by HIV seropositivity on the additive scale. To do so, we developed a R statistical software package for full reporting of effect modification and interaction.

2 LITERATURE REVIEW

This section includes a narrative literature review on HPV, tobacco smoking and HIV. It opens with a description of the phylogenetic classification of HPV, emphasizing the most relevant types to head and neck carcinogenesis and some specific anatomic features that especially predisposes some parts of the head and neck region to its infection and propagation, and finishes with a primer for effect measure modification.

2.1 Human Papillomaviruses

The papillomaviridae is a family of non-enveloped circular double-stranded DNA viruses of about 8kb in size (14). They exhibit strict species-specific tropism, which means, for example, papillomavirus (PV) that infects rabbits cannot infect cats (or any other specie for that matter) (15). This family of viruses is phylogenetically classified into genera, species, and types (Figure 1). For example, HPV-4 is a *type* under Gamma-1 *specie* belonging to the gammapapillomavirus *genus* (Figure 1). Three main genera of the papillomaviridae infects humans: alphapapillomavirus (α), betapapillomavirus (β) and Gammapapillomavirus (γ) (Figure 1) (14). Of these, the α-genus is the most extensively studied, being responsible for approximately 90% of all HPV-related HNCs (16), with HPV-16 as the most prominent type. Alongside HPV-16, other carcinogenic α-genus types include 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59; together these are designated as high-risk HPV by the International Association for Research on Cancer (IARC) for their high propensity for anogenital carcinogenicity (17).

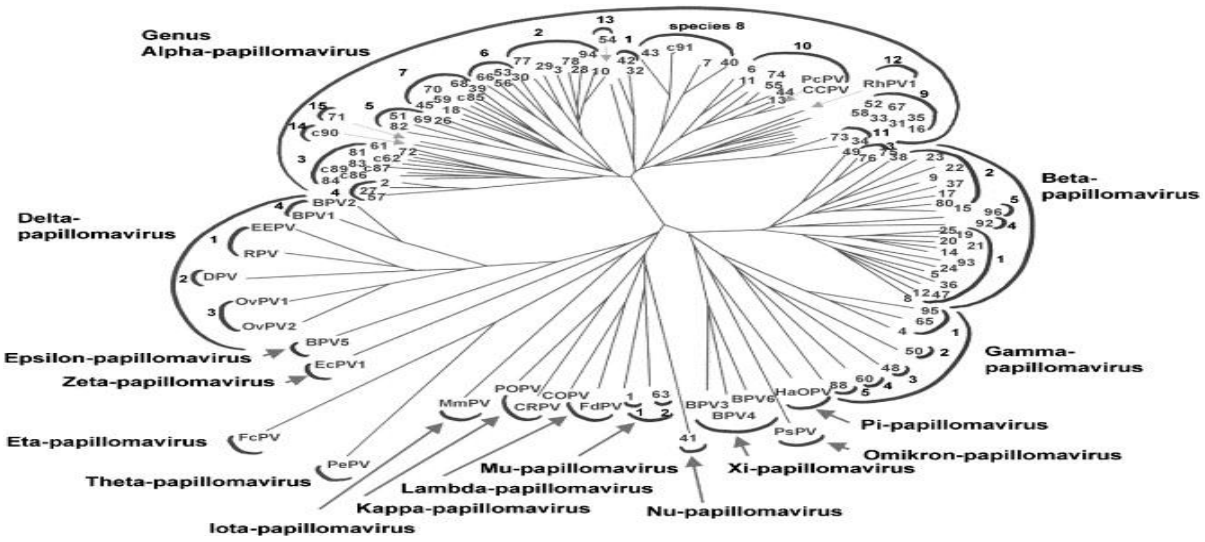

Figure 1: Phylogenetic classification of the Papillomaviridae (14).

Until recently, viruses in the β and γ genera were not thought to infect the mucosae (6); in fact, they were colloquially referred to as the 'cutaneous HPV types' (18). However, they are now not only known to infect the oral mucosae (6), but some of them have been associated with HNCs (17, 19). A recent prospective study conducted in the US reported an association between β1-HPV-5 type, γ11- and γ12-HPV species and HNCs (19), thus expanding the HPV-related HNCs' viral landscape.

Despite all HPVs sharing the same molecular structure, the main difference between the types that causes cancer and those that do not is in the carcinogenic potential of their oncoproteins (e.g., E6 and E7; to be reviewed later)

2.1.1 Brief History of HPV

Skin and genital warts (papilloma) have been recorded for millennials (20), and even from this ancient period, it was associated with sexual promiscuity, hence considered potentially infectious. The first experimental confirmation of the infectious and viral nature (21) – a virus later identified

as an HPV- was provided in Italy in 1907. Subsequently, in 1949, Strauss et al. demonstrated viral particles in skin papilloma for the first time with electron microscopy (22).

In 1842, the Italian physician Rigoni-Stern made the astute observation of a sexually transmitted-basis of cervical cancer (23). He observed that compared to a rarity among nuns and virgins, cervical cancer was highly frequent among prostitutes, widows, and married women. Although the oncogenic potential of PV was demonstrated in domestic rabbit as early as 1934 (24), and confirmed in cottontail rabbits in 1944 (25), HPV role in the aetiology of cervical cancer was largely overlooked in favor of the Herpes Simplex Virus (HSV). The latter was more associated with 'promiscuity' than HPV probably because the progression to cancer of the ubiquitous human wart was not yet observed at that time. However, the combined lack of evidence from both molecular and epidemiological studies to support the HSV hypothesis (26, 27), and the anecdotal reports of malignant transformation of genital warts into squamous cell carcinomas turned the attention of some researchers to HPV (28, 29). In 1983, zur Hausen and co-workers isolated the novel HPV from a cervical cancer biopsy and designated it HPV-16 (30). About a year later, they reported another novel PV, HPV-18 (31), providing the first set of mechanistic evidence for the oncogenicity of HPV – an achievement that earned zur Hausen a shared Nobel prize in Medicine in 2008 (32). In short order, this evidence was rapidly corroborated by other mechanistic evidence such as the demonstration that the E6 and E7 oncoproteins of these high-risk HPV types cooperate to immortalize human epithelial cells such as cervical keratinocytes (33).

Notwithstanding this overwhelming mechanistic evidence for HPV oncogenicity, the early epidemiologic evidence was weak at best (34). Franco (35) attributed the disparity in the strength of evidence between the mechanistic and epidemiologic to severe misclassification bias in the epidemiologic studies, due to differences in technology employed for HPV DNA testing.

Subsequently, epidemiologic studies testing with more sensitive methods such as the Polymerase Chain Reaction (PCR) corroborated the overwhelming evidence from laboratory studies (36). Based on these body of evidence, the IARC classified HPV 16 and 18 as a Group 1 human carcinogen in 1995; other HPV types was added to the high-risk group in the subsequent monographs (37). It is now known that HPV is a necessary cause of cervical cancer (i.e., 100%), and also responsible for about 25% of all HNCs (2). In all, HPV is responsible for 5% of all human cancers.

2.1.2 Molecular mechanism of HPV carcinogenicity

To understand the carcinogenic pathway of HPV, its important to clarify its molecular structure. As previously mentioned, HPVs are double-stranded DNA viruses. Contained within a non-enveloped icosahedral capsid, this DNA encodes 8 open reading frames (ORF) that can be subdivided into three functional parts: the early (E) region, which contains information for the production of the early proteins (E1-E7) responsible for the replication of the virus; the late (L) region, that encodes information for the proteins (L1-L2) necessary for virion assembly; and the long control region (LCR), a non-coding section containing important elements for the regulation of viral expression (38-40) (Figure 2). Through these molecular actors (the E and L proteins especially), HPV exerts its influence on important pathways in the hallmark of carcinogenesis (38, 41).

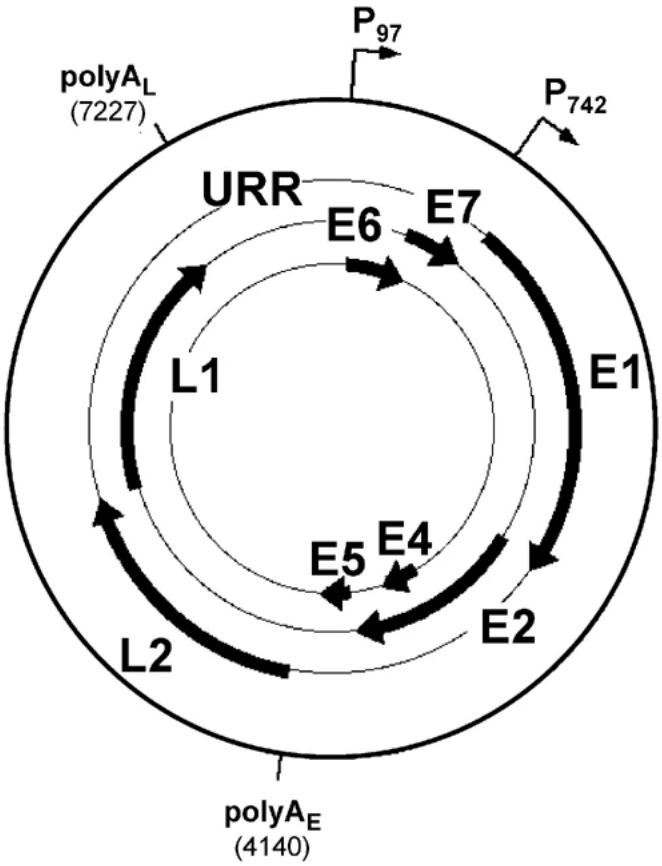

Figure 2: Circular double-stranded DNA of HPV showing the ORFs. Image reproduced with permission from reference (40).

HPV is highly epitheliotropic (i.e., it preferentially infects the mucosa and cutaneous epithelium) and the human keratinocyte is its primary host (37, 42). In a normal human epithelium, the population of keratinocytes are arranged in successive compartments from the base to the surface according to their stage of cellular differentiation: the basal cells are the proliferating pool while the cells at the surface are terminally differentiated (43). During basal cell proliferation, cells that lose contact with the basement lamina move to the supra-basal compartment exiting the cell cycle and becoming committed to terminal differentiation. As the cell moves through the epithelial

layers, it becomes more matured until it reaches the surface as an adult keratinocyte (43, 44). The HPV life cycle is tightly integrated to this process of epithelial differentiation (45, 46). HPV infection and propagation require cell division, thus the virus cannot infect a committed keratinocytes, only basal cells (47). To generate infection, the virus must interact with these basal cells; first, HPV binds to the epithelial basement membrane through the L1 major capsid protein (48, 49), which is followed by gaining entry into the basal cells through L2 minor protein.

Once infection is established, the E6 and E7 oncoproteins induces a state of genomic instability, leading to multiple uncontrolled breaks in the double stranded DNA of both the host and the virus allowing the integration of these oncoproteins in the host's genome (50). Once integration is achieved, the persistent expression of E6 and E7 further interferes with the normal regulators of the cell cycle. First, E6 binds to p53 degrading this crucial tumour suppressor protein thereby deregulating the cell cycle and allowing unchecked proliferation of the basal cells (51). In parallel, E7 binds to the retinoblastoma protein (pRb), another tumour suppressor, dissociating it from the E2F transcription factor that drives the cell unchecked through the cell cycle (52). In addition, E6 and E7 inhibits the suppressors of the human telomerase reverse transcriptase (hTRT), an enzyme that prevents the shortening of the cell telomeres, thereby rendering the cell immortal (53). Furthermore, these oncoproteins, alongside E5, targets differentially, non-coding small microRNAs (miRNA) by downregulating the tumour suppressive miRNAs, and upregulating the tumour promoters (54-57).

Taken together, these viral oncoprotein activities drive the host cells into cancers, and their persistent expression is required to keep the cells in a carcinogenic state (58). To compound this, E6, E7 and to some extent E5 proteins, further act to help these immortalized and uncontrollably dividing cells to evade the immune system, preventing them from destruction. First, E6 and E7

inhibits the interferon immune signalling pathway by binding to the factor-1 and -3 proteins of the interferon regulatory system. Also, E7 binds to a crucial death receptor on the cell surface, thereby inhibiting the FAP-mediated cell apoptosis.

The above described pathways are shared by all carcinogenic HPVs, although have been characterized mainly through the extensive study of HPV-16 and -18. However, recent investigations show one important difference for beta genus HPV-5 and HPV-8 (See Figure 1). Whereas persistent infection by alpha high-risk HPV (mainly through the constant expression of E6 and E7 oncogenes) is required to sustain a cancerous state, beta HPVs appears to act only at the beginning of the carcinogenic process in the so-called hit-and-run mechanism (58, 59). In fact, the DNAs of these HPVs are rarely found in skin cancer samples; they achieve this by acting as a co-factor, potentiator and accumulator of ultraviolet-induced DNA damage in cells by derailing p53 and Notch genes at the inception of carcinogenesis. Once this is achieved, E6 and E7 becomes dispensable (58-60). Put simply, they merely facilitate the induction of carcinogenesis, but are not required for the maintenance of the cancer phenotype.

2.1.3 HPV-related HNCs

HNCs represents approximately 4.8% of all cancer incidence and mortality worldwide (61). The vast majority of these cancers (>90%) arise from the squamous epithelium that extensively lines the mucosae of the head and neck region. This includes: (i) the oral cavity, including the lips, tongue, buccal mucosae and the floor of the mouth; (ii) the oropharynx, including the base of the tongue, lingual and palatine tonsils, and the back of the throat; and (iii) the larynx, including the supraglottic, glottic, and subglottic larynx. The epithelial tropism of HPV makes it well suited to infect these tissues (37, 39). For the remainder of this thesis, HPV that occurs in the oral cavity, oropharyngeal and laryngeal region is referred to as oral HPV.

Despite the decreasing incidence of tobacco smoking related HNCs in the developed world (due to the decreasing prevalence of smoking habit), the incidence of HPV-related HNCs is on the rise (62). As mentioned previously, approximately 25% of these cancers worldwide are now due to oral HPV infection (2), with HPV-16 and -18 responsible for about 85% of them (2, 63). The distribution of HPV-related HNCs worldwide varies by anatomical site with the oropharynx harboring the majority of them (33.6%), followed by oral cavity (22.2%) and larynx (20.2%) (2, 63). However, the proportion of HNCs attributable to oral HPV by anatomical sites differs markedly from North America. The attributable fraction (AF) for oral HPV in the US (all HPV types) and Canada (for oral HPV-16 alone) are 70.1%, 32% and 20.9% and 60.2%, 8.2%, and 12.7% for cancers of the oropharynx, oral cavity and larynx, respectively, (64) (65).

2.1.4 Epidemiology of oral HPV

HPV is the most common sexually transmitted infection(67); the majority of individuals will have at least one incident α-HPV infection of any type in their lifetime (66). The worldwide prevalence of α-types oral HPV in healthy individuals varies between 4.5%-4.9%, with a prevalence of 2.6%-3.9% and 1%-1.3% for high-risk types and HPV-16, respectively (68-71). Although estimates for different oral HPV types varies in different parts of the world, oral HPV-16 is the most prevalent singular type detected in all regions (68). However, the global prevalence for oral HPV from β and γ genus is low with an estimation of 0.7% and 1.8%, respectively (68). It is suggested that the prevalence is underestimated (6, 19).

The estimate for overall prevalence in the US general population is 6.9% for any α-type oral HPV, 3.7% for all high-risk alpha types, and 1% for HPV-16 (72). This prevalence, apparently low in the general population, dramatically increases in at-risk populations, especially PLWH, with an estimated prevalence of 16%-60% for oral α-HPV (6, 11, 73-76). Whereas population prevalence

estimates for and γ genus HPV are lacking, evidence to date also suggests detection is also higher among PLHW (6).

In addition, oral HPV exhibits a bimodal age distribution in the general population, with multiple studies reporting an initial peak among 30-34 years another peak at 60-64 years (77, 78). The first peak may represent peak sexual activities, while the second peak potentially represents a reactivation of latent infection probably due to immunosenescence (79). Finally, there is evidence for gender difference in this prevalence, with the overall prevalence of oral HPV 2.8 times higher (10.1% vs 3.6%), high-risk oral HPV prevalence 5.3 times higher (12.7% vs 3.6%) and oral HPV-16 six times higher (1.8% vs 0.3%) in men than in women (78, 80).

2.1.5 Risk factors for oral HPV

Sexual activity is the primary risk factor for the acquisition of oral alpha HPV (81), with increases in incidence and prevalence observed with higher lifetime number of sexual partners, oral sexual experience (ever or never), and lifetime number of oral sexual partners (4, 73, 78, 82, 83). The oral sexual transmission route is mainly attributed to men performing oral sex on women, which is supported by the observation of a higher oral HPV infection in heterosexual men compared to men who have sex with men (MSM) (84). The association between sexual activity and oral HPV may partly explain why men are at a higher risk of oral HPV compared to women – i.e., men are more likely to have more lifetime sexual and oral sexual partners (82). Gender differences in immune responses mounted against an oral HPV infection may also explain the high prevalence of these viruses among males (85).

Beyond sexual activities and gender, other risk factors for oral HPV include age, race, tobacco smoking and HIV (9, 75, 78, 80). The latter two risk factors are discussed in the following sections.

2.2 Tobacco smoking

Used for millennia, tobacco consumption is the preeminent human habit (86). The most common form of this habit is the smoked tobacco, predominantly, cigarette smoking (87). Globally, over 1.1 billion people are current smokers (942 million men and 175 million women) (88). Although smoking prevalence is reducing in the developed countries, this is offset by an increasing prevalence in the developing world (89).

Cigarette smoking was first associated with any form of cancer in 1950 (90, 91), and since then has been established as the primary aetiology for a wide-range of cancers affecting multiple organs of the human body, including the head and neck region (92).

2.2.1 Association of tobacco smoking and oral HPV

The evidence in the literature also show tobacco smoking as a risk-factor for oral HPV infection. Multiple epidemiologic studies have demonstrated an association between smoking and oral HPV, including a study showing that smoking increases the persistence of incident oral HPV infection in men after a follow-up period of 7 years (5, 9, 10, 78).

The mechanistic evidence for this is not far-fetched. As mentioned previously, to establish an infection, HPV requires access to the epithelial basement membrane. Cigarette smoking extracts have been demonstrated to damage epithelial integrity (93), which will facilitate virus entry. Also, HPV, unlike most other viruses, has a prolonged replication rate (94). Therefore, it needs to persist in the epithelium for a long time without being detected by the host's immune system. The virus has developed a complex immune signalling inhibition system (briefly reviewed in section 2.1.2) that helps to achieve this. In addition, the chronic state of immunosuppression and immunomodulation induced by cigarette smoking may contribute further to this function (95).

2.3 Human Immunodeficiency Virus (HIV)

HIV is one of the main causes of morbidity and mortality worldwide (96). Approximately 37.9 million people live with the infection globally (97), with 1.7 million incident cases in 2018 (97). Although the largest burden of this disease is concentrated in low and middle-income countries (98), developed countries are not spared with an estimated over 1.1 million PLWH in the US (99).

2.3.1 Risk of HPV-related HNCs in PLWH

As with most other malignancies affecting humans, PLWH are at a higher risk of HPV-related HNCs. This has been attributed to immunodeficiency, immune dysfunction/senescence, chronic inflammation and some unique behavioral and biological characteristics specific to this population (100). Compared to the general population, PLWH have a standardized incidence ratios (SIRs) of 1.5 to 4-fold higher risk in HPV-related HNCs (101-104). However, the majority of these SIR estimates are from registry-based studies where confounding variables other than age and smoking could not be adjusted for. In a study in which important confounders (e.g., smoking) was adjusted for, the risk disparity between PLWH and the HIV-negative population was more modest at 1.4-fold higher (105).

2.3.2 Association of HIV and oral HPV

The prevailing factors that makes PLWH at higher risk of human cancers also makes them predisposed to oral HPV infections. Indeed, as previously mentioned, PLWH have a remarkably higher prevalence of oral α-HPV compared to the general population (6, 11, 73-76). In addition, a study reported that HIV infection increases the risk of incident oral HPV infection (5). However, results from the same study also showed that HIV does not influence the clearance of incident infections and postulated that the very high prevalence of oral HPV in PLWH is due more to the early involvement of HIV in establishing an HPV infection, rather than its persistence. The

findings that HIV proteins disrupt tight mucosal epithelial junctions thereby facilitating entry and easy access of HPV to the basement epithelium supports this assertion (106, 107).

In the HAART era, the burden of the vast majority of opportunistic infections in PLWH has been reduced. However, this ameliorating effect does not extend to oral HPV infections. In fact, opposite findings have been reported, that is, an increase in oral HPV DNA detection even after antiviral therapy initiation (108). This suggests that oral HPV infection (and its associated malignancies) will continue to be a significant public health issue among PLWH even in the HAART era (7).

2.4 Overview of Effect Modification

As described above, while tobacco smoking and HIV are strong individual risk-factors for oral HPV, their effects acting together (or in the presence of one another) on oral HPV might be stronger (or weaker) than the mere sum of their individual effects: this is known as effect modification. In this section, we briefly introduce the concept of effect modification and/or interaction.

2.4.1 Effect Modification – A Primer

Effect modification occurs when the effect of an exposure on an outcome differs within the strata of another exposure (109); for example, result from the Million Women Study shows that the use of hormone replacement therapy (the main exposure) reduces the risk of endometrial carcinoma (the outcome) only in women with a BMI (the second exposure) of ≥ 30 kg/m^2 and not in non-obese women (110). Effect Modification is sometimes distinguished from interaction, in which the joint effect of two independent exposures on an outcome is assessed against their separate individual effects (111); for example, using data from a case control study, Rothman reported a

higher joint effect of the combination of alcohol and smoking on oral cancer than their individual effects (112). However, in practice, both terms are often used interchangeably (111).

Importantly, the assessment of effect modification/interaction is scale dependent: multiplicative or additive (109). On the multiplicative scale, effect modification is assessed based on how much the combined effects of the two exposures differs from the *product* of the exposures considered separately: this employs ratio measures such as risk ratio (RR) and odds ratio (OR) (113). On the other hand, the additive scale is when effect modification is based on how much the joint effect differs from the *sum* of the individual exposures using risk difference (RD) (113).

In epidemiological practice, effect modification/interaction is assessed by introducing a product term for the two exposures under study in a statistical model (113). The coefficient of this product term (also called the 'interaction' term) and its confidence interval (CI) is the statistical interaction (113). For example, let S represent smoking and A represent alcohol and we are to assess the effect of their interaction on oral cancer (OC). Fitting a linear risk model for RD[1], log-linear model for RR[2], and a logistic model for OR[3], respectively, we would have:

$$P(OC|S,A) = \beta_0 + \beta_1 S + \beta_2 A + \beta_3 S*A \qquad [1: RD]$$

$$\text{Log } (P(OC|S,A)) = \beta_0 + \beta_1 S + \beta_2 A + \beta_3 S*A \qquad [2: RR]$$

$$\text{Logit } (P(OC|S,A)) = \beta_0 + \beta_1 S + \beta_2 A + \beta_3 S*A \qquad [3: OR]$$

In model [1], the coefficients β_1 and β_2 represents the individual absolute risks of smoking and alcohol on oral cancer respectively, while β_3 represents the statistical interaction of these two exposures. However, because this is an absolute risk model (i.e., RD), then β_3 is the statistical interaction on the additive scale. For model [2] and [3], the exponents of the coefficients β_1 and β_2 represents the individual effects of smoking and alcohol respectively, while exponentiated β_3 is the

statistical interaction on the multiplicative scale, because the models are based on ratios. Crucially, the phenomenon of statistical interaction on either of the two scales only represents effect modification in the absence of bias (109).

Furthermore, it has been suggested that interaction on the additive scale, unlike multiplicative scale, is more consistent with biological interaction as encoded in the sufficient cause model proposed by Rothman (109, 113). Also, an important motivation for assessing effect modification is to identify a high-risk subgroup for public health interventions. This can be achieved on the additive scale rather than the multiplicative scale (111, 113-115), making the former more important from the perspective of public health.

However, most statistical modelling in the medical literature is performed by estimating ORs from logistic regression models (a multiplicative scale) because linear risk and log-linear models often run into problems with convergence in the presence of covariates (113), making additive interaction not readily accessible. A way around this is computing the relative excessive risk due to interaction (RERI) which estimates additive interaction from multiplicative measures (RR or OR) (116). The RERI – first introduced by Rothman (117) – is now the standard approach in the literature for the evaluation of additive interaction.

3 RATIONALE

HPV-related HNCs is a growing public health concern in North America (3, 118). The estimated incidence and deaths for these cancers in 2020 in the US alone is 53,260 and 10,750 cases, respectively (119). They are one of the most expensive cancers to manage, imposing a heavy financial toll on the healthcare system (120).

Secondary prevention through screening by HPV testing has been important in the fight against cervical cancer (121) – another HPV-related human cancer. However, unlike cervical cancer, HPV testing is not yet a useful – whether primary or adjunct – screening tool for HPV-related HNCs (122). This is because of the previously mentioned low prevalence of oral HPV in 'healthy' individuals (i.e., people that do not already have a diagnosis of HNCs) (122). For HNC screening based on oral HPV testing to be viable, a high-risk subgroup for oral HPV needs to be identified and characterized (122, 123).

PLWH are possibly one such group; as compared to the general population, they have a remarkably higher prevalence of oral HPV infection. What is more, they also have a disproportionately higher prevalence of tobacco smoking habits – another risk factor for oral HPV infection – compared to the general population (124, 125). So, although previous studies have demonstrated that HIV infection and tobacco smoking are independent risk factors for oral HPV infection, and that there is a higher prevalence of smoking habits in PLWH, no previous study have investigated how the presence of these two exposures (smoking and HIV) together in an individual contributes to oral HPV infection. Manuscript I address this gap in the literature by investigating the effect modification of the tobacco smoking – oral HPV relationship by HIV infection.

In addition, the reporting of effect modification/interaction analysis is inadequate in the medical literature (12). For example, among other things, many authors present interaction only on the

multiplicative scale, despite the fact that additive interaction is more relevant to public health (12, 13, 112). To remedy this, Knol and Vanderweele proposed a set of recommendations for the reporting of interaction analyses – which was an improvement on previous guidelines – in order to enable readers to fully evaluate the reported interaction (115). However, to abide by these recommendations (as was performed in Manuscript I), some additional steps needed to be taken to compute effect estimates (such as RERI) that are not readily available from the output of regression models of standard statistical software such as R (126). These additional steps could have deterred authors from conducting in prior studies (127). Manuscript II presents a novel and easy to use R package that enables one to compute and report interactive effect estimates like RERI; this package was developed in the course of preparing Manuscript I. The package improves on existing R functions (127, 128) that only allow for computation of measures of additive interaction and do not provide the remaining estimates to fulfill the aforementioned recommendations. It also, for the first time, enables the estimation of CIs for additive interaction measures in R using multiple approaches including alternatives to the delta method (129).

4 HYPOTHESIS AND OBJECTIVES

We hypothesize that the effect of tobacco smoking on oral HPV infection will be higher in PLWH compared to HIV-negative individuals, and that the prevalence of all types of oral HPV (not just α-types) will be higher in PLWH.

The primary objectives of this thesis are:

1. To evaluate the modification of the effect of tobacco smoking on oral HPV infection by HIV status (Manuscript I).
2. To estimate the prevalence of alpha, beta, and gamma oral HPV types by HIV status in an at-risk population in New York (Manuscript I)

The secondary objective is:

1. To develop and test a novel R package for full reporting of interaction (Manuscript II)

5 METHODS

5.1 Study Design

Data from the Multicenter Oropharyngeal Squamous Atypical Lesion (MuCOSAL) study was used for this project. The MuCOSAL project, an international prospective cohort study conducted in the US, Canada and Brazil, was set up to study the natural history of potentially malignant oral lesions. Participants with and without oral lesions were recruited from large health centers in cities in these three countries and followed-up for a year with three visits: baseline, 6 months, and 12 months. However, only the US site recruited PLWH and at-risk HIV-negative individuals, therefore, the baseline data for this site was used for this project.

5.2 Study Setting and Population

Between 2004 and 2013 inclusive, 177 PLWH and 148 at-risk HIV negative individuals for a total of 325 people with and without oral lesions were recruited from the Infectious Disease, Otolaryngology, Dental and Oral Medicine outpatient clinics of the Montefiore Medical Center (Bronx, NY) and Rutgers School of Dental Medicine (Newark, NJ). Where possible, HIV-negative participants were matched to PLWH by age (±10 years).

5.3 Eligibility Criteria

To be eligible, participants had to be 21 years of age and above with no history of treatment for cancers nor oral lesions in the preceding year

5.4 Ethical Approval and Informed Consents

The MuCOSAL study protocol was approved by the Institutional Review Boards of the participating hospitals and all the participants provided written informed consents (Appendix I – Study Consent Form) when they were recruited into the study.

5.5 Data Collection

5.5.1 Recruitment

A research nurse at the participating clinics approached attending patients to perform an initial eligibility screening based on the inclusion criteria. Those eligible were invited to participate and those who agreed filled the informed consent forms. The research nurse then carried out a baseline interview, after which each participant underwent a comprehensive oral examination a follow-up visit was scheduled.

5.5.2 Study Instruments

5.5.2.1 Interview

The interview questionnaire (Appendix I – Study Questionnaire) was adapted from instruments used by the IARC. Participants underwent a detailed structured interview that collected information on socio-demographic characteristics including age, sex, race/ethnicity and occupation; lifestyle habits including cigarette smoking and alcohol consumption; sexual practices and history (e.g. lifetime number of sexual partners, practice of oral sex); and history of various sexual-transmitted diseases.

5.5.2.2 Clinical Oral Examination

Each participant underwent a comprehensive oral examination carried out by the participating dentists. Presence of oral lesions were noted and described. If oral lesions were present, participants were then classified based on presentation of the most severe oral lesion, including benign/papilloma (e.g., oral warts, frictional keratosis), and potentially malignant lesions (e.g., lichen planus, leukoplakia and/or erythroplasia).

5.5.2.3 Oral Rinse

At each visit, each participant provided a 30-second gargle and oral rinse sample with Scope® mouthwash for HPV genotyping. The samples were kept on ice until stored at -20°C within six hours of collection and thereafter transferred to the lab for DNA isolation.

5.6 Definition of Variables and Measurements

The following section describes the definition of the outcome and the main exposure variables alongside the definitions and selection of covariates a priori with directed acyclic graph (DAG).

5.6.1 Oral HPV

Oral HPV is the outcome for this study. HPV DNA detection and typing was performed using a next-generation sequencing (NGS) assay developed by the New York research team (130). DNA extracted from the oral rinse sample and brush biopsy specimens were tested for the presence of α-, β- and γ-HPV types using two different assays; (i) PCR based using primers targeting specific sequences (or short PCR fragments) within the HPV genome (130), which were published previously (131); and (ii) NGS which the results were the focus of the current project. These NGS assay operates similarly to the 16S microbiome assays that amplify a specific fragment using bar-coded primers. For α-, β- and γ-HPV detection, a unique 12 bp Golay DNA barcode was assigned to each subject and introduced into the PCR amplicons. Barcoded PCR products from multiple subjects were pooled at approximately equal molar DNA concentrations and sequenced on an Illumina platform (Illumina, San Diego CA). The millions of reads produced by NGS were de-multiplexed with individual samples being reconstructed by sorting barcodes into unique files. The reads were then matched with known or novel HPV sequences using BLAST, and a specialized HPV database developed by our lab. Positive and negative controls were included with each PCR, as well as samples previously identified to have a representation of different α-, β- and γ-HPV

types. Identification of individual HPV types was based on parameters including total read number, percent of total HPV reads and distribution of reads for each HPV type.

For the specific outcomes, HPV was classified phylogenetically into α, β and γ HPV genera, and the relevant species corresponding to these genera: α-3, α-9; β-1 , β-2, β-3; γ-7, γ-8, γ-9 (14). Furthermore, we grouped together α-HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59 as high-risk α-types following the IARC classification (132), and we singled out α-HPV-16 – for its predominant role in HNCs – as well as the recently implicated β-1 species types, HPV-5 and HPV-8 (17, 19).

5.6.2 Tobacco Smoking

Tobacco smoking was the main exposure variable. Detailed smoking history was collected by asking participants to divide their lifetime consumption into varying periods of duration and intensity. For each period, information on age at start, age at cessation, and the number of cigarettes smoked per day in that period was collected. From this, smoking duration and intensity was encoded into pack-years computed as the number of cigarettes smoked daily divided by 20 and multiplied by the total duration smoked in years. One pack-year of tobacco smoking is equivalent to smoking 20 cigarettes per day for one year or smoking 40 cigarettes daily for half a year and so on. Ever-smokers were defined as those who smoked tobacco for at least one year at any point in their lives.

5.6.3 HIV

HIV infection was the putative effect modifier. Participants were identified by self-report with history of HIV infection confirmed through their medical records.

5.6.4 Covariates and Directed Acyclic Graph (DAG)

Covariates to be included in the final statistical model assessing for the relationship between the variables mentioned above were selected a priori using causal directed acyclic graph (DAG). A priori model specification with DAGs is widely used in the epidemiological literature to control for confounding. It has been shown to be superior to other methods such as significant testing in univariate analysis, step-forward nested model selection and the change-in-estimates approach when valid a priori assumptions can be formulated (133-135), therefore a brief introduction is in order.

DAGs use arrows drawn in a timeline that characterize the causal and temporal relationship between variables (Nodes) (136). They are directed, which means that each line has a single arrowhead emanating from a variable indicating the variable's effect on another; for example, 'X → Y' indicates that variable X effects Y. They are acyclic, which means there are no feedback loops of arrows because a variable cannot be it's own descendant in time – due to temporality (136); for example, 'X → Y → A → X' is not a valid DAG because of the feedback loop from variable A back to X. From these simple building blocks, we can encode our assumptions of how variables relate to each other. A hypothetical example (Figure 3) is presented for illustration.

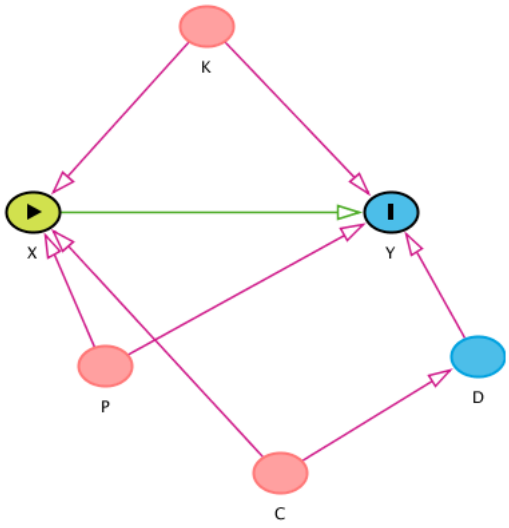

Figure 3: Hypothetical Directed Acyclic Graph

Figure 3 indicates that X is a cause of Y (or Y is a descendant of X), K and P are common causes of X and Y (i.e., classic confounders), and though C is a direct cause of X, it's an indirect cause of Y through the mediator D. Any path that leaves X through arrows entering it and ends in Y satisfies the so-called "backdoor" criterion (136). To ensure that the effect of X on Y is not confounded, all backdoor paths must be blocked through covariates adjustments, leaving the direct path from X to Y (134, 136). In our example, these backdoor paths are: X-K-Y; X-P-Y; and X-C-D-Y, and each of these paths will be blocked if variables in the path are controlled or adjusted for. The minimum set of variables sufficient to completely control for confounding in our example is {K, P, C} or {K, P, D}.

5.6.4.1 Number of Sexual Partners and History of Oral Sex

As reviewed earlier, lifetime number of sexual partners and oral sexual history is a risk factor for oral HPV (4, 73, 78, 82, 83). This was collected using the "intimate relationship and marriage" section of the questionnaire instrument. Self reported lifetime number of sexual partners was

recorded and categorized into ≤5, 6-20, and ≥20. Oral sexual history was recorded as "ever" for participants that reported performing oral sex at least once in their lifetime, and "never" otherwise.

5.6.4.2 Sex, Age, Race and Oral Lesions

Other important variables of interest included age, sex, black race, and the presence of oral lesions, all of which are associated with oral HPV in the literature (9, 78, 80).

5.6.4.3 The Directed Acyclic Graph (DAG)

Taken together, these covariates are considered possible confounders of the smoking-oral HPV relationship in this cross-sectional analysis. This is shown in the causal DAG below (Figure 4) and represent the minimally sufficient adjustment set of variables required to de-confound the relationship between smoking and oral HPV detection. In the DAG, HIV is a direct effect modifier (114).

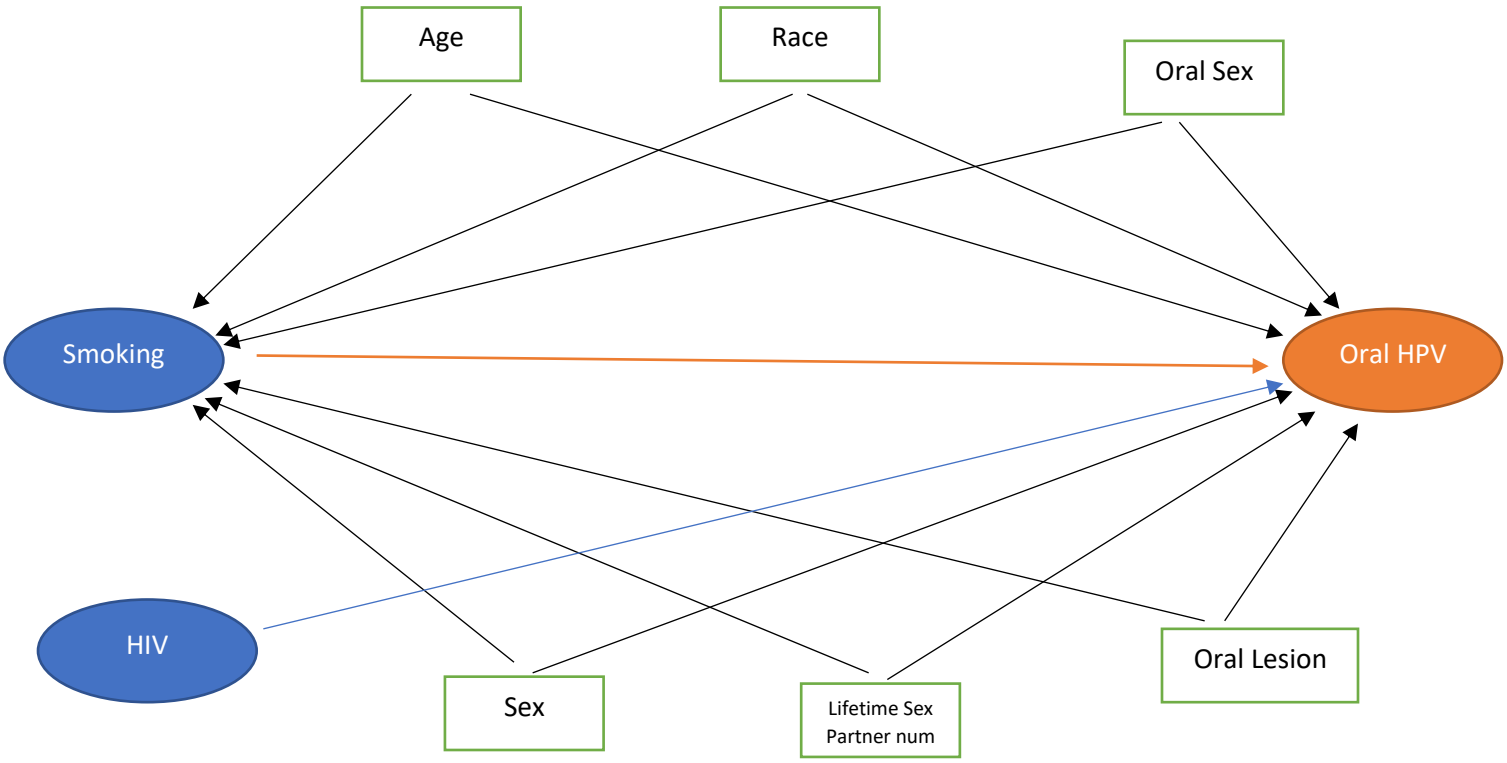

Figure 4: Directed Acyclic Graph of the smoking-HPV relationship with a priori confounders. HIV represents a direct effect modifier (See Vanderweele and Robins (114)).

5.7 Overview of Statistical Analysis

The next section provides details of the statistical strategies employed in Manuscript 1. All analyses were conducted in R Studio (V 1.2.1335) using the R statistical programming language (126).

5.7.1 Modeling of Tobacco Smoking

Smoking was modelled following the recommendation by Leffondre et al (137). First, a centered continuous smoking pack-years variable was created by deducting the mean pack-years from each 'ever-smoker' while keeping the value for 'never-smokers' at zero. Then an indicator variable for 'never' and 'ever' smoker was added as previously defined – this is the binary smoking exposure variable interpretable in the final model. Modelling smoking this way accounted for the quantitative (duration) and qualitative (intensity) differences between a 'never' and 'ever' smoker, while allowing for the interpretation of the coefficient of the indicator variable as the effect of average smoking pack-years among 'ever-smokers' on oral HPV detection compared to 'never-smokers' (137). This transformation and modelling strategy provides a superior model fit (as compared to modelling duration or intensity alone, smoking pack-years alone, or ever and never smoking alone), aids interpretation, and does not alter the regression coefficients (137, 138).

5.7.2 Binary Logistic Regression Model with Interaction Term

A binary logistic regression model was fitted to data to model the association between smoking and the oral HPV outcomes adjusting for confounders as specified in the DAG (Figure 3). A binary logistic regression is a type of generalized linear model that predicts the probability of an outcome

variable given a set of exposures. The employed logistic regression model included an interaction term for smoking and HIV in order to assess for effect modification as follows:

$$\text{Logit}\,(\Pr(\text{HPV}_i \mid S, H, C)) = \beta_0 + \beta_1 \times S_i + \beta_2 \times H_i + \beta_3 \times S_i{*}H_i + \sum_{i=1}^{k} \beta \times C_I \quad [4]$$

Where Pr represents the probability of oral HPV given smoking (S), HIV (H) and the adjusted covariates (C), S_i represents the binary smoking exposure status (ever vs never), H_i represents the binary putative effect modifying exposure variable HIV, and C_i represents the a priori confounding variables alongside the re-centered continuous smoking pack-years variable, respectively for the i^{th} participant. The following important effect measures are estimated from the parameters in model [4]:

- The effect of smoking on oral HPV = $OR_{01} = e^{\beta 1}$
- The effect of HIV on oral HPV = $OR_{10} = e^{\beta 2}$
- The joint effect of smoking and HIV on oral HPV = $OR_{11} = e^{\beta 1 + \beta 2 + \beta 3}$

5.7.3 Effect Modification Analysis

Effect modification by HIV was assessed and reported based on the recommendations by Knol and Vanderweele (115). This mainly includes reporting the following:

1. ORs with CIs for each stratum of smoking and HIV with the stratum of lowest risk as the reference category (OR_{00}, OR_{01}, OR_{10}, and OR_{11})
2. ORs and CIs for the effect of smoking on oral HPV within the strata of HIV
3. Measures of additive effect modification such as RERI and attributable proportion due to Interaction (AP)
4. The confounders for which the association of smoking and oral HPV was adjusted for

5.7.3.1 Relative Excess Risk due to Interaction (RERI)

RERI is the departure from additivity of the effects of smoking and HIV on oral HPV (109, 113, 139), i.e., the extent to which the combined effects of smoking and HIV exceeds the sum of the smoking considered alone and HIV considered alone. This was calculated as follows:

$$RERI = OR_{11} - OR_{01} - OR_{10} + 1 \quad (113)$$

The RERI will be greater than zero, if and only if additive interaction is positive, less than zero, if and only if additive interaction is negative, and zero if there is no additive interaction (140).

5.7.3.2 Attributable Proportion due to Interaction (AP)

The AP is the proportion of the oral HPV outcome in those exposed to both smoking and HIV that is due to the interaction of the two exposures (140), calculated as:

$$AP = RERI \div OR_{11}$$

The CIs for RERI and AP were computed using the delta method (129).

5.7.4 Handling of Missing Data

325 participants were originally recruited into the study. However, oral HPV results were not available for 30 participants for a missing rate of 9.23% (30/325), due to insufficient DNA for additional NGS analysis. While the most common rule of thumb is that 5% missingness is negligible, different analyses have shown that missingness becomes a threat to a study's validity only when the proportion of missing data is greater than 10% and the underlying pattern of missingness is not random (141, 142). The idea being that a complete case analysis after the removal of missing data represents an unbiased sample of the target data (141, 142). Furthermore, it has been shown that if only the dependent variable is missing (as is the case here), and auxiliary variables cannot be identified, then a complete case analysis is also valid; in fact, multiple

imputation in this case will only lead to an increase in uncertainty around the effect estimates (143, 144).

Since the missing rate is less than 10% in our data, and it is reasonable to assume that this missingness is not systematic (due to the reason given above), a listwise deletion of missing data points was conducted, resulting in a complete dataset for 295 participants which was included in the final analysis.

5.7.5 Post-hoc Power Considerations

In this section, we consider a post-hoc power analysis for the final sample size included in Manuscript I. There is scant literature on sample size and power calculations for additive interaction. Here, we use the excel spreadsheet provided as an appendix by Vanderweele in an article that addressed this topic (145). The article provided two versions of the spreadsheet: cohort and case-control; we adapt the case-control version of the spreadsheet and applied this to the high-risk HPV outcome (i.e., high-risk HPV positive as cases and negative as controls). The spreadsheet was populated with the following required inputs:

- Alpha (significance level) – 0.05
- Number of Cases – 46
- Number of Controls – 249
- OR_{10} – 1.56
- OR_{01} – 1.59
- Multiplicative Interaction – 2.21
- Marginal prevalence of HIV – 0.57
- Marginal prevalence of Ever smokers – 0.71

- OR relating HIV and Ever smoker – 1.73

The result of the post-hoc analysis showed that our final sample size of 295 for Manuscript I had a 2-sided power of 30% to detect a $RERI_{OR}$ > or < than 0.

6 MANUSCRIPT I

HIV modifies the effect of tobacco smoking on oral human papillomavirus infection

Babatunde Y. Alli BDS[1], Robert D. Burk MD[2,3], Mahnaz Fatahzadeh DMD MSD[4], Robert M. Grossberg MD[6], Richard V. Smith MD[7], Thomas J. Ow MD MS[7,8], Mauricio Wiltz DMD[5], Jacqueline Polanco DDS[5], Marie-Claude Rousseau PhD[1,9], Belinda Nicolau DDS PhD[1], Nicolas F. Schlecht PhD[1,3,8,10*]

1. Faculty of Dentistry, McGill University, Montreal, Canada
2. Departments of Pediatrics (Genetics), Microbiology & Immunology, Obstetrics Gynecology & Women's Health, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY, USA
3. Department of Epidemiology and Population Health, Albert Einstein College of Medicine and Montefiore Medical Center, New York, United States
4. Department of Diagnostic Sciences, Rutgers School of Dental Medicine, Newark, NJ, USA
5. Department of Dentistry, Albert Einstein College of Medicine and Montefiore Medical Center, New York, United States
6. Department of Infectious Diseases, Albert Einstein College of Medicine and Montefiore Medical Center, New York, USA
7. Department of Otorhinolaryngology-Head and Neck Surgery, Albert Einstein College of Medicine and Montefiore Medical Center, New York, United States
8. Department of Pathology, Albert Einstein College of Medicine and Montefiore Medical Center, New York, USA
9. Epidemiology and Biostatistics Unit, Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique (INRS), Laval, Quebec, Canada
10. Department of Cancer Prevention and Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA

***Corresponding author:**
Department of Cancer Prevention and Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA

This manuscript was published on 26 March 2020 by the Journal of Infectious Diseases as:

Babatunde Y Alli, BDS, Robert D Burk, MD, Mahnaz Fatahzadeh, DMD MSD, Julie Kazimiroff, DDS, Robert M Grossberg, MD, Richard V Smith, MD, Thomas J Ow, MD, MS, Mauricio Wiltz, DMD, Jacqueline Polanco, DDS, Marie-Claude Rousseau, PhD, Belinda Nicolau, DDS, PhD, Nicolas F Schlecht, PhD, HIV modifies the effect of tobacco smoking on oral human papillomavirus infection, The Journal of Infectious Diseases, jiaa135. Available at https://doi.org/10.1093/infdis/jiaa135

Abstract

Introduction: People living with HIV (PLWH) are more likely to smoke and harbor oral human papillomavirus (HPV) infections, putting them at higher risk for head and neck cancer. We sought to investigate the joint effects of HIV and smoking on risk of oral HPV.

Methods: Consecutive PLWH (n=169) and at-risk HIV-negative individuals (n=126) were recruited from two large health centers in United States. Lifetime smoking history was collected using questionnaires. Participants provided oral rinse samples for HPV genotyping by Next-Generation Sequencing. We used multivariable logistic regression models with interaction terms for HIV to test for smoking effect on oral HPV and reported the Relative Excess Risk due to Interaction (RERI).

Results: PLWH were more likely to harbor oral HPV than HIV-negative individuals, including α (39% vs. 28%), β (73% vs. 63%), and γ-types (33% vs. 20%). HIV infection positively modified the association between smoking and detection of high-risk oral HPV with odds ratios for smoking of 3.46 [95%CI:1.01, 11.94] and 1.59 [95%CI:0.32, 8.73] among PLWH and HIV-negative individuals, respectively, and an observed RERI of 3.34 [95%CI:-1.51, 8.18]. The RERI for HPV-16 was 1.79 [95%CI:-2.57, 6.16], and 2.78 for β1-HPV [95%CI:-0.08, 5.65].

Conclusion: Our results show tobacco smoking as a risk factor for oral HPV among PLWH.

Key words: HIV, Tobacco, HPV, Smoking

Introduction

Human Papillomavirus (HPV) infection is an established risk factor for a distinct subset of Head and Neck Cancers (HNCs), accounting for an increasing share of incident HNCs in North America, including in particular cancers originating in the oropharyngeal palatine and lingual tonsils (1, 4). Three species groups of HPV are well-known to infect humans: alpha (α), beta (β) and gamma (γ); with the α-HPV types being most extensively studied due to their established associations with cervical and anogenital diseases (14). Although the α-HPV types – overwhelmingly HPV-16 – are responsible for over 90% of HPV-positive HNCs, recent evidence suggests an etiologic role for types that were until recently not known to infect the oral cavity (6), including β (HPV-5 and -8) and γ (γ-11 and γ-12) types (17, 19).

People living with HIV (PLWH) have a higher prevalence of all types of oral HPV (5, 6), and the molecular and mechanistic interactions that underpin this relationship have been previously described (146). As with most cancers, PLWH are also at higher risk of HPV-related HNCs (7). Tobacco smoking, a strong and independent risk factor for HNCs prevalent among PLWH (8), is also an important risk factor for oral HPV infection (9, 10, 147). However, this association has only been demonstrated for the detection of α-HPV types (especially high-risk types) known to infect the mucosal epithelia of the anogenital tract. Furthermore, whereas differences in prevalence and risk of oral HPV in PLWH compared to HIV-negative populations have been demonstrated, the interactive effects of HIV and smoking have not been fully explored (5).

Many studies have suggested that smoking has adverse effects on immunity, both systemic and mucosal (95, 148). We hypothesized that the synergistic effects of HIV infection and smoking combined will translate into higher risks of oral HPV infection, and that the risk of all types of oral HPV will be greater among PLWH compared to HIV-negative individuals.

Materials and Methods

Study Population

The Multicenter Oro-pharyngeal Squamous Atypical Lesion (MuCOSAL) study was a prospective study conducted between 2004 and 2013 inclusive of HIV-positive and HIV-negative patients presenting with and without oral lesions to the Infectious Disease, Dental, Oral Medicine, and Otolaryngology outpatient clinics at the Montefiore Medical Center (Bronx, NY) and Rutgers School of Dental Medicine (Newark, NJ). Where possible, HIV-negative participants were frequency matched to HIV-positive patients on age (±10 years) and sex. Exclusion criteria were minors aged less than 21 years and cancer treatment in the preceding year. Participants were followed-up for approximately 1 year, with three visits at baseline, 6 months and 12 months. The study protocol was approved by the Institutional Review Boards at the participating institutions, and all participants provided written informed consent prior to participating.

Data Collection

The main components of the data collection process for this study have been described previously (131). Briefly, participants underwent a structured interview administered by trained research personnel. We collected information on socio-demographic characteristics (such as age, sex, race/ethnicity and occupation); lifestyle habits (e.g. cigarette smoking and alcohol consumption); sexual practices and history (e.g. lifetime number of sexual partners, practice of oral sex); and history of various sexual-transmitted diseases.

In addition, each participant underwent a comprehensive clinical oral examination. Patients were classified based on presentation of the most severe oral lesion, if detected, including

benign/papilloma (e.g. oral warts, frictional keratosis), and potentially malignant lesions (e.g. lichen planus, leukoplakia and/or erythroplasia) (149).

Participants provided a 30-second gargle and rinse sample at each visit using Scope® mouthwash for HPV genotyping. The samples were kept on ice until stored at -20°C within six hours of collection and thereafter transferred to the lab for DNA isolation. In the current analyses, baseline questionnaire, clinical oral examination, and HPV genotyping data were used to test for the effect of smoking on HPV detection by HIV serostatus.

Smoking history and covariates

Detailed smoking history was collected by asking participants to divide their lifetime consumption into varying periods of duration and intensity. For each period, information on age at start, age at cessation, and the number of cigarettes smoked per day in that period was collected. From this, smoking duration was calculated, and pack-years computed as the number of cigarettes smoked daily divided by 20 and multiplied by the duration. Ever-smokers were defined as those who smoked tobacco for at least one year at any point in their lives.

Age, sex, race/ethnicity, lifetime number of sexual partners and oral sexual history are factors that have been shown to be independently associated with oral HPV detection in the literature (4, 9, 78, 131). Also, HPV is associated with the presence of both benign and potentially malignant lesions (150). Therefore, we considered these variables as possible confounders in our study. Also, we considered HIV infection a direct effect modifier of the smoking-HPV relationship. We represented these factors and their relationships with the exposure (smoking) and outcome (HPV detection) in our proposed causal directed acyclic graph for the overall smoking-HPV relationship

(Figure 3), while estimating the direct effect modification by HIV infection using the approach proposed by Vanderweele and Robbins (114).

HPV DNA Testing

HPV DNA detection and typing was performed using a next-generation sequencing (NGS) assay developed in our research group (130). DNAs extracted from the oral rinse sample and brush biopsy specimens were tested for the presence of *beta*- and *gamma*-HPV types using two different PCR assays, termed "FAP and E1". These assays are similar to the "SPF" assay, which was designed to amplify *alpha*-HPVs recently described in Fonseca et al. (130). These assays operate similarly to the 16S microbiome assays that amplify a specific fragment using bar-coded primers. For *beta*- and *gamma*-HPV detection, a unique 12 bp Golay DNA barcode was assigned to each subject and introduced into the PCR amplicons. Barcoded PCR products from multiple subjects were pooled at approximately equal molar DNA concentrations and sequenced on an Illumina platform (Illumina, San Diego CA). The millions of reads produced by NGS were de-multiplexed with individual samples being reconstructed by sorting barcodes into unique files. The reads were then matched with known or novel HPV sequences using BLAST, and a specialized HPV database developed by our lab. Positive and negative controls were included with each PCR, as well as samples previously identified to have a representation of different *beta*- and *gamma*-HPV types. Identification of individual HPV types was based on parameters including total read number, percent of total HPV reads and distribution of reads for each HPV type.

For the study outcomes, we classified HPV phylogenetically into α, β and γ HPV genera, and the relevant species corresponding to these genera: α-3, α-9; β-1 , β-2, β-3; γ-7, γ-8, γ-9 (14). Furthermore, we grouped together α-HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59 as high-risk α-types following the IARC classification (132), and we singled out α-HPV-16 – for

its predominant role in HNCs – as well as the recently implicated β-1 species types, HPV-5 and HPV-8 (17, 19).

Statistical Analyses

Unconditional logistic regression modelling was used to estimate the odds ratio (OR) and confidence interval (CI) for the smoking-HPV relationship while adjusting for previously specified confounders. Smoking was modelled following the recommendation by Leffondre et al (137). First, we included a centered continuous smoking pack-years variable by deducting the mean pack-years from each 'ever-smoker' while keeping the value for 'never-smokers' as zero. Then we added an indicator variable for 'never' and 'ever' smoker. Modelling smoking this way accounted for the quantitative (duration) and qualitative (intensity) differences between a 'never' and 'ever' smoker, while allowing for the interpretation of the coefficient of the indicator variable as the effect of average smoking pack-years among 'ever-smokers' on oral HPV detection compared to 'never-smokers'. This transformation and modelling strategy provides a superior model fit, aids interpretation, and does not alter the regression coefficients (137, 138).

We assessed for effect modification based on an epidemiological definition: that the effect of one variable (smoking – the exposure) on another (oral HPV – the outcome) varies across the strata of a third (HIV – the effect modifier) (114), by including an interaction term for HIV in the logistic model, and reporting this following the four steps recommended by Knol and Vanderweele (115). This included estimating ORs for each stratum of HIV and smoking and their joint effects, alongside the effect of smoking on HPV within the strata of HIV and the relative excess risk due to interaction (RERI).

RERI represents the difference between the joint OR and the individual contributions by smoking and HIV, calculated as: $OR_{11} - OR_{01} - OR_{10} + 1;$ whereas a RERI of 0 indicates perfect additivity (i.e. no effect modification), a value of greater or less than 0 indicates positive or negative additive effect modification respectively (151). In addition, we estimated the attributable proportion (AP) of oral HPV risk associated with both smoking and HIV, which was calculated as: $RERI \div OR_{11}$ (151). For the few outcomes where either of the exposure was protective (i.e. either OR_{01} or OR_{10} was < 1), the exposure was recoded as a risk factor in order to estimate the correct RERI and AP (152). Confidence intervals and p-values were computed from the standard error estimates using the delta method (129).

As sensitivity analyses, we modelled smoking simply as never, past and current, and retested the effect modification by HIV for high-risk for α-HPV, HPV-16 and β-1. All analyses were conducted in R Studio (V 1.2.1335) using the R statistical programming language (126).

Results

A total of 325 participants were recruited into the study. HPV genotyping results were missing for 30 subjects, leaving 295 adults as the final sample size in this analysis – 126 HIV-negative and 169 HIV-positive individuals. The mean age of the study participants was 52.1 years (±11.8 standard deviation), with approximately half being male (51.2%) and half identifying as African American (49.1%). Table 1 further describes the characteristics of the study participants stratified by HIV serostatus. HIV-positive participants were younger on average, had a higher number of lifetime sexual partners, smoked more cigarettes, and drank more alcohol across their lifetimes, but less likely to have oral lesions. Among HIV-negative individuals, prevalence of any α-HPV or β-1 HPV was associated with the presence of benign/papilloma or precursor/dysplastic lesions

respectively (Appendix II – Table S1); there were no significant differences in oral HPV prevalence by oral lesion status among PLWH (Appendix III – Table S2).

Table 1: Selected characteristics of study participants by HIV status

	HIV-Negative n= 126	HIV-Positive n=169	
Age (Mean (SD))	52.25 (14.10)	49.83 (8.76)	0.071
Sex (%)			
Female	65 (51.6)	79 (46.7)	0.481
Male	61 (48.4)	90 (53.3)	
Race (%)			
Non-Hispanic White	24 (19.0)	15 (8.9)	0.071
Hispanic	39 (31.0)	60 (35.5)	
African-American	57 (45.2)	88 (52.1)	
Other	6 (4.8)	6 (3.6)	
Ever-smoker (%)			
Never	44 (34.9)	40 (23.7)	0.047
Ever	82 (65.1)	129 (76.3)	
Smoking Pack-Years (among ever-smokers) (Mean (SD))	20.50 (22.84)	20.17 (20.47)	0.914
Smoking History (%)			
Never	44 (34.9)	40 (23.7)	0.089
Current	50 (39.7)	84 (49.7)	
Past	32 (25.4)	45 (26.6)	
Lifetime Ethanol Use in Liters (Median [IQR])	17 [0.00, 158.50]	30 [0.00, 336.00]	0.196
Lifetime Ethanol Use Categories (%)			
Up to 10L	61 (48.4)	69 (40.8)	0.422
>10 – 124L	28 (22.2)	39 (23.1)	

>124 – 369L	15 (11.9)	19 (11.2)	
>369L	22 (17.5)	42 (24.9)	
Type of Oral Lesion			
None	79 (62.7)	132 (78.1)	0.011
Benign/Papilloma	19 (15.1)	18 (10.7)	
Precursor/Dysplastic	28 (22.2)	19 (11.2)	
Lifetime Number of Sexual Partners			
≤5	61 (48.4)	50 (29.6)	<0.001
6 - 20	45 (35.7)	54 (32.0)	
≥ 21	20 (15.9)	61 (36.1)	
NA	0(0.0)	4(2.4)	
Sexual Preferences			
WSM	61 (48.4)	67 (39.6)	NA
WSMW	3 (2.4)	11 (6.5)	
MSM	0 (0.0)	22 (13.0)	
MSWM	57 (45.2)	64 (37.9)	
NA	5 (4.0)	5 (3.0)	
Oral Sex History			
No	36 (28.6)	27 (16.0)	0.024
Yes	90 (71.4)	141 (83.4)	
NA	0 (0.0)	1 (0.6)	
CD4 Count			
500 + T-Cells / mm3	-	50 (29.96)	-
>200 - < 500 T-Cells / mm3	-	54 (32.0)	
≤200 T-Cells / mm3	-	17 (10.1)	
NA		48 (28.4)	
HIV Viral Load			-
<4000 copies / ml	-	130 (76.9)	
4000 + copies / ml	-	13 (7.7)	
NA	-	26 (15.4)	

Abbreviations: MSM, men who have sex with men only; MSWM, men who have sex with women only or with women and men; WSM, women who have sex with men only; WSWM, women who have sex with women only or with women or men; NA, not available.

Figure 5 shows the distribution of oral HPV types and genera detected by HIV status. Consistently, the prevalence of HPV was higher among HIV-positive participants for all genera and types, except α-14 types. In addition, no α-11, γ-3, or high-risk α-HPV 31 or 45 types, were detected in oral rinse samples of HIV-negative participants.

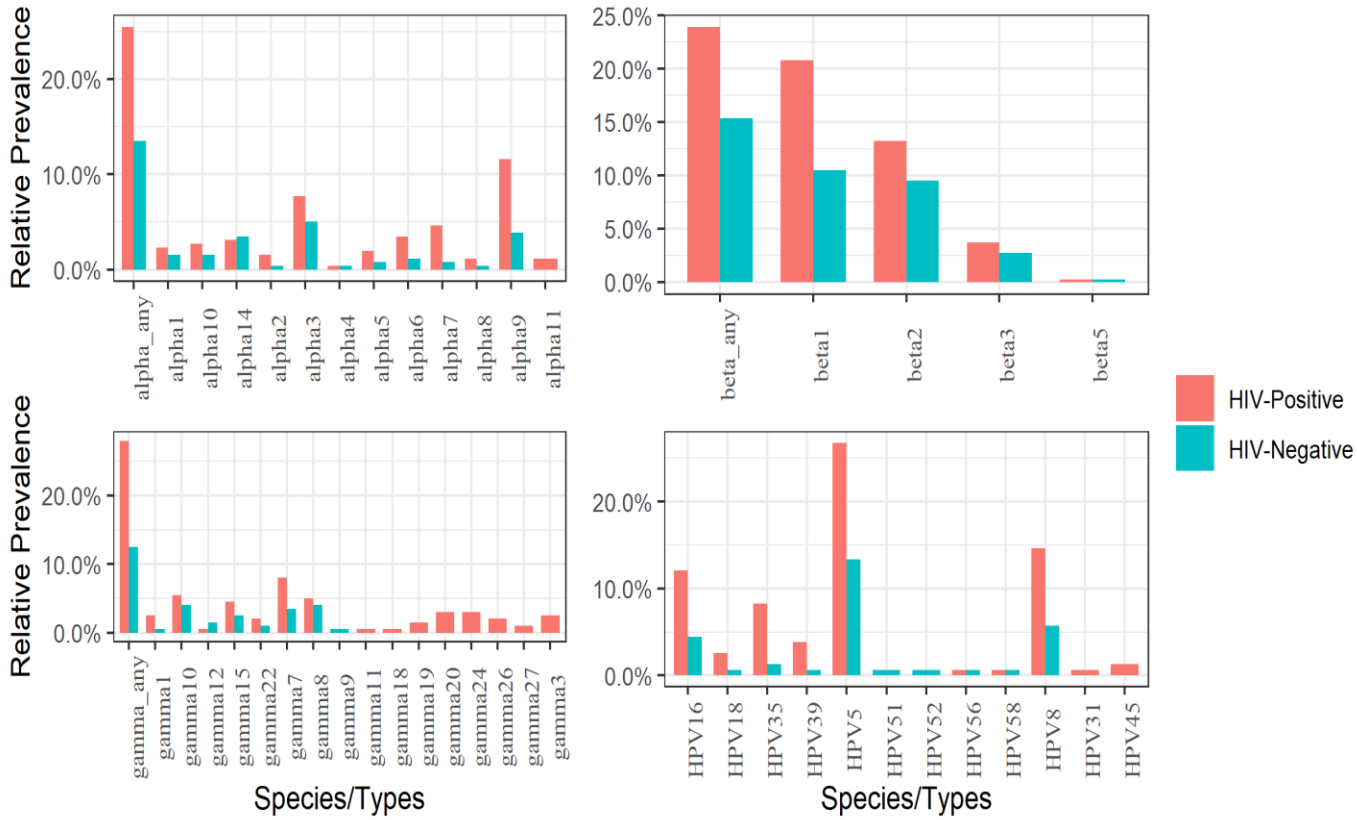

Figure 5: Unadjusted relative prevalence of oral HPV (genus, species and types) by HIV status

Among HIV-positive participants, an ever-smoker with 15.4 pack-years (reflecting an average number of pack-years in our study population) had 3.46 times higher odds of presenting with a high-risk α-HPV type [95% CI: 1.01, 11.94] compared to a never-smoker HIV-positive after

adjusting for confounders. Increased associations were also observed for HPV-16 alone (Table 2) and β-1 HPV species (Table 3).

For α-genus HPV, the joint effect estimates (i.e. ever smoker + HIV positive, OR_{11}) for all the outcomes were consistently larger than the individual effects (i.e. ever smoker alone, OR_{01}; and HIV positive alone OR_{10}) (Table 2). On the additive scale, HIV positively modified the effect of smoking on the detection of oral α-HPV types, including HPV-16 – with the strongest measures of effect modification observed for high-risk α-types (RERI = 3.34, 95CI: 1.51, 8.18]; and AP = 0.61, 95%CI: 0.12, 1.10) (Table 2). The results indicated a negative effect modification for the detection of α-3 and α-9 species types (except HPV-16).

Table 2: Modification of the effect of smoking on oral α-HPV (genus/specie/types) by HIV

	Never smoker OR [95% CI]	Ever smoker OR [95% CI]	OR [95% CI] for the effect of smoking on oral HPV within the strata of HIV
Any α-HPV			
HIV negative	1	1.67 [0.68, 4.32]	1.67 [0.68, 4.32]
HIV positive	1.32 [0.47, 3.76]	3.05 [1.29, 7.73]	2.32 [0.96, 5.56]
	RERI [95% CI] = 1.06 [-0.81, 2.95] p=0.13 AP [95% CI] = 0.35 [-0.21, 0.91] p=0.11		
α-3			
HIV negative	1	1.49 [0.42, 6.23]	1.49 [0.42, 6.23]
HIV positive	1.26 [0.27, 5.93]	1.61 [0.48, 6.58]	1.28 [0.36, 4.59]
	RERI [95% CI] = -0.14 [-2.50, 2.22] p=0.54 AP [95% CI] = -0.09 [-1.53, 1.36] p=0.45		
α-9			
HIV negative	1	0.96 [0.24, 4.21]	0.96 [0.24, 4.21]
HIV positive	2.03 [0.51, 8.88]	3.32 [1.01, 13.54]	1.64 [0.53, 5.03]

	*****RERI [95% CI] = -1.40 [-5.71, 2.91] p=0.73 *****AP [95% CI] = -0.66 [-2.58, 1.26] p=0.24		
α-high-risk HPV[T]			
HIV negative	1	1.56 [0.39, 7.86]	1.56 [0.39, 7.86]
HIV positive	1.59 [0.32, 8.73]	5.48 [1.58, 26.23]	3.46 [1.01, 11.94]
	RERI [95% CI] = 3.34 [-1.51, 8.18] p=0.08 AP [95% CI] = 0.61 [0.12, 1.10] p=0.007		
α-HPV 16			
HIV negative	1	1.60 [0.30, 12.24]	1.60 [0.30, 12.24]
HIV positive	1.75 [0.26, 14.41]	4.14 [0.89, 30.93]	2.37 [0.53, 10.54]
	RERI [95% CI] = 1.79 [-2.57, 6.16] p=0.20 AP [95% CI] = 0.43 [-0.40, 1.26] p=0.15		

ORs adjusted for age, sex, race, smoking pack-years, oral sex history and lifetime number of sexual partners.

[T] *HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59*

**At least one exposure was preventive, so it was recoded as a risk factor to correctly estimate RERI and AP*

Table 3: Modification of the effect of smoking on oral β-HPV (genus/specie/types) by HIV

	Never smoker OR [95% CI]	Ever smoker OR [95% CI]	OR [95% CI] for the effect of smoking on oral HPV within the strata of HIV
Any β-HPV			
HIV negative	1	0.87 [0.37, 2.02]	0.87 [0.37, 2.02]
HIV positive	1.00 [0.39, 2.55]	2.50 [1.07, 5.83]	2.49 [1.04, 5.98]
	*****RERI [95% CI] = -1.85 [-4.64, 0.93] p=0.90 *****AP [95% CI] = -1.62 [-3.92, 0.68] p=0.08		
β-1			
HIV negative	1	1.12 [0.49, 2.60]	1.12 [0.49, 2.60]
HIV positive	1.70 [0.67, 4.39]	4.60 [2.03, 10.86]	2.70 [1.16, 6.28]
	RERI [95% CI] = 2.78 [-0.08, 5.65] p=0.02 AP [95% CI] = 0.60 [0.24, 0.96] p=0.0005		
β-2			
HIV negative	1	0.94 [0.42, 2.15]	0.94 [0.42, 2.15]
HIV positive	0.80 [0.31, 2.02]	1.27 [0.58, 2.87]	1.60 [0.69, 3.69]
	*****RERI [95% CI] = -0.67 [-2.31, 0.97] p=0.78 *****AP [95% CI] = -0.57 [-1.87, 0.73] p=0.19		
β-3			
HIV negative	1	0.31 [0.09, 1.08]	0.31 [0.09, 1.08]
HIV positive	0.54 [0.13, 2.05]	0.42 [0.13, 1.39]	0.78 [0.21, 2.91]
	*****RERI [95% CI] = -1.83 [-5.91, 2.25] p=0.81 *****AP [95% CI] = -1.05 [-3.34, 1.23] p=0.18		
HPV 5			
HIV negative	1	0.75 [0.26, 2.32]	0.75 [0.26, 2.32]
HIV positive	1.97 [0.65, 6.30]	1.05 [0.39, 3.09]	0.53 [0.20, 1.39]
	*****RERI [95% CI] = 0.89 [-1.04, 2.83] p=0.18 *****AP [95% CI] = 0.34 [-0.35, 1.03] p=0.16		
HPV 8			
HIV negative	1	1.14 [0.26, 6.03]	1.14 [0.26, 6.03]

HIV positive	1.78 [0.36, 9.95]	2.93 [0.79, 14.79]	1.64 [0.46, 5.92]
	RERI [95% CI] = 1.00 [-1.93, 3.94] p=0.25 AP [95% CI] = 0.34 [-0.59, 1.28] p=0.23		

ORs adjusted for age, sex, race, smoking pack-years, oral sex history and lifetime number of sexual partners.

**At least one exposure was preventive, so it was recoded as a risk factor to correctly estimate RERI and AP*

HIV also showed super-additive (positive) effect modification on the associations between smoking and oral β-HPV types HPV-8 and -5, and β-1 species types, but negative effect modification for β-2 and β-3 types (Table 3), as well as with γ-8 and γ-10 species types (Supplemental Table 1). Overall, trend estimates indicated stronger smoking-HPV associations among PLWH compared to HIV negative individuals (right-most columns of Tables 2, 3 and 4), consistent with our hypothesis of effect modification by HIV.

In the sensitivity analyses, where smoking history was modelled simply as never, past and current (Table 4), the joint effect estimates for current smoker PLWH were higher than for HIV-positive past smokers, for high-risk α-HPV types (including HPV-16) and β-1 species types (Table 4). These dose-response associations were also reflected by the measures of additive effect modification (i.e. RERI and AP for each outcome).

Table 4: Sensitivity analyses for the modification of the effect of smoking on oral HPV (α-high risk, HPV 16, and β-1 specie) by HIV

		Smoking			OR [95%CI] for past smoker on oral HPV within strata of HIV	OR [95%CI] for current smoker on oral HPV within strata of HIV
		Never	Past	Current		
		OR [95%CI]	OR [95%CI]	OR [95%CI]		
α-high-risk HPV[Ŧ]						
	HIV Negative	1.0	1.52 [0.24, 9.69]	1.53 [0.32, 8.53]	1.52 [0.24, 9.69]	1.53 [0.32, 8.53]
	HIV Positive	1.58 [0.32, 8.72]	4.33 [1.05, 22.84]	6.24 [1.72, 30.76]	3.33 [0.85, 13.10]	3.55 [0.98, 12.80]
		RERI [95%CI]:	2.23 [-2.53, 6.99] p=0.17	4.12 [-2.03, 10.28] p=0.09		
		AP [95% CI]:	0.51 [-0.25, 1.28] p=0.09	0.66 [0.19, 1.13] p=0.002		
α-HPV 16						
	HIV Negative	1.0	1.45 [0.15, 14.19]	1.69 [0.24, 14.70]	1.45 [0.15, 14.19]	1.69 [0.24, 14.70]
	HIV Positive	1.76 [0.26, 14.53]	3.29 [0.55, 27.48]	4.85 [0.96, 38.10]	2.12 [0.40, 11.30]	2.54 [0.53, 12.14]
		RERI [95% CI]:	1.08 [-3.73, 5.90] p=0.33	2.41 [-3.26, 8.07] p=0.20		
		AP [95% CI]:	0.33 [-0.94, 1.60] p=0.31	0.50 [-0.30, 1.29] p=0.11		
β-1[†]						
	HIV Negative	1.0	1.21 [0.43, 3.47]	1.06 [0.42, 2.69]	1.21 [0.43, 3.47]	1.06 [0.42, 2.69]
	HIV Positive	1.70 [0.67, 4.38]	3.85 [1.47, 10.55]	5.03 [2.11, 12.53]	2.29 [0.86, 6.15]	2.93 [1.20, 7.14]
		RERI [95%CI]:	1.95 [-1.44, 5.34] p=0.13	3.27 [-0.24, 6.78] p=0.03		
		AP [95% CI]:	0.51 [-0.11, 1.12] p=0.05	0.65 [0.30, 1.00] p=0.0001		

ORs adjusted for age, sex, race, oral sex history and lifetime number of sexual partners.

[Ŧ] HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59

[†]Includes HPV 5 and

Discussion

In this study, we examined how the association between smoking and oral HPV is modified by HIV. We show positive effect modification by HIV on the smoking for high-risk α-HPV types, including HPV-16 alone, and for β-1 species types, with a high attributable proportion for HPV detection in the oral cavity associated with exposure to both smoking and HIV. In addition, we found a higher (unadjusted) prevalence for almost all α-HPV types in PLWH compared to HIV-negative individuals that was consistent with the literature (5, 11, 131, 153). In addition, we found higher prevalence of β- and γ-HPV types among PLWH, which was novel. We also report a higher prevalence of oral lesions among the HIV-negative individuals, despite their lower oral HPV prevalence. This is likely due to recruiting HIV-negative individuals from dental and oral medicine clinics at the early stage of the study, which may have oversampled patients with oral lesions; importantly however, all our reported effect estimates were adjusted for each individual's lesion status (none, benign/papilloma or precursor/dysplastic).

The immunosuppressive effects of cigarette smoke have been well described (154, 155), including adverse effects on both systemic and mucosal immunity (95, 148). The combination of tobacco-induced and HIV-induced immunosuppression may explain the positive RERI and high AP for oral HPV detection observed in our study. This suggests that PLWH who smoke may be at higher risk of HPV-associated head and neck cancer (113).

Deployment of oral HPV testing/screening for HNC prevention will be challenging, due to the low prevalence of high-risk oral HPV in the general population (122). Consequently, for oral HPV screening to be viable and useful, there needs to be identification of high-risk subgroups (122, 123); and there is a growing public health interest in identifying these populations (156, 157). Our study findings suggest that PLWH who smoke may be such a candidate group.

To the best of our knowledge, our study is the first to show effect modification by HIV of the smoking-oral HPV relationship. A previous study by Beachler et al. (11) found no relationship between smoking intensity and duration, and detection of α-type oral HPV in both HIV-positive and negative individuals. This contrasts with our finding of an association between smoking and α-high-risk oral HPV among PLWH but HIV-negative individuals, or the observed risk disparity with β-1 oral HPV. The differing results could be due to how the two studies modelled smoking history, and/or differences in HPV DNA testing methods; in their study, Beachler et al. tested for α-types by PGMY09/11 PCR, whereas we employed NGS. However, their smoking modelling strategy was not described.

There are a few limitations to consider in the interpretation of our results. First, since our analyses were based on the baseline data (i.e. were cross-sectional), the temporal relationship between smoking, HPV and HIV infection is unclear. This notwithstanding, our conclusions are plausible, given that the effects of smoking on oral HPV has been shown prospectively and with increasing smoking pack-years (10), and smoking has been shown to be associated with a reduced clearance rate of oral HPV (5). Second, the NGS method we used to test for HPV is a highly sensitive assay with elevated risk for false positives; this, combined with recall error inherent in measuring lifetime cigarette smoking and other self-reported covariates, increases the potential for misclassification bias in our results.

In conclusion, while previous studies have shown an association between smoking and detection of alpha-HPV in the oral cavity, this was the first to show associations for non-alpha HPV types, and to provide evidence of effect modification by HIV on the relationship between smoking and oral HPV, with a significant attributable proportion of high-risk HPV associated with combined exposure to smoking and HIV. Although we have previously reported a role in HNC risk for

detection of several β- and γ-HPV types in the oral cavity (19), further evidence is needed from studies of HIV-infected populations, who are at higher risk of HNC (7). Also, importantly, we showed that PLWH, with an average of 15.4 smoking pack-years, are at higher risk of presenting with oral HPV compared to HIV-positive never-smokers. If our findings are confirmed, this may have public health implications if high-risk PLWH can be targeted for preferential HNC screening and smoking cessation interventions.

Preface to manuscript II

As described previously, the reporting of effect modification analysis is inadequate in the literature. A plausible reason for this is the extra work and sample size required to obtain estimates. This was my experience during the preparation and analysis stage of manuscript I. Clearly, there is a need to make this task easier for authors – most of whom effect modification analysis is often a secondary aim rather than a primary one – thereby encouraging and increasing full reporting of effect modification analysis. In manuscript II, I presented a new R package that meets this need.

7 MANUSCRIPT II

Title: InteractionR: An R Package for Full Reporting of Effect Modification and Interaction Analyses

Babatunde Y. Alli BDS[1*]

1Faculty of Dentistry, McGill University, Montreal, Canada.

*Corresponding author at

Faculty of Dentistry, McGill University

2001 McGill College Avenue, Suite 500

Montreal, QC, Canada

H3A 1G1

babatunde.alli@mail.mcgill.ca

Abstract word count: 149

Text word count: 1971

Manuscript to be submitted to *Computer Methods and Programs in Biomedicine*

Abstract

Effect modification and/or Interaction are frequently assessed in epidemiological research. However, in most cases, authors do not present sufficient information for the readers to fully assess the extent and significance of interaction on both additive and multiplicative scale. Also, due to being readily available in most software, the delta method has proliferated in the literature for the estimation of confidence intervals (CIs) for measures of additive interactions; despite its well documented poor performance compared to alternative methods.

We introduce *interactionR*, an open-source R package with user-friendly functions that ensures full reporting of effect modification or interaction based on recommended guidelines. In addition to the simple asymptotic delta method, the package also allows for estimation of CIs for additive interaction measures using the variance recovery and percentile bootstrapping methods. The package is available from the Comprehensive R Archives Network (CRAN) [https://cran.r-project.org/package=interactionR] and can be installed from within R using *install.packages("interactionR")*.

Key words: Effect modification, Interaction, R, variance recovery, delta method, percentile bootstrapping.

Introduction

Although effect modification analysis is ubiquitous in the epidemiologic literature, its reporting is grossly inadequate (12). For example, whereas departure from additivity is more relevant to disease prevention and public health intervention (13), most studies only report interaction on the multiplicative scale because this can be easily extracted from the exponentiated coefficient of the product term of two exposures in a logistic regression model (12).

To improve this situation, Knol and Vanderweele proposed a set of recommendations for the reporting of analyses of effect modification (115) – an improvement on earlier proposals such as STROBE (Strengthening The Reporting of Observational Study) (158). In their recommendation, they distinguished effect modification from interaction. Effect modification is when the effect of an exposure on an outcome differs within the strata of a second exposure, while interaction is the causal effect of two exposures together on an outcome (111). For effect modification analysis with three dichotomous variables: the effect of exposure A on an outcome Y and a putative modifying variable M of this effect, they recommended reporting mainly the following (Table 5a):

i) Individual (A on Y (OR_{01}) and M on Y (OR_{10})) and joint effect estimates (A and M on Y (OR_{11})) with confidence intervals (CI), with the stratum of lowest risk (OR_{00}) as the reference

ii) Estimates and CIs of the effect of A on Y within the strata of M and

iii) Measures of additive (e.g., Relative Excess Risk due to Interaction, RERI) and multiplicative effect modification with CIs.

If interaction of A and M on Y is of interest, then to report in addition to the above, estimates and CI of the effect of M on Y within the strata of A (Table 5b). Presenting the results this way allows the reader to fully interpret all dimensions of interaction (115).

Table 5: Recommended presentation formats for effect modification and interaction analyses (Adapted with permission from Tables 1 and 3 in Knol and Vanderweele (115)).

a*.

	A = 0	A = 1	A = 1 vs A = 0 within the strata of M
M = 0	OR_{00} (95% CI) [Reference]	OR_{01} (95% CI)	$OR_{A=1 \text{ vs } A=0}$ [M = 0] (95% CI)
M = 1	OR_{10} (95% CI)	OR_{11} (95% CI)	$OR_{A=1 \text{ vs } A=0}$ [M = 1] (95% CI)
Effect modification measure on additive scale (95% CI); Effect modification on multiplicative scale (95% CI); ORs adjusted for …			

b*.

	A = 0	A = 1	A = 1 vs A = 0 within the strata of M
M = 0	OR_{00} (95% CI) [Reference]	OR_{01} (95% CI)	$OR_{A=1 \text{ vs } A=0}$ [M = 0] (95% CI)
M = 1	OR_{10} (95% CI)	OR_{11} (95% CI)	$OR_{A=1 \text{ vs } A=0}$ [M = 1] (95% CI)
M = 1 vs M = 0 within the strata of A	$OR_{M=1 \text{ vs } M=0}$ [A = 0] (95% CI)	$OR_{M=1 \text{ vs } M=0}$ [A = 1] (95% CI)	
Interaction measure on additive scale (95% CI); Interaction on multiplicative scale (95% CI); ORs adjusted for …			

*Though OR (odds ratio) is used here, this could also denote risk ratios (RR) or risk difference (RD)

Fitting a regression model in R (126) will produce some direct estimates required to fill these tables, however, estimates such as RERI, $OR_{A=1 \text{ vs } A=0}$ [M = 1], $OR_{M=1 \text{ vs } M=0}$ [A = 1] and their CIs are not readily available from the regression model and need to be produced with additional methods and coding.

Complicating things further are the competing methods in the literature for the estimation of CIs around these measures of additive interaction (i.e. RERI, Attributable Proportion (AP) and Synergy Index (SI)). The first such method was the simple asymptotic delta method proposed by Hosmer and Lemeshow (159), which is the most widely used in the literature (128, 160-162) despite its well-documented poor performance (163, 164). To overcome the shortcomings of this method, Assmann et al. (163) suggested a non-parametric bootstrapping approach for CI of RERI, where RERI is estimated each time in a specified number (usually 1000) of bootstrap replications (with replacement) of the original sample. Then the 95% CI of RERI is the 2.5th and 97.5th percentile of the resulting distribution. However, this method, although has a superior performance if the sample size is large enough (163, 164), falls apart at typical sample sizes at which most observational studies are performed at due to the inevitable sparse cells in some of the bootstrap simulations (165). However, the variance recovery ('MOVER') method that was subsequently proposed by Zou (164) demonstrated better performance to the delta method at typical sample sizes and a comparable performance to bootstrapping at larger sample sizes.

Although, there are R functions (128, 160) to estimate relevant measures of additive interactions, they compute CIs using the easily implemented delta method, while the alternative methods are buried in the appendices of their publications as spreadsheets, Stata and SAS codes (163, 164). The easy availability of the delta method for estimation of CIs of additive interaction measures across the major statistical software may explain its prominence in the literature. In this paper, we introduce 'interactionR', an R package that ensures full reporting of interaction (Table 5a or b) in easy steps. It also allows the option of computing the CIs for measures of additive interaction with all of the methods described above.

Implementation and Usage

The interactionR is implemented in the R statistical software environment (126); R is installable from https://www.r-project.org/. The main user-facing function of the package, *interactionR()*, accepts the following arguments:

- *model*: a regression model fitted by the user with interaction term for the two exposures under consideration. This may be an object of class glm with a valid link for logistic regression or approximants of risk ratio, class clogit or class coxph. It can also include confounders adjustment as is usually the case.
- *exposure_names*: A character vector of two named binary exposure variables present in the fitted model.
- *ci.type*: A character string ("delta" or "mover") specifying the method to use for the estimation of CI for the measures of additive interaction. Default is "delta".
- *ci.level*: Magnitude of the returned CI level.
- *em*: TRUE, for effect modification assessment. FALSE, for interaction.
- *recode*: If TRUE, recodes the exposures - if at least one of the exposures is protective - such that the stratum with the lowest risk becomes the new reference category when the two exposures are considered jointly (See Knol et al. (152)).

Example 1: The joint effect of alcohol and smoking on oral cancer.

Consider the case-control data from Rothman and Keller (112), which studied the joint effect of alcohol and smoking on oral cancers. This dataset was used previously by Hosmer and Lemeshow (129), and Zou (164) and its included in our package as ‘OCdata’ for easy replication of our analysis by the user. It contains the two exposures, alcohol (‘alc’) and smoking (‘smk’), and the outcome, oral cancer (‘oc’) as dichotomous variables (0 or 1, indicating absence or presence

respectively). Suppose we are considering interaction and we want to compute the CIs for the measures of additive interaction using the MOVER method, we will start by fitting the following logistic regression model with an interaction term for alcohol and smoking on oral cancer:

$$\text{Logit}(\Pr(OC_i \mid \text{Alcohol, Smoking})) = \alpha + \beta_1.\text{Alcohol}_i + \beta_2.\text{Smoking}_i + \beta_3.\text{Alcohol*Smoking}_i$$

in R as:

- `m = glm(oc~alc*smk, family = binomial(link = 'logit'), data = OCdata)`

The fitted 'm' object is then passed to the *'interactionR()'* function:

- `table_object = interactionR(m, exposure_names = c("alc", "smk"), ci.type = "mover", ci.level = 0.95, em=F, recode = F)`

This returns a list object of class *interactionR* which includes a data frame containing all the effect estimates that are necessary to fully report effect modification or interaction. This data frame and other components of the list are accessible to the user for further manipulation, if desired. Importantly, the output object is formatted in such a way that the data frame can be processed by the tabling function *interactionR_table()*. An example is:

- `interactionR_table(table_object)`

The tabling function will save a publication-ready table as a word document (Table 6) to the user's working directory, if desired, which corresponds to Table 5b with CIs for RERI, AP and SI estimated with the MOVER method (the set of arguments selected in the original *interactionR()* call). The returned point estimate and CI for RERI (3.74; 95% CI: -11.43, 21.87) (Table 6) is as reported by Zou for this data (164) – If the 'ci.type' argument in *interactionR()* had been set to

"delta", the CIs for this trio of interaction measures would be as reported by Hosmer and Lemeshow (129). Following the same steps above, when the package was applied to the data of a recently published article on the modification of the effect of tobacco smoking on oral HPV by HIV (166), the results were the same (data not shown).

Table 6: Interaction of alcohol and smoking on oral cancers as returned by the *'interactionR_table()'* function.

	smk absent	**smk present**	**Effect of smk within the strata of alc**
	OR [95% CI]	OR [95% CI]	OR [95% CI]
alc absent	1 [Reference]	2.96 [0.68, 12.91]	2.96 [0.68, 12.91]
alc present	3.33 [0.7, 15.86]	9.04 [2.64, 30.91]	2.71 [1, 7.37]
Effect of alc within the strata of smk	3.33 [0.7, 15.86]	3.05 [1.29, 7.18]	
Multiplicative scale	0.91 [0.15, 5.42]		
RERI	3.74 [-11.43, 21.87]		
AP	0.41 [-0.38, 0.81]		
SI	1.87 [0.65, 5.42]		

In addition to the main function described here, the package also provides standalone functions for each of the CI estimating methods: *interactionR_mover(), interactionR_delta(),* and *interactionR_boot().* The latter implements the percentile bootstrapping of CIs of additive interaction measures as described by Assmann et al (163) and its described below.

Example 2: Effect of sports participation and smoking on herniated lumbar disc.

To illustrate the *interactionR_boot()* function, consider the case-control data of the effect of sports participation and smoking on herniated lumbar disc examined by Assmann et al. in their analysis. The dataset is available in the package as 'HDiscdata' and contains three binary variables: i) the outcome herniated lumbar disc, 'h'; and ii & iii) the exposures sports participation, 'ns', and smoking 'smk'. The function accepts the following arguments:

- *model*: A fitted model of class glm. Requires the exposures with interaction term to be listed first before any other covariates/confounders (if applicable).
- *ci.level; em; recode*: As described for *interactionR()* above.
- *seed*: The random number seed to use for generating the bootstrap samples (for reproducibility). Default is 12345 but can be set to any number.
- *s*: Number of bootstrap resampling. Default is 1000

Again, first, we fit a regression model:

```
m2 = glm(h ~ ns*smk, family = binomial(link = 'logit'), data = HDiscdata)
```

Then, pass the object to the *interactionR_boot()* function as follows:

```
table_object2 = interactionR_boot(m2, ci.level = 0.95, em = F, recode = F, seed = 12345, s = 1000)
```

This runs a non-parametric bootstrap sample 1000 times with replacements and a percentile CI. The function also returns a list object of class *interactionR* containing all the desired estimates and manipulable by the tabling function - *interactionR_table()* - described earlier. Calling *interactionR_table()* on the returned object produces Table 7, with estimates for the CI of RERI and AP similar to that reported by Assmann et al. for this data.

Table 7: Interaction of sports participation and smoking on herniation of lumbar disc as returned by the *'interactionR_table()'* function.

	smk absent	**smk present**	**Effect of smk within the strata of ns**
	OR [95% CI]	OR [95% CI]	OR [95% CI]
ns absent	1 [Reference]	1.88 [1.29, 2.73]	1.88 [1.29, 2.73]
ns present	2.38 [1.27, 4.46]	1.98 [1.12, 3.48]	0.83 [0.3s9, 1.75]
Effect of ns within the strata of smk	2.38 [1.27, 4.46]	1.05 [0.61, 1.83]	
Multiplicative scale	0.44 [0.19, 1.02]		
RERI	-1.28 [-3.63, 0.54]		
AP	-0.65 [-2.21, 0.18]		
SI	0.43 [0.07, 1.41]		

Furthermore, some base R functions are available to the user to further manipulate some parts of the output object from *interactionR_boot()*. A simple example is:

- `hist(table_object2$bootstrap)`

which produces histograms of the distribution of each of the three bootstrapped parameters (RERI, AP and SI), allowing the user to inspect the overall performance and accuracy of the returned estimates.

Discussion

In this report, we introduce the 'interactionR' package in R which provides a suit of user-friendly functions that makes assessment and full reporting of effect modification and interaction a one-stop activity. Also, to the best of our knowledge, the package provides the first set of functions in R that implements alternative techniques for CIs estimation for additive interaction measures other than the delta method. Future iterations of the package will include implementations of the likelihood-based CI estimation for RERI from a linear risk model as proposed by Richardson and Kaufman (165) and updates for any additional robust methods that may be developed.

The examples given in this report computed additive interaction measures from odds ratios estimated from logistic regression models of case-control datasets used in previous analyses on the subject matter (129, 163, 164). This has become the fairly standard approach in the literature (113). However, the resultant RERI and SI from this approach will be an exaggerated estimate of any underlying additive interaction if the OR from the study does not closely approximate RR (167). Therefore, users of this approach must ensure that the conditions (such as – but not limited to – rare disease assumption) required for ORs to closely approximate RRs are fully satisfied in their study. Although, models that directly estimate RR are not as stable and versatile especially in the presence of confounding adjustments as are logistic regression models, RERI, AP, and SI that are based on RR are better estimants and should be used if the study design permits (164). We show how to use the package with RR estimated from a log-binomial model (168) in the supplementary file (Appendix II).

In conclusion, we provide herein an easily applicable tool to encourage reporting of interaction with multiple options for estimating Cis in epidemiological studies. The package is available on the Comprehensive R Archive Network (CRAN) at https://cran.r-

project.org/package=interactionR and can be installed from within R using *install.packages("interactionR")*.

Acknowledgments

I thank Sreenath A. Madathil for being the first tester and for critical feedback about the design of the earlier version of the package, and Belinda Nicolau for critical comments regarding the manuscript.

8 DISCUSSION

This section provides a summary of the two manuscripts included in this project and in addition, expands on the limitations to the interpretation of our results and finally, provides a discussion on the public health and epidemiological research implications of our findings.

8.1 Summary of Research

The overarching goal of this thesis was to characterize a high-risk group that will make feasible any future oral HPV screening for prevention of HNCs. This was achieved in Manuscript I through effect modification analysis of the effect of smoking on oral HPV by HIV. However, current statistical programs do not facilitate the sets of analyses necessary to fully report effect modification showed in Tables 2, 3, and 4. Therefore, to be able to conduct the analysis presented in Manuscript I, I introduced a novel R package (Manuscript II). The package allows users to generate a publication-ready interaction table following easy steps as recommended by the guidelines. Furthermore, it also implements for the first time in R, alternatives to the delta method for the estimation of CIs for measures of additive interaction.

While previous studies have shown smoking and HIV are independent risk factors for oral HPV (α-types), our findings show for the first time that the effect of smoking on oral HPV is more pronounced in PLWH.

We demonstrated a positive effect modification for most HPV types linked to HPV-related HNCs including high-risk α-types, HPV-16 (although not as strong) and β-1 species (HPV-5 and -8). Importantly, when we conducted sensitivity analysis modelling the tobacco smoking variable as never, past and current smoker, the positive effect modification by HIV indicated a dose-response relationship between this variable categorization and three outcomes (i.e., high-risk α-types, HPV-

16, and β-1 species). However, effect modification was not evident for all the assessed oral HPV outcomes (e.g., γ-genus/types).

8.2 Limitations and Strengths

In addition to the limitations already identified in Manuscript I, viz. lack of temporality due to cross-sectional analysis, and the possibility of increased oral HPV false positives due to the high sensitivity of the NGS method, other issues deserve attention and they are described below.

8.2.1 Statistical power

The post-hoc power considerations showed that the study had a 30% power to detect additive interaction due to the relatively small sample size. This probably explains the poor precision around RERI for most of the oral HPV outcomes assessed. However, the direction of RERI (i.e., greater, lesser or equal to zero) is enough to interpret an underlying additive interaction (positive, negative or none) as shown in the examples in the article by Knol and Vanderweele (115); this is because unless a study is specifically powered for interaction, relying on lack of statistical significance is a poor way to exclude differential effects (169).

It's also important to note that a larger sample size is required for the detection of an interaction effect (i.e., a sub-group analysis) than for a corresponding overall effect (169). In other words, if the magnitude of an interaction effect is equal to an overall effect, then a study designed with 80% power to detect an overall effect will have just 29% power to detect any interaction effect (169, 170). Consequently, although the MuCOSAL study is adequately powered for the pre-planned cohort design, it is under-powered for an interaction analysis. To achieve adequate statistical power, the sample size needs to be increased four-fold, and even more if the magnitude of the interaction effect is less than the overall effect (169, 170). Evidently, this is not the case for most studies that carries out sub-group analyses in the literature.

8.2.2 Information bias

Information bias or misclassification bias ensues from incorrect determination or measurement of exposure or outcome, or both (171). This bias is a threat to the validity of all observational studies (171) and the MuCOSAL study is no exception. The self-reported HIV status – one of the two main exposures in Manuscript I – was ascertained from clinical records, therefore, little to no misclassification was anticipated for this exposure. However, measurement of tobacco smoking, the other main exposure, was strictly based on the self-reported lifetime smoking history by the participants, making this exposure susceptible to information bias; especially, since tools that improve recall, such as the life-grid technique (172) was not employed in the study. However, all dimensions of adult self-reported smoking history have been shown to have a high reliability (173). Moreover, the MuCOSAL study was designed as a prospective cohort study, thus any misclassification of this exposure will be more likely non-differential, in which case the results are more likely biased towards the null (i.e., conservative estimates) rather than exaggerated (174, 175), this is in contrast to case-control designs that are more susceptible to recall bias, leading to differential misclassification and an exaggerated effect estimate (175-177).

8.2.3 Other methodological issues

The cross-sectional analysis of the baseline data of a planned cohort study – as was done in Manuscript I – is susceptible to a couple of methodological challenges. For example, to minimize selection bias, cross-sectional studies – in particular, those that aim to assess the prevalence of an exposure or a disease in a population – requires the study sample to be representative of the underlying population under study (178), this is achieved through strategies such as employing random sampling techniques, very large sample sizes and ensuring high response rates, among others (175, 178-180).

In contrast, representativeness issues are not a threat to the validity of prospective cohort studies (181), and this has been demonstrated empirically in the literature (182) – the principal source of selection bias for cohort studies is in differential loss to follow-up (183, 184). Due to this, techniques that are normally employed to ensure representativeness for cross-sectional studies are not the focus while establishing a cohort at baseline and this is a potential source of bias for analysing such data cross-sectionally.

However, the primary aim for Manuscript I was to compare oral HPV by smoking and HIV status and not to generate population prevalence estimates. Therefore, for internal comparisons to be valid, individuals needed to come from the same source population, and the selection of exposed and unexposed individuals should not be done in a way that is systematically related to the outcome of interest (in this case HPV detection by NGS assay), as was the case with our study.

8.3 Implications for Public Health

There are three levels of prevention in public health: primary, secondary, and tertiary (185, 186). Primary prevention attempts to prevent a disease before it occurs – e.g., smoking cessation to prevent lung cancer – while secondary prevention aims to detect a disease in its nascent stage (prior to the appearance of signs or symptoms) to prevent further morbidities and ensure full recovery – e.g., mammograms to detect breast cancers (185, 186). Tertiary prevention simply aims to ameliorate the impact of an ongoing disease (usually chronic diseases) to reduce morbidity, improve quality of life and prevent mortality.

Furthermore, there are two main paradigms of disease prevention: The high-risk or targeted approach and the population-based approach (187-189). The high-risk strategy aims to reduce the burden of a disease by targeting prevention strategies (primary or secondary) to the groups of individuals that present with risk-factors responsible for that disease (187-189). In contrast, the

population-based approach considers the entire population as the focus of prevention efforts rather than just a high-risk group. A more refined variant of this population-wide strategy is the so-called common risk factor approach, where multiple diseases in the population are prevented by targeting the risk factors that are common to them (190, 191). For example, because high consumption of refined sugars is a common risk factor for obesity, diabetes, and dental caries, focusing on this single risk factor – e.g., through a sugar tax – can reduce the burden of these diseases in the population (190).

Whereas there are clear advantages and disadvantages to both approaches (i.e., high-risk vs population-based) (186, 188), the findings reported in this thesis are more aligned with the high-risk prevention strategy. Results from Manuscript I identify PLWH who smoke as a targetable group for public health interventions. For example, this group due to their concomitant higher oral HPV prevalence could be preferentially recruited for HNCs screening based on oral HPV tests (secondary prevention). In addition, public health strategies such as smoking cessation and education will be more impactful in this group (primary prevention).

8.4 Implications for Epidemiological Research

It is also anticipated that the tools introduced in Manuscript II will help authors of epidemiological research in general to better report all dimensions of assessed interactions. This is not a trivial matter as shown by a very recent article (192) illustrating the issues solved by the package. On page 7 of the article, the authors reported two results of exposure interaction on Psoriasis: sex and HPV, and age and HPV. They only reported p-values for interaction: ‘non-significant’ for the former and ‘significant’ for the latter. While no statement was given on the scale in which the interactions were assessed on, an educated guess based on the analytical strategy used (Cox proportional model) was strongly suggestive of a multiplicative scale. Importantly, if two

exposures affect an outcome, then the absence of interaction in one scale necessarily implies its presence in another (109, 113). Therefore, because the authors dismissed interactions between sex and HPV solely based on a multiplicative scale assessment, they may have missed a more important interaction on the additive scale – since they did not check on this latter scale. If additive interaction was assessed, their results might perhaps show that males are more at risk for HPV-induced Psoriasis than females (indeed, this was suggested by the Table 5 in their article), which will influence the clinical management of this condition. However, due to lack of full reporting by the authors, the readers are deprived of the effect estimates necessary to come to this important conclusion; the R package introduced in Manuscript II solves this issue.

9. CONCLUSIONS

The following conclusions could be made from the manuscripts included in this thesis:

1. PLWH who smoke are a distinct high-risk group targetable for any future oral HPV screening application for HNC prevention
2. Public health interventions on smoking for oral HPV prevention may be more rewarding in PLWH
3. PLWH have a higher prevalence of all types of oral HPV (α, β and γ) compared to HIV-negative individuals
4. The *interactionR* package confers notable advantages to existing R functions for effect modification and interaction analyses.

9 REFERENCES

1. Laprise C, Madathil SA, Schlecht NF, Castonguay G, Soulières D, Nguyen-Tan PF, et al. Human papillomavirus genotypes and risk of head and neck cancers: Results from the HeNCe Life case-control study. Oral Oncology. 2017;69:56-61.

2. de Martel C, Plummer M, Vignat J, Franceschi S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. Int J Cancer. 2017;141(4):664-70.

3. Mahal BA, Catalano PJ, Haddad RI, Hanna GJ, Kass JI, Schoenfeld JD, et al. Incidence and Demographic Burden of HPV-associated Oropharyngeal Head and Neck Cancers in the United States. Cancer Epidemiology Biomarkers & Prevention. 2019:cebp.0038.2019.

4. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, et al. Human Papillomavirus and Rising Oropharyngeal Cancer Incidence in the United States. Journal of Clinical Oncology. 2011;29(32):4294-301.

5. Beachler DC, Sugar EA, Margolick JB, Weber KM, Strickler HD, Wiley DJ, et al. Risk Factors for Acquisition and Clearance of Oral Human Papillomavirus Infection Among HIV-Infected and HIV-Uninfected Adults. American Journal of Epidemiology. 2015;181(1):40-53.

6. Bottalico D, Chen Z, Dunne A, Ostoloza J, McKinney S, Sun C, et al. The Oral Cavity Contains Abundant Known and Novel Human Papillomaviruses From the Betapapillomavirus and Gammapapillomavirus Genera. 2011;204(5):787-92.

7. Beachler DC, D'Souza G. Oral human papillomavirus infection and head and neck cancers in HIV-infected individuals. Current Opinion in Oncology. 2013;25(5):503-10.

8. Kirk GD, Merlo CA. HIV Infection in the Etiology of Lung Cancer: Confounding, Causality, and Consequences. Proceedings of the American Thoracic Society. 2011;8(3):326-32.

9. Fakhry C, Gillison ML, D'Souza G. Tobacco use and oral HPV-16 infection. JAMA. 2014;312(14):1465-7.

10. Kero K, Rautava J, Syrjänen K, Willberg J, Grenman S, Syrjänen S. Smoking increases oral HPV persistence among men: 7-year follow-up study. European Journal of Clinical Microbiology & Infectious Diseases. 2014;33(1):123-33.

11. Beachler DC, Weber KM, Margolick JB, Strickler HD, Cranston RD, Burk RD, et al. Risk factors for oral HPV infection among a high prevalence population of HIV-positive and at-risk HIV-negative adults. Cancer Epidemiol Biomarkers Prev. 2012;21(1):122-33.

12. Knol MJ, Egger M, Scott P, Geerlings MI, Vandenbroucke JP. When One Depends on the Other: Reporting of Interaction in Case-Control and Cohort Studies. Epidemiology. 2009;20(2):161-6.

13. ROTHMAN KJ, GREENLAND S, WALKER AM. CONCEPTS OF INTERACTION. American Journal of Epidemiology. 1980;112(4):467-70.

14. De Villiers E-M, Fauquet C, Broker TR, Bernard H-U, Zur Hausen H. Classification of papillomaviruses. Virology. 2004;324(1):17-27.

15. Cladel NM, Budgeon LR, Balogh KK, Cooper TK, Hu J, Christensen ND. Mouse papillomavirus MmuPV1 infects oral mucosa and preferentially targets the base of the tongue. Virology. 2016;488:73-80.

16. Kreimer AR, Clifford GM, Boyle P, Franceschi S. Human Papillomavirus Types in Head and Neck Squamous Cell Carcinomas Worldwide: A Systematic Review. Cancer Epidemiology Biomarkers & Prevention. 2005;14(2):467-75.

17. Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, et al. A review of human carcinogens--Part B: biological agents. Lancet Oncol. 2009;10(4):321-2.

18. Pierce Campbell CM, Messina JL, Stoler MH, Jukic DM, Tommasino M, Gheit T, et al. Cutaneous human papillomavirus types detected on the surface of male external genital lesions: a case series within the HPV Infection in Men Study. J Clin Virol. 2013;58(4):652-9.

19. Agalliu I, Gapstur S, Chen Z, Wang T, Anderson RL, Teras L, et al. Associations of Oral α-, β-, and γ-Human Papillomavirus Types With Risk of Incident Head and Neck Cancer. JAMA Oncology. 2016;2(5):599.

20. Bäfverstedt B. Condylomata acuminata--past and present. Acta dermato-venereologica. 1967;47(5):376.

21. Ciuffo G. Innesto positivo con filtrato di verruca volgare. Giorn Ital Mal Venereol. 1907;48(1):12-7.

22. Strauss MJ, Shaw EW, Bunting H, Melnick JL. "Crystalline" Virus-Like Particles from Skin Papillomas Characterized by Intranuclear Inclusion Bodies. Proceedings of the Society for Experimental Biology and Medicine. 1949;72(1):46-50.

23. Rigoni-Stern D. Fatti statistici relativi alle malattie cancerose. Giorn Prog Patol Terap. 1842;2:507-17.

24. Rous P, Beard J. Carcinomatous changes in virus-induced papillomas of the skin of the rabbit. Proceedings of the Society for Experimental Biology and Medicine. 1935;32(4):578-80.

25. Rous P, Friedewald WF. The effect of chemical carcinogens on virus-induced rabbit papillomas. The Journal of experimental medicine. 1944;79(5):511-38.

26. Zur Hausen J, Schulte-Holthausen H, Wolf HJ, Dörries K, Egger H. Attempts to detect virus-specific DNA in human tumors. II. Nucleic acid hybridizations with complementary RNA of human herpes group viruses. International journal of cancer= Journal international du cancer. 1974;13(5):657-64.

27. Krcmar M, Suchankova A, Kanka J, Vonka V. Prospective study on the relationship between cervical neoplasia and herpes simplex type 2 virus. III. Presence of herpes simplex type-2 antibody in sera of subjects who developed cervical neoplasia later in the study. Int J Cancer. 1986;38(2):161-5.

28. Zur Hausen H. Human papillomaviruses and their possible role in squamous cell carcinomas. Current topics in microbiology and immunology: Springer; 1977. p. 1-30.

29. Zur Hausen H. Condylomata acuminata and human genital cancer. Cancer res. 1976;36(2 pt 2):794.

30. Dürst M, Gissmann L, Ikenberg H, zur Hausen H. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. Proc Natl Acad Sci U S A. 1983;80(12):3812-5.

31. Boshart M, Gissmann L, Ikenberg H, Kleinheinz A, Scheurlen W, zur Hausen H. A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer. EMBO J. 1984;3(5):1151-7.

32. Dlugonska H. [Harald zur Hausen--a scientist with passion. Vaccine against cervical cancer]. Wiad Parazytol. 2009;55(3):191-4.

33. Hawley-Nelson P, Vousden KH, Hubbert NL, Lowy DR, Schiller JT. HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. EMBO J. 1989;8(12):3905-10.

34. Munoz N, Bosch X, Kaldor J. Does human papillomavirus cause cervical cancer? The state of the epidemiological evidence. British Journal of Cancer. 1988;57(1):1-5.

35. Franco E. Measurement errors in epidemiological studies of human papillomavirus and cervical cancer. IARC scientific publications. 1992(119):181-97.

36. Bosch FX, Lorincz A, Muñoz N, Meijer C, Shah KV. The causal relation between human papillomavirus and cervical cancer. Journal of clinical pathology. 2002;55(4):244-65.

37. Papillomaviruses H. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans.

38. Estêvão D, Costa NR, Gil da Costa RM, Medeiros R. Hallmarks of HPV carcinogenesis: The role of E6, E7 and E5 oncoproteins in cellular malignancy. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2019;1862(2):153-62.

39. Graham SV, Faizo AAA. Control of human papillomavirus gene expression by alternative splicing. Virus Res. 2017;231:83-95.

40. Fehrmann F, Laimins LA. Human papillomaviruses: targeting differentiating epithelial cells for malignant transformation. Oncogene. 2003;22(33):5201-7.

41. Fouad YA, Aanei C. Revisiting the hallmarks of cancer. Am J Cancer Res. 2017;7(5):1016-36.

42. Griffin LM, Cicchini L, Xu T, Pyeon D. Human keratinocyte cultures in the investigation of early steps of human papillomavirus infection. Methods Mol Biol. 2014;1195:219-38.

43. Groeger S, Meyle J. Oral Mucosal Epithelial Cells. Frontiers in Immunology. 2019;10(208).

44. Hutton E, Paladini RD, Yu QC, Yen M, Coulombe PA, Fuchs E. Functional differences between keratins of stratified and simple epithelia. J Cell Biol. 1998;143(2):487-99.

45. Stubenrauch F, Laimins LA. Human papillomavirus life cycle: active and latent phases. Semin Cancer Biol. 1999;9(6):379-86.

46. zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. Nat Rev Cancer. 2002;2(5):342-50.

47. Pyeon D, Pearce SM, Lank SM, Ahlquist P, Lambert PF. Establishment of human papillomavirus infection requires cell cycle progression. PLoS Pathog. 2009;5(2):e1000318.

48. Joyce JG, Tung JS, Przysiecki CT, Cook JC, Lehman ED, Sands JA, et al. The L1 major capsid protein of human papillomavirus type 11 recombinant virus-like particles interacts with heparin and cell-surface glycosaminoglycans on human keratinocytes. J Biol Chem. 1999;274(9):5810-22.

49. Giroglou T, Florin L, Schafer F, Streeck RE, Sapp M. Human papillomavirus infection requires cell surface heparan sulfate. J Virol. 2001;75(3):1565-70.

50. Oyervides-Muñoz MA, Pérez-Maya AA, Rodríguez-Gutiérrez HF, Gómez-Macias GS, Fajardo-Ramírez OR, Treviño V, et al. Understanding the HPV integration and its progression to cervical cancer. Infect Genet Evol. 2018;61:134-44.

51. Martinez-Zapien D, Ruiz FX, Poirson J, Mitschler A, Ramirez J, Forster A, et al. Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. Nature. 2016;529(7587):541-5.

52. Tomaic V. Functional Roles of E6 and E7 Oncoproteins in HPV-Induced Malignancies at Diverse Anatomical Sites. Cancers (Basel). 2016;8(10).

53. Katzenellenbogen R. Telomerase Induction in HPV Infection and Oncogenesis. Viruses. 2017;9(7).

54. Zheng ZM, Wang X. Regulation of cellular miRNA expression by human papillomaviruses. Biochim Biophys Acta. 2011;1809(11-12):668-77.

55. Wang X, Wang HK, Li Y, Hafner M, Banerjee NS, Tang S, et al. microRNAs are biomarkers of oncogenic human papillomavirus infections. Proc Natl Acad Sci U S A. 2014;111(11):4262-7.

56. Wongjampa W, Ekalaksananan T, Chopjitt P, Chuerduangphui J, Kleebkaow P, Patarapadungkit N, et al. Suppression of miR-22, a tumor suppressor in cervical cancer, by human papillomavirus 16 E6 via a p53/miR-22/HDAC6 pathway. PLoS One. 2018;13(10):e0206644.

57. Olatunji I. Potential application of tumor suppressor microRNAs for targeted therapy in head and neck cancer: A mini-review. Oral Oncol. 2018;87:165-9.

58. Tommasino M. HPV and skin carcinogenesis. Papillomavirus Res. 2019;7:129-31.

59. Viarisio D, Müller-Decker K, Accardi R, Robitaille A, Dürst M, Beer K, et al. Beta HPV38 oncoproteins act with a hit-and-run mechanism in ultraviolet radiation-induced skin carcinogenesis in mice. PLoS Pathog. 2018;14(1):e1006783.

60. Rollison DE, Viarisio D, Amorrortu RP, Gheit T, Tommasino M. An Emerging Issue in Oncogenic Virology: the Role of Beta Human Papillomavirus Types in the Development of Cutaneous Squamous Cell Carcinoma. J Virol. 2019;93(7):e01003-18.

61. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2018;68(6):394-424.

62. Sturgis EM, Cinciripini PM. Trends in head and neck cancer incidence in relation to smoking prevalence: an emerging epidemic of human papillomavirus-associated cancers? Cancer. 2007;110(7):1429-35.

63. Ndiaye C, Mena M, Alemany L, Arbyn M, Castellsague X, Laporte L, et al. HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers: a systematic review and meta-analysis. Lancet Oncol. 2014;15(12):1319-31.

64. Saraiya M, Unger ER, Thompson TD, Lynch CF, Hernandez BY, Lyu CW, et al. US assessment of HPV types in cancers: implications for current and 9-valent HPV vaccines. J Natl Cancer Inst. 2015;107(6):djv086-djv.

65. Volesky KD, El-Zein M, Franco EL, Brenner DR, Friedenreich CM, Ruan Y. Cancers attributable to infections in Canada. Preventive Medicine. 2019;122:109-17.

66. Gellin B, Modlin JF, Barr E, Tamms G. Quadrivalent Human Papillomavirus Vaccine. Clinical Infectious Diseases. 2007;45(5):609-17.

67. Satterwhite CL, Torrone E, Meites E, Dunne EF, Mahajan R, Ocfemia MC, et al. Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008. Sex Transm Dis. 2013;40(3):187-93.

68. Mena M, Taberna M, Monfil L, Arbyn M, de Sanjosé S, Bosch FX, et al. Might Oral Human Papillomavirus (HPV) Infection in Healthy Individuals Explain Differences in HPV-Attributable Fractions in Oropharyngeal Cancer? A Systematic Review and Meta-analysis. The Journal of infectious diseases. 2019;219(10):1574-85.

69. Kreimer AR, Bhatia RK, Messeguer AL, Gonzalez P, Herrero R, Giuliano AR. Oral human papillomavirus in healthy individuals: a systematic review of the literature. Sex Transm Dis. 2010;37(6):386-91.

70. Rieth KKS, Gill SR, Lott-Limbach AA, Merkley MA, Botero N, Allen PD, et al. Prevalence of High-Risk Human Papillomavirus in Tonsil Tissue in Healthy Adults and Colocalization in Biofilm of Tonsillar Crypts. JAMA Otolaryngology–Head & Neck Surgery. 2018;144(3):231-7.

71. Knight GL, Needham L, Roberts S, Ward D. Pilot study investigating the prevalence of oral Human Papilloma Viral (HPV) infection in young adults. 2016.

72. Chung CH, Bagheri A, D'Souza G. Epidemiology of oral human papillomavirus infection. Oral oncology. 2014;50(5):364-9.

73. Kreimer AR, Alberg AJ, Daniel R, Gravitt PE, Viscidi R, Garrett ES, et al. Oral human papillomavirus infection in adults is associated with sexual behavior and HIV serostatus. The Journal of infectious diseases. 2004;189(4):686-98.

74. Fakhry C, D'souza G, Sugar E, Weber K, Goshu E, Minkoff H, et al. Relationship between prevalent oral and cervical human papillomavirus infections in human immunodeficiency virus-positive and-negative women. Journal of clinical microbiology. 2006;44(12):4479-85.

75. Coutlee F, Trottier AM, Ghattas G, Leduc R, Toma E, Sanche G, et al. Risk factors for oral human papillomavirus in adults infected and not infected with human immunodeficiency virus. Sex Transm Dis. 1997;24(1):23-31.

76. D'Souza G, Fakhry C, Sugar EA, Seaberg EC, Weber K, Minkoff HL, et al. Six-month natural history of oral versus cervical human papillomavirus infection. International journal of cancer. 2007;121(1):143-50.

77. Gillison ML, Chaturvedi AK, Anderson WF, Fakhry C. Epidemiology of Human Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma. J Clin Oncol. 2015;33(29):3235-42.

78. Gillison ML, Broutian T, Pickard RK, Tong ZY, Xiao W, Kahle L, et al. Prevalence of oral HPV infection in the United States, 2009-2010. JAMA. 2012;307(7):693-703.

79. García-Piñeres AJ, Hildesheim A, Herrero R, Trivett M, Williams M, Atmetlla I, et al. Persistent human papillomavirus infection is associated with a generalized decrease in immune responsiveness in older women. Cancer research. 2006;66(22):11070-6.

80. Sonawane K, Suk R, Chiao EY, Chhatwal J, Qiu P, Wilkin T, et al. Oral Human Papillomavirus Infection: Differences in Prevalence Between Sexes and Concordance With Genital Human Papillomavirus Infection, NHANES 2011 to 2014. Annals of Internal Medicine. 2017;167(10):714-24.

81. Ryser MD, Rositch A, Gravitt PE. Modeling of US Human Papillomavirus (HPV) Seroprevalence by Age and Sexual Behavior Indicates an Increasing Trend of HPV Infection Following the Sexual Revolution. The Journal of infectious diseases. 2017;216(5):604-11.

82. Chaturvedi AK, Graubard BI, Broutian T, Pickard RK, Tong ZY, Xiao W, et al. NHANES 2009-2012 Findings: Association of Sexual Behaviors with Higher Prevalence of Oral Oncogenic Human Papillomavirus Infections in U.S. Men. Cancer Res. 2015;75(12):2468-77.

83. Schnelle C, Whiteman DC, Porceddu SV, Panizza BJ, Antonsson A. Past sexual behaviors and risks of oropharyngeal squamous cell carcinoma: a case-case comparison. Int J Cancer. 2017;140(5):1027-34.

84. Darwich L, Canadas MP, Videla S, Coll J, Molina-Lopez RA, Cobarsi P, et al. Oral human papillomavirus type-specific infection in HIV-infected men: a prospective cohort study among men who have sex with men and heterosexual men. Clin Microbiol Infect. 2014;20(9):O585-9.

85. Markowitz LE, Sternberg M, Dunne EF, McQuillan G, Unger ER. Seroprevalence of human papillomavirus types 6, 11, 16, and 18 in the United States: National Health and Nutrition Examination Survey 2003-2004. J Infect Dis. 2009;200(7):1059-67.

86. Gately I. La diva nicotina: the story of how tobacco seduced the world: Scribner; 2002.

87. Humans IWGotEoCRt. (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, No. 100E.) TOBACCO SMOKING Lyon (FR): International Agency for Research on Cancer; 2012 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK304395/.

88. Organization WH. Prevalence of tobacco smoking. World Health Available at: https://wwwwhoint/gho/tobacco/use/en/ (Accessed on 17 March 2020). 2016.

89. Organization WH. WHO global report on trends in prevalence of tobacco smoking 2000-2025: World Health Organization; 2018.

90. Doll R, Hill AB. Smoking and carcinoma of the lung. British medical journal. 1950;2(4682):739.

91. Wynder EL, Graham EA. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma: a study of six hundred and eighty-four proved cases. Journal of the American medical association. 1950;143(4):329-36.

92. Sasco A, Secretan M, Straif K. Tobacco smoking and cancer: a brief review of recent epidemiological evidence. Lung cancer. 2004;45:S3-S9.

93. Heijink IH, Brandenburg SM, Postma DS, van Oosterhout AJM. Cigarette smoke impairs airway epithelial barrier function and cell–cell contact recovery. European Respiratory Journal. 2012;39(2):419.

94. Reinson T, Henno L, Toots M, Ustav M, Jr., Ustav M. The Cell Cycle Timing of Human Papillomavirus DNA Replication. PLoS One. 2015;10(7):e0131675.

95. Sopori M. Effects of cigarette smoke on the immune system. Nature Reviews Immunology. 2002;2(5):372-7.

96. Abubakar I, Tillmann T, Banerjee A. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015;385(9963):117-71.

97. Services UDoHaH. The Global HIV/AIDS Pandemic US: HIV.gov; 2019 [updated July 31, 2019. Available from: https://www.hiv.gov/hiv-basics/overview/data-and-trends/global-statistics.

98. Shao Y, Williamson C. The HIV-1 epidemic: low- to middle-income countries. Cold Spring Harb Perspect Med. 2012;2(3):a007187-a.

99. CDC. HIV Surveillance Report US: CDC; 2019 [updated November 21, 2019. Available from: https://www.cdc.gov/hiv/statistics/overview/index.html.

100. Dubrow R, Silverberg MJ, Park LS, Crothers K, Justice AC. HIV infection, aging, and immune function: implications for cancer risk and prevention. Current opinion in oncology. 2012;24(5):506-16.

101. Shiels MS, Cole SR, Kirk GD, Poole C. A meta-analysis of the incidence of non-AIDS cancers in HIV-infected individuals. J Acquir Immune Defic Syndr. 2009;52(5):611-22.

102. Chaturvedi AK, Madeleine MM, Biggar RJ, Engels EA. Risk of human papillomavirus-associated cancers among persons with AIDS. J Natl Cancer Inst. 2009;101(16):1120-30.

103. Patel P, Hanson DL, Sullivan PS, Novak RM, Moorman AC, Tong TC, et al. Incidence of types of cancer among HIV-infected persons compared with the general population in the United States, 1992-2003. Ann Intern Med. 2008;148(10):728-36.

104. Frisch M, Biggar RJ, Goedert JJ. Human papillomavirus-associated cancers in patients with human immunodeficiency virus infection and acquired immunodeficiency syndrome. J Natl Cancer Inst. 2000;92(18):1500-10.

105. Silverberg MJ, Chao C, Leyden WA, Xu L, Horberg MA, Klein D, et al. HIV infection, immunodeficiency, viral replication, and the risk of cancer. Cancer Epidemiol Biomarkers Prev. 2011;20(12):2551-9.

106. Tugizov SM, Herrera R, Chin-Hong P, Veluppillai P, Greenspan D, Michael Berry J, et al. HIV-associated disruption of mucosal epithelium facilitates paracellular penetration by human papillomavirus. Virology. 2013;446(1-2):378-88.

107. Sufiawati I, Tugizov SM. HIV-associated disruption of tight and adherens junctions of oral epithelial cells facilitates HSV-1 infection and spread. PLoS One. 2014;9(2):e88803.

108. Shiboski CH, Lee A, Chen H, Webster-Cyriaque J, Seaman T, Landovitz RJ, et al. Human papillomavirus infection in the oral cavity of HIV patients is not reduced by initiating antiretroviral therapy. AIDS (London, England). 2016;30(10):1573-82.

109. Rothman KJ, Greenland S, Lash TL. Modern epidemiology: Lippincott Williams & Wilkins; 2008.

110. Beral V, Bull D, Reeves G. Endometrial cancer and hormone-replacement therapy in the Million Women Study. Lancet. 2005;365(9470):1543-51.

111. VanderWeele TJ. On the Distinction Between Interaction and Effect Modification. Epidemiology. 2009;20(6):863-71.

112. Rothman K, Keller A. The effect of joint exposure to alcohol and tobacco on risk of cancer of the mouth and pharynx. Journal of Chronic Diseases. 1972;25(12):711-6.

113. VanderWeele TJ, Knol MJ. A tutorial on interaction. Epidemiologic Methods. 2014;3(1):33-72.

114. VanderWeele TJ, Robins JM. Four Types of Effect Modification: A Classification Based on Directed Acyclic Graphs. Epidemiology. 2007;18(5):561-8.

115. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol. 2012;41(2):514-20.

116. VanderWeele TJ, Vansteelandt S. Invited Commentary: Some Advantages of the Relative Excess Risk due to Interaction (RERI)—Towards Better Estimators of Additive Interaction. American Journal of Epidemiology. 2014;179(6):670-1.

117. Rothman KJ. Modern epidemiology: Lipincot Williams and Wilkins; 1986.

118. Habbous S, Chu KP, Lau H, Schorr M, Belayneh M, Ha MN, et al. Human papillomavirus in oropharyngeal cancer in Canada: analysis of 5 comprehensive cancer centres using multiple imputation. CMAJ. 2017;189(32):E1030-E40.

119. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: A Cancer Journal for Clinicians. 2020;70(1):7-30.

120. Wissinger E, Griebsch I, Lungershausen J, Foster T, Pashos CL. The Economic Burden of Head and Neck Cancer: A Systematic Literature Review. PharmacoEconomics. 2014;32(9):865-82.

121. Arbyn M, Ronco G, Anttila A, Meijer CJLM, Poljak M, Ogilvie G, et al. Evidence Regarding Human Papillomavirus Testing in Secondary Prevention of Cervical Cancer. Vaccine. 2012;30:F88-F99.

122. D'Souza G, McNeel TS, Fakhry C. Understanding personal risk of oropharyngeal cancer: risk-groups for oncogenic oral HPV infection and oropharyngeal cancer. Ann Oncol. 2017;28(12):3065-9.

123. Zhao M, Rosenbaum E, Carvalho AL, Koch W, Jiang W, Sidransky D, et al. Feasibility of quantitative PCR-based saliva rinse screening of HPV for head and neck cancer. Int J Cancer. 2005;117(4):605-10.

124. Elf JL, Variava E, Chon S, Lebina L, Motlhaoleng K, Gupte N, et al. Prevalence and Correlates of Smoking Among People Living With HIV in South Africa. Nicotine Tob Res. 2018;20(9):1124-31.

125. Mdodo R, Frazier EL, Dube SR, Mattson CL, Sutton MY, Brooks JT, et al. Cigarette Smoking Prevalence Among Adults With HIV Compared With the General Adult Population in the United States: Cross-sectional Surveys. Annals of Internal Medicine. 2015;162(5):335-44.

126. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.

127. Mathur MB, VanderWeele TJ. R Function for Additive Interaction Measures. Epidemiology (Cambridge, Mass). 2018;29(1):e5-e6.

128. Källberg H, Ahlbom A, Alfredsson L. Calculating measures of biological interaction using R. Eur J Epidemiol. 2006;21(8):571-3.

129. Hosmer DW, Lemeshow S. Confidence Interval Estimation of Interaction. Epidemiology. 1992;3(5):452-6.

130. Fonseca AJ, Taeko D, Chaves TA, Amorim LDdC, Murari RSW, Miranda AE, et al. HPV Infection and Cervical Screening in Socially Isolated Indigenous Women Inhabitants of the Amazonian Rainforest. PLOS ONE. 2015;10(7):e0133635.

131. Muller K, Kazimiroff J, Fatahzadeh M, Smith RV, Wiltz M, Polanco J, et al. Oral Human Papillomavirus Infection and Oral Lesions in HIV-Positive and HIV-Negative Dental Patients. J Infect Dis. 2015;212(5):760-8.

132. Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, Ghissassi FE, et al. A review of human carcinogens—Part B: biological agents. The Lancet Oncology. 2009;10(4):321-2.

133. Greenland S, Pearl J, Robins JM. Causal Diagrams for Epidemiologic Research. Epidemiology. 1999;10(1):37-48.

134. Weng H-Y, Hsueh Y-H, Messam LLM, Hertz-Picciotto I. Methods of Covariate Selection: Directed Acyclic Graphs and the Change-in-Estimate Procedure. American Journal of Epidemiology. 2009;169(10):1182-90.

135. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. BMC Medical Research Methodology. 2008;8(1):70.

136. Merchant AT, Pitiphat W. Directed acyclic graphs (DAGs): an aid to assess confounding in dental research. Community Dentistry and Oral Epidemiology. 2002;30(6):399-404.

137. Leffondre K. Modeling Smoking History: A Comparison of Different Approaches. American Journal of Epidemiology. 2002;156(9):813-23.

138. Sharma S, Kleinbaum DG, Kupper LL. Applied Regression Analysis and Other Multivariate Methods. 1978;15(3):498.

139. Nie L, Chu H, Li F, Cole SR. Relative Excess Risk Due to Interaction: Resampling-based Confidence Intervals. Epidemiology. 2010;21(4):552-6.

140. VanderWeele TJ. Reconsidering the denominator of the attributable proportion for interaction. Eur J Epidemiol. 2013;28(10):779-84.

141. Bennett DA. How can I deal with missing data in my study? Australian and New Zealand Journal of Public Health. 2001;25(5):464-9.

142. Dong Y, Peng C-YJ. Principled missing data methods for researchers. Springerplus. 2013;2(1):222-.

143. Garson GD. Missing values analysis and data imputation. Asheboro, NC: Statistical Associates Publishers. 2015.

144. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. BMC Medical Research Methodology. 2017;17(1):162.

145. VanderWeele TJ. Sample Size and Power Calculations for Additive Interactions. Epidemiologic methods. 2012;1(1):159-88.

146. Verma M, Erwin S, Abedi V, Hontecillas R, Hoops S, Leber A, et al. Modeling the Mechanisms by Which HIV-Associated Immunosuppression Influences HPV Persistence at the Oral Mucosa. PLOS ONE. 2017;12(1):e0168133.

147. Gaester K, Fonseca LAM, Luiz O, Assone T, Fontes AS, Costa F, et al. Human papillomavirus infection in oral fluids of HIV-1-positive men: prevalence and risk factors. Sci Rep. 2014;4:6592-.

148. Calvo M, Laguno M, Martinez M, Martinez E. Effects of tobacco smoking on HIV-infected individuals. AIDS Rev. 2015;17(1):47-55.

149. Van Der Waal I. Oral Leukoplakia: Present Views on Diagnosis, Management, Communication with Patients, and Research. Current Oral Health Reports. 2019;6(1):9-13.

150. Shiels MS, Althoff KN, Pfeiffer RM, Achenbach CJ, Abraham AG, Castilho J, et al. HIV Infection, Immunosuppression, and Age at Diagnosis of Non-AIDS-Defining Cancers. Clin Infect Dis. 2017;64(4):468-75.

151. VanderWeele TJ, Knol MJ. A Tutorial on Interaction. Epidemiologic Methods. 2014;3(1).

152. Knol MJ, VanderWeele TJ, Groenwold RHH, Klungel OH, Rovers MM, Grobbee DE. Estimating measures of interaction on an additive scale for preventive exposures. Eur J Epidemiol. 2011;26(6):433-8.

153. Mooij SH, Boot HJ, Speksnijder AG, Stolte IG, Meijer CJ, Snijders PJ, et al. Oral human papillomavirus infection in HIV-negative and HIV-infected MSM. AIDS. 2013;27(13):2117-28.

154. Hernandez CP, Morrow K, Velasco C, Wyczechowska DD, Naura AS, Rodriguez PC. Effects of cigarette smoke extract on primary activated T cells. Cell Immunol. 2013;282(1):38-43.

155. Desrichard A, Kuo F, Chowell D, Lee K-W, Riaz N, Wong RJ, et al. Tobacco Smoking-Associated Alterations in the Immune Microenvironment of Squamous Cell Carcinomas. JNCI: Journal of the National Cancer Institute. 2018;110(12):1386-92.

156. Moyer VA. Screening for Oral Cancer: U.S. Preventive Services Task Force Recommendation Statement. Annals of Internal Medicine. 2014;160(1):55-60.

157. Syrjänen S. Human Papillomaviruses in Head and Neck Carcinomas. New England Journal of Medicine. 2007;356(19):1993-5.

158. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. Epidemiology. 2007;18(6):805-35.

159. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. Epidemiology. 1992;3(5):452-6.

160. Mathur MB, VanderWeele TJ. R Function for Additive Interaction Measures. Epidemiology. 2018;29(1):e5-e6.

161. Li R, Chambless L. Test for Additive Interaction in Proportional Hazards Models. Annals of Epidemiology. 2007;17(3):227-36.

162. Andersson T, Alfredsson L, Källberg H, Zdravkovic S, Ahlbom A. Calculating measures of biological interaction. Eur J Epidemiol. 2005;20(7):575-9.

163. Assmann SF, Hosmer DW, Lemeshow S, Mundt KA. Confidence intervals for measures of interaction. Epidemiology. 1996:286-90.

164. Zou GY. On the Estimation of Additive Interaction by Use of the Four-by-two Table and Beyond. American Journal of Epidemiology. 2008;168(2):212-24.

165. Richardson DB, Kaufman JS. Estimation of the relative excess risk due to interaction and associated confidence bounds. American journal of epidemiology. 2009;169(6):756-60.

166. Alli BY, Burk RD, Fatahzadeh M, Kazimiroff J, Grossberg RM, Smith RV, et al. HIV modifies the effect of tobacco smoking on oral human papillomavirus infection. The Journal of Infectious Diseases. 2020.

167. Kalilani L, Atashili J. Measuring additive interaction using odds ratios. Epidemiologic Perspectives & Innovations. 2006;3(1):5.

168. Knol MJ, Le Cessie S, Algra A, Vandenbroucke JP, Groenwold RHH. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. CMAJ. 2012;184(8):895-9.

169. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. Health Technol Assess. 2001;5(33):1-56.

170. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol. 2004;57(3):229-36.

171. Grimes DA, Schulz KF. Bias and causal associations in observational research. The Lancet. 2002;359(9302):248-52.

172. Berney LR, Blane DB. Collecting retrospective data: Accuracy of recall after 50 years judged against historical records. Social Science & Medicine. 1997;45(10):1519-25.

173. Soulakova JN, Hartman AM, Liu B, Willis GB, Augustine S. Reliability of adult self-reported smoking history: data from the tobacco use supplement to the current population survey 2002-2003 cohort. Nicotine Tob Res. 2012;14(8):952-60.

174. Pearce N, Checkoway H, Kriebel D. Bias in occupational epidemiology studies. Occupational and Environmental Medicine. 2007;64(8):562-8.

175. Colditz GA. Overview of the Epidemiology Methods and Applications: Strengths and Limitations of Observational Study Designs. Critical Reviews in Food Science and Nutrition. 2010;50(sup1):10-2.

176. Schulz KF, Grimes DA. Case-control studies: research in reverse. The Lancet. 2002;359(9304):431-4.

177. Chouinard E, Walter S. Recall bias in case-control studies: An empirical analysis and theoretical framework. Journal of Clinical Epidemiology. 1995;48(2):245-54.

178. Allen M. The SAGE Encyclopedia of Communication Research Methods. 2017.

179. Levin KA. Study design III: Cross-sectional studies. Evidence-Based Dentistry. 2006;7(1):24-5.

180. Setia MS. Methodology Series Module 3: Cross-sectional Studies. Indian J Dermatol. 2016;61(3):261-4.

181. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol. 2013;42(4):1012-4.

182. Nohr EA, Liew Z. How to investigate and adjust for selection bias in cohort studies. Acta Obstet Gynecol Scand. 2018;97(4):407-16.

183. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ, Jr. Selection Bias Due to Loss to Follow Up in Cohort Studies. Epidemiology (Cambridge, Mass). 2016;27(1):91-7.

184. Kristman V, Manno M, Côté P. Loss to Follow-Up in Cohort Studies: How Much is Too Much? Eur J Epidemiol. 2004;19(8):751-60.

185. Last JM, Harris SS, Thuriaux MC, Spasoff RA. A dictionary of epidemiology: International Epidemiological Association, Inc.; 2001.

186. Offord DR. Selection of levels of prevention. Addictive Behaviors. 2000;25(6):833-42.

187. Platt JM, Keyes KM, Galea S. Efficiency or equity? Simulating the impact of high-risk and population intervention strategies for the prevention of disease. SSM Popul Health. 2016;3:1-8.

188. Babu RB, Alam M, Helis E, Fodor JG. Population-based versus high-risk strategies for the prevention of cardiovascular diseases in low- and middle-income countries. Indian Heart J. 2012;64(5):439-43.

189. ROSE G. Sick Individuals and Sick Populations. Int J Epidemiol. 1985;14(1):32-8.

190. Sheiham A, Watt RG. The Common Risk Factor Approach: a rational basis for promoting oral health. Community Dentistry and Oral Epidemiology. 2000;28(6):399-406.

191. Do LG, Scott JA, Thomson WM, Stamm JW, Rugg-Gunn AJ, Levy SM, et al. Common risk factor approach to address socioeconomic inequality in the oral health of preschool children – a prospective cohort study. BMC Public Health. 2014;14(1):429.

192. Chen M-L, Kao W-M, Huang J-Y, Hung Y-M, Wei JC-C. Human papillomavirus infection associated with increased risk of new-onset psoriasis: a nationwide population-based cohort study. Int J Epidemiol. 2020.

193. Knol MJ, van der Tweel I, Grobbee DE, Numans ME, Geerlings MI. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. Int J Epidemiol. 2007;36(5):1111-8.

10 APPENDIX I

10.1 Study Consent Form

INFORMATION LETTER AND CONSENT FORM

COMMON LESIONS IN THE MOUTH AND THROAT :
WHAT MAKES THEM MORE OR LESS SERIOUS?

Project title:	**Natural history of oropharyngeal cancer precursors: A pilot study**
Principal Investigators:	**Drs Marie-Claude Rousseau and Belinda Nicolau, Institut national de la recherche scientifique (INRS)-Institut Armand-Frappier**
Co-Investigators:	**Dr Paul Allison and Dr Eduardo Franco, McGill University** **Drs Nicolas Schlecht and Thomas Belbin, Albert Einstein College of Medicine, NY, USA** **Dr François Coutlée, Centre hospitalier de l'Université de Montréal (CHUM)** **Drs Martin Black and Michael Hier, Jewish General Hospital**
Project sponsored by:	**Canadian Institutes of Health Research**

1. OBJECTIVE

This study, funded by the Canadian Institutes of Health Research, is run by researchers who work in the medical field at the INRS–Institut Armand-Frappier, McGill University, Université de Montréal and the Albert Einstein College of Medicine (New York, USA). The principal investigators of this study are Drs Marie-Claude Rousseau and Belinda Nicolau, at INRS–Institut Armand-Frappier.

Our research team is studying common lesions in the mouth and throat. We want to determine what factors influence some of these lesions to eventually become cancerous, since it is not clear why some lesions progress to cancer and what influences this progression. The factors that we are studying include genetic changes, viral infections, and lifestyle habits. Knowing which factors increase the risk will allow us to propose ways to identify the problematic lesions earlier, and to prevent the development of mouth and throat cancers. Our study will also help doctors make treatment decisions by informing them about which lesions are the most aggressive ones. We are asking for your participation in this research project.

2. PROCEDURES

Three (3) research centres are participating in patient recruitment for this research project and are located in Montreal (Canada), New York (USA) and São Paulo (Brazil). In Montreal, participants will be recruited at the McGill University Health Centre and affiliated teaching hospitals. The study coordination centre is based at the Epidemiology and Biostatistics Unit at INRS–Institut Armand-Frappier. You will be one (1) of forty (40) persons from Montreal who

will be participating in this study. Your participation in the study will last approximately one year.

At your first visit, a research assistant will determine if you can participate in this research study. Your eligibility for the study will be confirmed by the doctor at the time of the clinical exam. If you agree to participate, we will collect a sample of cells from the inside of your cheeks and you will be asked to answer a questionnaire administered by the research assistant.

The cell samples will be collected by your doctor after you have rinsed your mouth with mouthwash. He/she will gently scrape the inside of your cheek at the site of any visible plaques (little red or white spots that can occur on the inside of your mouth or tonsil) with a soft brush, similar to a toothbrush. This will be done at least twice, as well as once on the side of your cheek opposite to the lesion. This procedure, rinsing and taking cells from the inside of your cheek, will last less than one minute. The brush is painless and usually does not cause bleeding. In total we may perform two (2) to six (6) brushings at each visit. You may be asked by your doctor to return for additional visits if he/she feels that your medical care requires it. We may also collect brushings of your lesion at these visits. We will not be conducting a biopsy of the lesion. However, if your doctor needs to do one for your medical follow-up, we are also seeking access to the results of that biopsy.

The research assistant will then conduct an interview, recording your smoking, alcohol drinking, eating habits, sexual behaviour, and oral hygiene habits. The interview will last approximately between 20 to 60 minutes, and your answers will be kept confidential.

You will be contacted for a repeat visit, six (6) months after the first visit and six (6) months after the second visit. Cells will be collected by brushing the inside of your cheeks at each visit, and a shorter interview about your recent social and eating habits will be conducted.

3. GENETIC RESEARCH

Using research laboratory methods, cells from the inside of your cheek will be tested in order to identify if some genes are activated differently in lesions as compared to normal tissue. Also, we will test for the presence of human papillomavirus (HPV), a virus that often infects the mouth without symptoms. The tests conducted under this study will reveal genetic information, but it will not be directly relevant to your health care or in determining your risk for cancer. All laboratory specimens and reports will be identified only by a coded number to ensure subject confidentiality. They will be kept in secured facilities and made accessible only to authorized study personnel (see section 7).

The genetic results will not be communicated to you. This information will be protected from use by anyone outside of the research team. Since it is still unknown if the genetic changes studied will be predictive of cancer development, genetic counselling will not be provided to you either.

4. POSSIBLE BENEFITS

If you agree to take part in this study, there will be no direct medical benefit to you. However, the information learned from this study may benefit other people in the future.

5. POSSIBLE RISKS/DISCOMFORTS

The collection of cells from the inside of your cheeks will only pose a minor risk. Sampling with the brush is painless and should not cause breakdown or cuts on the inside of your cheek. Risks include the possibility of minor discomfort and bleeding from these sites. If some bleeding occurs, you will be asked to rinse your mouth with salty water (sterile saline solution), and the bleeding will stop within minutes. It is important that you report any and all symptoms or adverse reactions to your doctor. You will be monitored for adverse reactions by the study staff.

Some of the questions that will be asked during the interview are of a personal nature and may cause you slight embarrassment or discomfort. You may see the questions before deciding whether or not to participate in this study. They are attached to this consent form.

6. VOLUNTARY PARTICIPATION

Your participation in this study is completely free and voluntary. You may take the time necessary to reflect on your decision and discuss your participation in the project with persons close to you before giving us your answer. Your participation and the results of the research will not appear in your medical record. You may choose not to participate and continue to receive standard medical care. If you do not join the study, you will not be penalized. If you join the study, you may withdraw at any time without affecting your future care at the McGill University Health Centre or its affiliated hospitals. If you decide to leave the study at any point, all you have to do is tell the doctor or the nurse/research assistant. No more collection of information or buccal cells will be conducted for the purpose of the study, and your doctor will continue to offer you the same medical care.

7. CONFIDENTIALITY (Who May See Your Records)

The study research records will be kept confidential and you will not be identified in any written or verbal reports. All laboratory specimens and reports will be identified only by a coded number to maintain subject confidentiality. The research records will be kept in a secured area and locked in a filing cabinet in the research offices of the Principal Investigators. Only research personnel authorized by the Principal Investigators will have access to these records.

As all the information in this research project is coded, your personal results cannot be made available to third parties such as employers, governmental organizations, insurance companies or educational institutions. This also applies to your spouse, other members of your family and your physician. However, the members of the research team, the sponsors of this research and other institutions who participate in this study may inspect your research records. These include: INRS–Institut Armand-Frappier, McGill University Health Centre and affiliated hospitals, Centre Hospitalier de l'Université de Montréal and Albert Einstein College of Medicine. All of these groups have been requested to maintain confidentiality.

Your records may also be inspected by the research ethics committees of INRS–Institut Armand-Frappier, Institutional Review Board of McGill University, McGill University Health Centre, Centre Hospitalier de l'Université de Montréal and Albert Einstein College of Medicine.

Cell samples may be shared with other scientists collaborating with the scientists in this department for the described study. No identifying information will be shared. Your identity will remain strictly confidential.

The results of this study may be published or communicated to scientific, clinical and general audiences. However, only statistics will be presented and it will be impossible to identify you or any of the other participants.

8. COMPENSATION
There will be no financial compensation for your participation in this study.

9. CONTACTS FOR QUESTIONS/ACCESS TO CONSENT FORM
If you have any questions about the study or any questions about side effects or injury from the research, you may call Dr. Marie-Claude Rousseau or Dr. Belinda Nicolau at (450) 687-5010. You can communicate with the research team to obtain information on the general progress or the results of the research project. Project updates will be sent by mail one year after the beginning of the recruitment and at the end of the project. However, we will not communicate any individual results to you.

For questions on your rights as a research subject:

You may call the Office of the Institutional Review Board (the committee that oversees research at INRS) at (418) 654-3845 or contact the President of this board:

Monsieur Michel Charbonneau
Président du Comité d'éthique en recherche avec des êtres humains
Institut national de la recherche scientifique
490, rue de la Couronne
Québec (Québec) G1K 9A9
Téléphone : (418) 654-3845
Courriel: michel.charbonneau@iaf.inrs.ca

You can also contact the ombudsman at the Jewish General Hospital at (514) 340-8222 #5833.

10. USE OF SAMPLES FOR FUTURE RESEARCH

Your specimens may be used for future analyses related to the objectives of this study. These may involve additional genetic analyses although the exact nature of these future investigations is not known at this time. We would like to store the samples for future use at INRS–Institut Armand–Frappier, Centre Hospitalier de l'Université de Montréal, and Albert Einstein College of Medicine. These samples will be coded with a unique identification number. The link to the sample will be maintained in a secure database with access limited to the researchers involved in this project. We need to ask your permission to store your samples.

10. USE OF SAMPLES FOR FUTURE RESEARCH (CONTINUING)

Do you give us permission to store your samples for 25 years and use them in future studies that are directly related to the current study?

Yes_____ No_____

Please note: If you agreed but change your mind, you may request in writing that any stored samples be destroyed.

Do you give us permission to have portions of the stored samples distributed to other investigators at INRS–Institut Armand–Frappier, Centre Hospitalier de l'Université de Montréal, Albert Einstein College of Medicine or other institutions for use in research related to the purpose of this study?

Yes_____ No_____

Please note: If you agreed but change your mind, you may request in writing that any shared samples be destroyed.

Do you give us permission to re-contact you in the future to obtain additional information in line with this study?

Yes_____ No_____

Do you give us permission to re-contact you in the future for possible participation in another research project?

Yes_____ No_____

11. SUMMARY AND SIGNATURES

I have read the above description of the research study. I have been told of the risks and benefits involved and all my questions have been answered to my satisfaction. Furthermore, I have been assured that a member of the research team will answer any future questions that I may have. I voluntarily agree to join this study and know that I can withdraw from the study at any time without penalty. By signing this form, I have not given up any of my legal rights.

The study has been explained to me and my questions have been answered to my satisfaction. Upon reflection, I agree to participate in this study.

Name of the Participant

Signature of the Participant ***Date***

Name of the person who explained the consent form

Signature of the person who explained the consent form ***Date***

The research project was approved by the research ethics committee of Institut national de la Recherche Scientifique (INRS) in December 2009 and McGill Institutional Review Board on _____ .

There are three copies of this consent form.
One is for our records, one is for you, and the third one is for the hospital.

10.2 Study Questionnaire

Identification number..... |__|__|-|__|__|__|

Center - Person #

English Questionnaire

Visit #1

Multi-Center Oro-pharyngeal Squamous Atypical Lesion (MuCOSAL) Study

Department of Epidemiology and Population Health

Albert Einstein College of Medicine – New York – USA

Epidemiology & Biostatistics Unit

INRS-Institut Armand-Frappier – Laval – Canada

Faculty of Dentistry & Department of Epidemiology

McGill University – Montreal – Canada

Department of Microbiology and Immunology

Centre de recherche du CHUM – Montreal – Canada

Hospital do Câncer-Departamento de Cirurgia de Cabeça e Pescoço

São Paulo – Brasil

- The identification number is composed of the values for center and person number.
- Person numbers are consecutive numbers within each centre.

1. General guidelines

- The columns should be filled in justified to the right (example: valid | 0 | 1| 2|;
  not valid | 1| 2| |)
- Avoid missing or unknown codes. Insist to get an answer even if it is only an estimation.
- If you do not succeed in getting an answer or estimation, the columns should be filled in with 999.

- When “specify” is written, note your answer on the uninterrupted line.
- For some variables (occupation, ICD code, etc...) a coding method is not provided. They will be coded by the coordinator after the interview or centrally. Boxes are shaded.

2. Local codes

- Local code [LC] change from one place to another. Please refer to the local study investigator (i.e. interviewer, hospital, town and district, etc.).

3. During the interview: Noting, observing, questioning

- If the question is multiple choice, circle or mark an “X” on the chosen answer.
- When “specify” is written and for some variables with “(LC)” (ex: occupation, income, ICD code etc.), note the verbatim answer on the interrupted line. You will confirm with a List of Codes afterwards.
- Avoid missing data or “don’t know” answers. Insist to get an answer even if it’s only an estimation.

4. After the interview: Reviewing, coding and proofreading

- Code all the answers by entering their appropriate code in the boxes. EVERY BOX must contain an entry.
- When “specify” is written and for some variables with “LC” (ex: occupation, income, ICD code etc.), refer to the List of Codes.
- If you did not succeed in getting an estimation and remain with missing data, you should fill the boxes with 99, 999 or 9999 accordingly.
- If the question was not asked or not applicable, code 88, 888 or 8888, accordingly.
- If you remain unsure as how to code your answer, contact the research coordinator to get help.

Index

MuCOSAL STUDY

A1. Identification number..... |__|__|-|__|__|__|

Center - Person #

(Person number = consecutive number, by centre)

A2. Participant's initials (Name – Surname) ..... |__|__|

A3. Country ..... |__|__|

USA

Canada

Brazil

A4. Hospital [LC] ..... |__|__|

A5. Department ..... |__|__|

(01) Otolaryngology

(02) Dentistry

(09) Other (specify) _____

A6. Interviewer [LC] ..... |__|__|

A7. Date of interview ..... |__|__|-|__|__|-|__|__|__|__|

Day Month Year

A8. Beginning of interview ..... |__|__| |__|__|

Hour Min.

Good morning.

My name is..... and first of all I would like to thank you for having accepted to participate in this study. We are conducting a study in _____ and in other countries in order to clarify if certain characteristics and habits of men and women are related to developing some lesions in the mouth and throat. For this purpose, we will interview many patients attending this and other hospitals.

If you agree, I will ask you several questions and the answers will be recorded on this form. I understand that some of these questions may be difficult for you to answer, and exact dates may be hard to remember. Please take as much time as you need so we can gather information, which is as accurate as possible.

I would like to reassure you that all that is said during the interview will be strictly confidential and that the information collected will only be used in scientific reports without any personal name or identifiers being mentioned.

Any likely benefits of the study for the well-being of the population rely on the accuracy of your answers. Therefore, if you do not understand the meaning of any of the questions, please don't be afraid to ask.

At any time you may refuse to continue or to answer specific questions. In addition to the interview, the study includes an examination of your mouth, and a collection of some cells from your mouth.

Can we start now?

Identification number..... |__|__|-|__|__|__|

GENERAL INFORMATION

B1. Sex ..... |__|__|

Male

Female

B2. In this list, which ethnic groups best describes you? ..... |__|__|

|__|__|

Check as many as applicable |__|__|

(00)	White (Caucasian)	ancestors from Europe or European-settled countries such as Argentina, Australia, Brazil, USA, Canada and Cuba
(01)	Hispanic	
(02)	African American	ancestors from Africa, Haïti, etc.
(03)	Arab	ancestors from Middle East
(04)	Asian Indian	ancestors from Indian subcontinent
(05)	East Asian	ancestors from China, Japan, Korea, Thailand, etc.
(06)	Pacific Islander	ancestors from Hawaii, New Zealand, Polynesia, etc.
(07)	First Nation, Inuit, American Indian, Alaska or Northwest Territory Native	
(08)	Mullato	White and Black ancestors
(09)	Mestizo	White and Asian ancestors
(10)	Amazon Indian	
(11)	Other, specify:_____	

B3. How old are you? ..... |__|__|

B4. What is your date of birth? ..... |__|__|-|__|__|-|__|__|__|__|

Day Month Year

B5. In which town or district and in which country were you born? [LC] ..... |__|__|__|__|__|__|

Specify: _____

Town and district Postal code Country

B6. In which town or district do you live now? [LC] ..... |__|__|__|__|__|__|

Specify: _____

B7. For how many years have you been living there? ..... |__|__|

(If less than a year code 00)

B8. Do you live in a rural (farm) or an urban (city) area? ..... |__|__|

Urban

Rural

Identification number..... |__|__|-|__|__|__|

EDUCATION AND OCCUPATION

C1. Did you ever attend school? ..... |__|__|

No (GO TO C3)

No, but I can read and write (GO TO C3)

Yes

C2. What is the highest level of education you have completed? ..... |__|__|

For USA:

Only nursery school

Elementary school

Middle school

High school or GED

Vocational or trade school

Community college

University

Don’t know

For Canada:

Only nursery school

Elementary school

Secondary school

Vocational or trade school

CEGEP or college

University

Don’t know

For Brazil:

Only nursery school

Fundamental school

Intermediate school

Vocational or trade school

University

Don’t know

C3. What was your longest occupation? [LC] ..... |__|__|

Specify:_____

From age To age

C4. What was your age when you started and when you ended this job .... |__|__|__| |__|__|__|

C5. Please describe your job at that time (main task description) ..... |__|__|__|

_____

_____

_____

C6. What was the specialty of the company you worked for? [LC] ..... |__|__|__|

Specify:_____

Identification number..... |__|__|-|__|__|__|

ORAL HEALTH

D1. Do you wear complete dentures? ..... |__|__|

No (GO TO D4)

Yes, bottom only (GO TO D3)

Yes, top only

Yes, top AND bottom

D2. At what age did you start wearing complete top dentures? (Years) ..... |__|__|__|

D3. At what age did you start wearing complete bottom dentures? (Years) ..... |__|__|__|

Interviewer reminder: if the answer at D1 = (02), do not answer D3 and code 888

D4. Do you wear partial dentures? ..... |__|__|

No

Yes, bottom only

Yes, top only

Yes, top AND bottom

D5. How often do you clean your teeth? ..... |__|__|

Never (GO TO D9)

Less than once a week

1-2 times a week

Every other day

Once a day

Twice or more a day

Not applicable (GO TO D9)

D6. Do you use any kind of substance to clean your teeth? ..... |__|__|

No

Toothpaste

Other, specify _____

D7. How often do you usually use mouthwash? ..... |__|__|

Never

I used mouthwash only in the past

Less than once a week

1-2 times a week

Every other day

Once a day

Twice or more a day

D8. Do your gums bleed when you clean your teeth? ..... |__|__|

No

Sometimes

Always or almost always

Not applicable

D9. Have you ever had pain or discomfort due to gum inflammation or periodontitis? ..... |__|__|

No

Yes

Don’t know

Not applicable

D10. Has any dentist/hygienist told you that you have gum diseases/periodontal disease?

…… |__|__|

No

Yes

Don’t know

Not applicable

D11. Has any dentist/hygienist told you that you have deep pockets (underneath your gums)?

…… |__|__|

No

Yes

Don’t know

Not applicable

D12. How often do you see a dentist? …….. |__|__|

(00) Never

(01) Every 6 months

(02) Every year

(03) Every 2-5 years

(04) Once every 5 years

D13. Have you ever had an oral biopsy? (excluding within this project) ..... |__|__|

No (GO TO NEXT SECTION)

Yes

D14. At what age did you have your last biopsy? ..... |__|__|

D15. What did it show? ……. |__|__|__|__|

Specify: _____

Identification number..... |__|__|-|__|__|__|

SMOKING AND CHEWING HABITS

E1. Have you ever smoked in your life? (or chewed, any product, any amount) ..... |__|__|

Never (GO TO NEXT SECTION)

Yes, I still do

Yes, but only in the past

Think of the periods in your life during which you smoked cigarettes, cigars, pipe, chewed tobacco products and/or marijuana, the amount you smoked/chewed/took and other details about the products. Please try to summarize the most important changes in the amount and type of product.

E2. Do you or did you smoke cigarettes? ..... (a) |__|__|

No (GO TO E4)

Yes

Please describe the periods in your life in which you smoke cigarettes and the amounts. Please try to summarize the most important changes in your life regarding the amount and type of product.

	From age (b)	To age (c)	Brand or "hand-rolled" (d)	Number per day (e)	Tobacco type (f)	Filter (g)
1						
2						
3						
4						
5						
					(f)	(g)

(01) Blond	(01) Filter
(02) Black	(02) Non filter
(99) Don't know	(99) Don't know

E3. How deeply do you usually inhale when you smoke cigarettes? ..... |__|__|

I don't inhale

Just into my mouth

Back into my throat

Into my lungs shallow

Into my lungs deep

Don't know (I smoke so little)

E4. Do you or did you smoke cigars? ..... (a) |__|__|

No (GO TO E5)

Yes

Please describe the periods in your life in which you smoke cigars and the amounts. Please try to summarize the most important changes in your life regarding the amount and type of product.

	From age (b)	To age (c)	Brand (d)	Number per day (e)
1				
2				
3				

E5. Do you or did you smoke pipe? ..... (a) |__|__|

No (GO TO E6)

Yes

Please describe the periods in your life in which you smoke pipe and the amounts. Please try to summarize the most important changes in your life regarding the amount and type of product.

	From age (b)	To age (c)	Brand (d)	Number per day (e)
1				
2				
3				

E6. Have you ever chewed tobacco, betel quid, areca nut, or pan massala every day for at least one year? ..... (a) |__|__|

Never (GO TO E7)

Yes, I still do

Only in the past

Please describe the periods in your life in which you chewed and the amounts. Please try to summarize the most important changes in your life regarding the amount and type of product.

	From age (b)	To age (c)	Product (d)	Number of times per day (e)
1				
2				
3				
4				
5				

(d)

(00) Tobacco

(01) Areca nut with tobacco

(02) Betel quid with tobacco

(03) Areca nut without tobacco

(04) Betel quid without tobacco

(05) Pan massala

(06) Other, specify _____

E7. Have you ever snuffed tobacco daily for at least one year? ..... (a) |__|__|

Never (GO TO E8)

Yes, I still do

Only in the past

Please describe the periods in your life in which you snuffed tobacco and the amounts. Please try to summarize the most important changes in your life regarding the amount.

	From age (b)	To age (c)	Number per day (d)
1			
2			
3			

E8. Do/did you ever smoke marijuana (marijuana = hashish = cannabis)? ..... (a) |__|__|

Never (GO TO E10)

Yes, I still do

Only in the past

Please describe the periods in your life in which you smoked marijuana and the amounts. Please try to summarize the most important changes in your life regarding the amount.

	From age (b)	To age (c)	Mixed with tobacco (d)	Method (e)	Unit (f)	Number per day (g)
1						
2						
3						
	If still smoking, write age at time of interview. If less than one year, write same age in (b) and (c)		(00) No (01) Yes (99) Not applicable	(00) Joints (01) Pipe (02) Bong (03) Other, specify : _____	(00) Grams (01) Joints	If less than daily, provide an average

E9. How deeply do you usually inhale when you smoke marijuana? ..... |__|__|

I don't inhale

Just into my mouth

Back into my throat

Into my lungs shallow

Into my lungs deep

Don't know (I smoke so little)

E10. Do/did you ever smoke or inhale drugs (other than marijuana)? ..... |__|__|

Never (GO TO NEXT SECTION)

Yes, I still do

Only in the past

If yes or in past, please list which drug: _____
_____

Identification number..... |__|__|-|__|__|__|

EATING HABITS & BEVERAGES

Coffee /Tea Consumption

F1. Have you ever consumed coffee or tea at least once a month for at least 1 year? ..... |__|__|

No (GO TO F3)

Yes

Don't know (GO TO F3)

F2. Please describe the periods in your life when you consumed coffee or tea.

Try and summarize the most important changes in your life regarding your consumption of coffee, and black or green tea while specifying the quantities. Please tell us about any interruptions lasting 1 year or longer.

Avoid overlapping years for the same beverage, i.e., enter 30-40, 41-45 instead of 30-40, 40-45 years old.

Unit of measurement: average cup (8 oz / 250 ml).

Exclude herbal teas

a) Coffee				c) Black tea			
From (age)	To (age)	Nb. of cups	Per day, week or month	From (age)	To (age)	Nb. of cups	Per day, week or month
			O day O week O month				O day O week O month

			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
b) Decaffeinated coffee				d) Green Tea			
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month

Consumption of beer, wine and spirits

F3. Have you ever consumed beer, wine, or spirits at least once a month for at least 1 year?

..... |__|__|

No (GO TO G1)

Yes

Don't know (GO TO G1)

F4. When do you drink? ..... |__|__|

With meals (including appetizer)

Between meals

Both

Only in social events

F5. Please describe the periods in your life when you consumed these beverages.

Try and summarize the most important changes in your life regarding your consumption of beer, wine, and spirits while specifying the quantities. Please tell us about any interruptions lasting 1 year or longer.

Avoid overlapping years for the same beverage, i.e., enter 30-40, 41-45 instead of 30-40, 40-45 years old.

Unit of measurement: bottle of beer (12 oz / 375 ml), glass of wine (4 oz / 125 ml), bottle of wine (6 glasses), spirits, i.e. hard liquor or aperitifs (1.5 oz / 45 ml).

Beer				Wine			
From (age)	To (age)	Nb. of beers	Per day, week or month	From (age)	To (age)	Nb. of glasses	Per day, week or month
			O day O week				O day O week

			O month				O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month

Spirits, hard liquors (>35%)				Aperitif (<35%)			
From (age)	To (age)	Nb. of glasses	Per day, week or month	From (age)	To (age)	Nb. of glasses	Per day, week or month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month
			O day O week O month				O day O week O month

Spirits: whisky, cognac, vodka, cachaça, grappa,… Aperitif: martini, Porto, sherry, vermouth

Identification number..... |__|__|-|__|__|__|

INTIMATE RELATIONSHIP & MARRIAGE

Now, I would like to ask you some questions about marriage and living as a couple.

G1. Have you ever been married or living in common law? ..... |__|__|

No (GO TO G3)

Yes

G2. Are you still married or living as married? ..... |__|__|

No, separated / divorced

No, widowed

Yes

The next section of the questionnaire deals with issues regarding sexual behaviour and lifestyle. As you know, our study is being conducted in many different countries and human behaviour in this regard can vary markedly from country to country. However, we have learned how important sexual behaviour is in the origin of many diseases. That is why it is considered an important topic to study.

We would be grateful if you would respond to this part of the questionnaire or as much of it as you feel comfortable doing. Remember that results of the study will be anonymised and your responses will be dealt with in strictest confidence.

Thanks again for your help.

G3. How many sexual partners (regular and casual) have you had in total in your life? ..... |__|__|

Please think of when you first became sexually active and of the periods in your life in which you may have been more sexually active or changed your sexual behaviour. For example from having more than one concurrent partner to a monogamous relationship, or from a monogamous relationship to more than one concurrent partner, or changed monogamous partners.

G4. Please try to summarize the most important changes in your life regarding your sexual partners

	From age (a)	To age (b)	Number (c)	If difficult to answer (d)
1				
2				
3				
4				
5				
6				

(d) If difficult to answer:

None

One

2-5

6-10

11-20

21-50

51-100

More than 100

Prefer not to say / Don't know

G5. How many of these were prostitute?

	From age (a)	To age (b)	Number (c)	If difficult to answer (d)
1				
2				
3				
4				
5				
6				

(d) If difficult to answer:

None

One

2-5

6-10

11-20

21-50

51-100

More than 100

Prefer not to say / Don't know

G6. How would you describe your sexual relationships? ..... |__|__|

With men only

With women only

With both men and women

G7. Have you ever performed oral sex? (your mouth on your partner's genitals) ..... |__|__|

No (GO TO NEXT SECTION)

Yes

G8. How old were you when you performed oral sex for the first time? (your mouth on your partner's genitals) ..... |__|__|

Prefer not to say / Don't know

G9. How many partners did you have oral sex with in total in your lifetime? (your mouth on your partner's genitals) ..... |__|__|

G10. How often have you had oral sex? (your mouth on your partner's genitals). Please try to summarize the most important changes in your life regarding your sexual partners

	From age (a)	To age (b)	Numbers of times (c)	Frequency (d)	In your sexual encounters, how often have you had oral sex? (e)
1				O day O week O month	
2				O day O week O month	
3				O day O week O month	
4				O day O week O month	
5				O day O week O month	
6				O day O week O month	

(e)

(00) Rarely (almost never)

(01)Occasionally (less than half the time)

(02) Often (more than half the time)

(03) Most of the time (almost every time)

(99) Prefer not to say / Don't know

Identification number..... |__|__|-|__|__|__|

HISTORY OF VARIOUS DISEASES

H1. Have you ever had skin warts? ..... |__|__|

No (GO TO H3)

Yes

Prefer not to say / Don't know (GO TO H3)

H2. If yes, where were they on your body?

	Where?
Hands	|__|__|
Feet	|__|__|
Head and Neck	|__|__|
Other, specify _____	|__|__|

(00) No

(01) Yes

(99) Prefer not to say / Don't know

H3. Since you started you sexual life, have you ever had a yeast infection (Candida Albicans)?

..... |__|__|

No (GO TO H5)

Yes

Prefer not to say / Don't know (GO TO H5)

H4. If yes, on what part of your body?

	Where?
Genitals	|__|__|
Mouth	|__|__|
Other, specify _____	|__|__|

(00) No

(01) Yes

(99) Prefer not to say / Don't know

H5. Since you started you sexual life, have you ever had herpetic lesions (cold sore)? .. |__|__|

No (GO TO H7)

Yes

Prefer not to say / Don't know (GO TO H7)

H6. If yes, on what part of your body?

	Where?
Genitals	|__|__|
Lip	|__|__|

Other, specify _____	|__|__|

(00) No

(01) Yes

(99) Prefer not to say / Don't know

H7. Have you ever had a sexually transmitted disease? ..... |__|__|

No (GO TO NEXT SECTION)

Yes

Prefer not to say / Don't know (GO TO NEXT SECTION)

H8. If yes, which one(s)?

Gonorrhea (discharge)	|__|__|
Syphilis (ulcer)	|__|__|
Condyloma (warts)	|__|__|
Chlamydia	|__|__|
HIV / AIDS	|__|__|

(00) No

(01) Yes

(99) Prefer not to say / Don't know

Identification number..... |__|__|-|__|__|__|

FAMILY HISTORY OF CANCER

Interviewer Reminder:

- Family includes these biological relatives: father, mother, brother, sister, son, daughter, aunt, uncle, grand-mother, and grand-father.

I1. How many biological brothers do you have? ..... |__|__|

I2. How many biological sisters do you have? ..... |__|__|

I3. How many biological daughters do you have? ..... |__|__|

I4. How many biological sons do you have? ..... |__|__|

I5. Has any member of your family (blood relatives) ever had cancer, including yourself? . |__|__|

No (GO TO NEXT SECTION)

Yes

Don't know (GO TO NEXT SECTION)

I6. Please describe …

	Relationship (a)	Status (b)	Current / last age (c)	Type of cancer (d)	Type of tumor (e) [LC]	Age at diagnosis (f)
1						
2						

3						
4						
5						
6						
Mother Father Sister Brother Daughter Son Grand-mother Grand-father Aunt/uncle Yourself		(00) Deceased (01) Alive	If alive, give present age. If deceased, give age at death (99) Don't know			(99) Don't know

I6. Continue as needed …

	Relationship (a)	Status (b)	Current / last age (c)	Type of cancer (d)	Type of tumor (e) [LC]	Age at diagnosis (f)
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						

Mother
Father
Sister
Brother
Daughter

(00) Deceased
(01) Alive

If alive, give present age.
If deceased, give age at death
(99) Don't know

(99) Don't know

Son

Grand-mother

Grand-father

Aunt/uncle

Identification number..... |__|__|-|__|__|__|

THANK YOU FOR HAVING AGREED TO ANSWER THIS QUESTIONNAIRE.

A9. End of interview ..... |__|__|-|__|__|

Hour Min.

A10. Quality of interview (to be established by interviewer) ..... |__|__|

Unsatisfactory

Questionable

Reliable

High quality

Comments (use additional pages if necessary)

.....

.....

.....

.....

11 APPENDIX II – Supplementary Materials for Manuscript I and II

Table S1: Prevalence of oral HPV in oral lesions among HIV-negative

	level	None	Benign/Papilloma	Precursor/Dysplastic	p
n		79	19	28	
Any alpha (%)	0	60 (75.9)	8 (42.1)	23 (82.1)	0.005
	1	19 (24.1)	11 (57.9)	5 (17.9)	
alpha3 (%)	0	72 (91.1)	16 (84.2)	25 (89.3)	0.670
	1	7 (8.9)	3 (15.8)	3 (10.7)	
Highrisk hpv (%)	0	74 (93.7)	16 (84.2)	25 (89.3)	0.387
	1	5 (6.3)	3 (15.8)	3 (10.7)	
hpv16 (%)	0	76 (96.2)	17 (89.5)	26 (92.9)	0.474
	1	3 (3.8)	2 (10.5)	2 (7.1)	
Any beta (%)	0	31 (39.2)	7 (36.8)	9 (32.1)	0.800
	1	48 (60.8)	12 (63.2)	19 (67.9)	
beta1 (%)	0	49 (62.0)	12 (63.2)	11 (39.3)	0.096
	1	30 (38.0)	7 (36.8)	17 (60.7)	
beta2 (%)	0	47 (59.5)	11 (57.9)	19 (67.9)	0.703
	1	32 (40.5)	8 (42.1)	9 (32.1)	
beta3 (%)	0	69 (87.3)	18 (94.7)	25 (89.3)	0.653
	1	10 (12.7)	1 (5.3)	3 (10.7)	
hpv5 (%)	0	68 (86.1)	16 (84.2)	21 (75.0)	0.399

	1	11 (13.9)	3 (15.8)	7 (25.0)	
hpv8 (%)	0	76 (96.2)	17 (89.5)	24 (85.7)	0.148
	1	3 (3.8)	2 (10.5)	4 (14.3)	
Any gamma (%)	0	66 (83.5)	12 (63.2)	23 (82.1)	0.129
	1	13 (16.5)	7 (36.8)	5 (17.9)	
gamma8 (%)	0	75 (94.9)	16 (84.2)	27 (96.4)	0.180
	1	4 (5.1)	3 (15.8)	1 (3.6)	
gamma10 (%)	0	75 (94.9)	16 (84.2)	27 (96.4)	0.180
	1	4 (5.1)	3 (15.8)	1 (3.6)	

Level: 0 = Negative, 1=Positive

Table S2: Prevalence of oral HPV in oral lesions among PLWH

	level	None	Benign/Papilloma	Precursor/Dysplastic	p
n		132	18	19	
Any alpha (%)	0	80 (60.6)	9 (50.0)	14 (73.7)	0.332
	1	52 (39.4)	9 (50.0)	5 (26.3)	
alpha3 (%)	0	116 (87.9)	15 (83.3)	18 (94.7)	0.549
	1	16 (12.1)	3 (16.7)	1 (5.3)	
Highrisk hpv (%)	0	103 (78.0)	13 (72.2)	18 (94.7)	0.179
	1	29 (22.0)	5 (27.8)	1 (5.3)	
hpv16 (%)	0	116 (87.9)	15 (83.3)	19 (100.0)	0.219
	1	16 (12.1)	3 (16.7)	0 (0.0)	
Any beta (%)	0	38 (28.8)	4 (22.2)	4 (21.1)	0.685
	1	94 (71.2)	14 (77.8)	15 (78.9)	
beta1 (%)	0	49 (37.1)	6 (33.3)	7 (36.8)	0.952
	1	83 (62.9)	12 (66.7)	12 (63.2)	
beta2 (%)	0	79 (59.8)	11 (61.1)	11 (57.9)	0.979
	1	53 (40.2)	7 (38.9)	8 (42.1)	
beta3 (%)	0	116 (87.9)	17 (94.4)	17 (89.5)	0.706
	1	16 (12.1)	1 (5.6)	2 (10.5)	

hpv5 (%)	0	97 (73.5)	15 (83.3)	15 (78.9)	0.610
	1	35 (26.5)	3 (16.7)	4 (21.1)	
hpv8 (%)	0	112 (84.8)	15 (83.3)	19 (100.0)	0.182
	1	20 (15.2)	3 (16.7)	0 (0.0)	
Any gamma (%)	0	91 (68.9)	11 (61.1)	11 (57.9)	0.545
	1	41 (31.1)	7 (38.9)	8 (42.1)	
gamma8 (%)	0	124 (93.9)	17 (94.4)	18 (94.7)	0.988
	1	8 (6.1)	1 (5.6)	1 (5.3)	
gamma10 (%)	0	122 (92.4)	18 (100.0)	18 (94.7)	0.461
	1	10 (7.6)	0 (0.0)	1 (5.3)	

To illustrate using the package with RRs, we consider data from a cohort study examined in Zou (164). The study evaluated the effect of age and BMI on hypertension (193).

```
d1 = data.frame(age = rep(1,1021),
                bmi = rep(1, 1021),
                hypertension = rep(c(1,0), c(278,743)))

d2 = data.frame(age = rep(1,681),
                  bmi = rep(0, 681),
                  hypertension = rep(c(1,0), c(100,581)))

d3 = data.frame(age = rep(0,1385),
                  bmi = rep(1, 1385),
                  hypertension = rep(c(1,0), c(153,1232)))

d4 = data.frame(age = rep(0,1810),
                  bmi = rep(0, 1810),
                  hypertension = rep(c(1,0), c(79,1731)))
hypdata = rbind(d1,d2,d3,d4) #study data
```

Next, we fit a log-binomial regression with interaction term for age and BMI, and hypertension as the outcome.

```
RRmodel = glm(hypertension ~ age*bmi, family = binomial(link = "log"), data = hypdata)
```

Then, we call the *interactionR()* function and generate the table:

```
library(interactionR)
v = interactionR(RRmodel, exposure_names = c("age", "bmi"), ci.type = "mover", em = F)
interactionR_table(v) #returns a publication-ready table.
```

Do you want to save a Microsoft Word copy of the em/interaction table to your working director y? (yes/No/cancel)

Interaction of age and bmi

	bmi absent	bmi present	Effect of bmi within the strata of age
	RR [95% CI]	RR [95% CI]	RR [95% CI]
age absent	1 [Reference]	2.53 [1.95, 3.29]	2.53 [1.95, 3.29]
age present	3.36 [2.54, 4.46]	6.24 [4.92, 7.91]	1.85 [1.51, 2.28]
Effect of age within the strata of bmi	3.36 [2.54, 4.46]	2.46 [2.06, 2.95]	

```
Multiplicative     0.73 [0.52,
scale                1.02]
RERI                 1.34 [0.31,
                   2.37]
AP                   0.22 [0.06,
                   0.35]
SI                   1.34 [1.07,
                   1.69]
------------------------------------------------------------------------------

Column names: c1, c2, c3, c4
```

Notice that the tabling function recognizes the model as RR, and updates the effect measure accordingly (instead of outputting "OR"). The returned point estimates and CI for RERI and AP is as reported in Zou (Figure 5) for this data (164).