Exploration and Characterization of the Microbiome in Oral Squamous Cell Carcinoma using 16s rRNA Gene Next-generation Sequencing

Shatha Abdulmohsen M AlNafea

University College London

Doctor of Philosophy

Declaration

I, Shatha AlNafea, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

_____

Signed: Shatha AlNafea

Abstract

Oral squamous cell carcinoma (OSCC) is an oral cancer with high prevalence rates in many countries, especially ones with high rates of risk factors, such as tobacco use. As OSCC is typically diagnosed in later stages and is responsible for high mortality rates, identifying biomarkers associated with OSCC could be instrumental in improving OSCC diagnosis, prognosis, and treatment. A comprehensive literature review on the use of 16s rRNA gene next generation sequencing (NGS) in studying OSCC biomarkers found that few studies to date have used rigorous scientific or biostatistical approaches, and that the high-quality studies identified have inconsistent findings. This dissertation investigated the oral microbiome of OSCC patients along with patients with a condition known to be a precursor to OSCC called oral epithelial dysplasia (OED) in the United Kingdom (UK), London. First, in substudy 1, which is a cross-sectional study, the oral microbiome of OSCC lesion tissue was compared to OED lesion tissue in terms of diversity and relative differential abundance. Next, in substudy 2, which is a case series study, the same metrics were used to compare OSCC and OED lesion tissue with healthy tissue in the same patient. Finally, in substudy 3, which is also a case series study, data obtained from a Chinese study of similar design to substudy 2 was reanalysed using the same bioinformatics and statistical approaches as used in substudy 2, and the results compared. The overall results of all three studies showed evidence that OSCC lesion tissues have statistically significantly more diversity than OED or healthy tissue. However, no specific members of

the oral microbiome could be identified that appeared to be responsible for OSCC in any of the substudy analyses. Future research must tackle the challenge of using rigorous scientific methods that are replicable to identify OSCC biomarkers that have clinical utility.

Impact Statement

There are three main areas where the results of the research from this dissertation could have a positive impact in terms of public benefit. First, a comprehensive review of the literature was done on the use of 16s rRNA gene next generation sequencing (NGS) on biomarker research in oral squamous cell carcinoma (OSCC). The results showed that – to date – no members of the oral microbiome have been reliably implicated in the aetiology of OSCC. In addition, this review revealed two important findings.

The first finding is that much of the research being done currently on the oral microbiome using 16s rRNA gene NGS for OSCC biomarker identification is not adhering to rigorous scientific standards. Second, the review revealed that there exists much guidance available in the scientific literature to improve the rigor of these studies, but this guidance is cross-disciplinary, and elaborated upon in a diversity of scientific journals. Raising the level of awareness of these two issues in the study of OSCC biomarkers is important, because it is necessary to conserve the public's resources devoted to cancer research. In order to benefit the public the most, scientific research should be designed optimally so as to produce clear and replicable results.

Secondly, in addition to the findings from the comprehensive review herein, this dissertation provides an example of methods that can be used in OSCC biomarker research that are both rigorous and replicable. Substudies 1 and 2 provide an example of OSCC microbiome study designs and biostatistical analyses that are scientifically valid,

although they possess features that could improved (such as expanding the sample size). Substudy 3 demonstrates a procedure for reanalysing publicly available OSCC microbiome data from a previously published study – that used similar design but different analysis method. By standardising the analysis protocol, a direct comparison of the results of studies of similar designs can be obtained. These examples also highlight the importance of maintaining high-quality online repositories of oral microbiome data, so scientists can improve their methods by learning from each other.

Thirdly, if scientists studying OSCC biomarkers are able to conduct rigorous research that identifies reliable OSCC biomarkers, then the public may benefit by having expanded diagnostic and treatment options. First, OSCC microbiome biomarkers that signal OSCC can be used as a less invasive method to diagnose OSCC as compared to incisional biopsy – the gold standard diagnostic technique. Next, Microbiome biomarkers in OSCC can be used to guide treatment decisions. In an ideal situation, microbiome biomarkers, once established to induce or promote OSCC, can be targeted directly to enhance both treatment and prognosis of the disease.

Gaining a better understanding of the oral microbiome in OSCC will lead to improved clinical care and outcomes. Raising the awareness for the need for collaborative multi-disciplinary efforts to build evidence - based on rigorous studies, can contribute to the positive impact of this research on public good.

Acknowledgements

The existence of this research thesis would not have been possible without the help and support of many individuals, to whom I would like to express my thankfulness.

First, I will start by extending my sincerest gratitude to my supervisors, Dr Sean Nair and Professor Stefano Fedele, for their invaluable guidance, full support, and consistent encouragement throughout my learning journey and research experience. Their insightful observations and constructive feedback have been instrumental in shaping this thesis. I would also like to express my deepest gratitude to Professor Stephen Porter for giving generously of his time, knowledge, and great expertise. His authentic support has made him a most influential mentor who added measurably to my professional and personal growth. In addition, I would like to thank Ms Clare Schilling for her unending support and helping me by facilitating recruitment and sample collection during my research.

My thanks also go to the following technicians in the microbiology lab: Dr Anna Tymon, Dr Haitham Hussain, Tracey Moss, and Ingrid Green. I also would like to thank Dr Mehmet Davrandi. I greatly value the assistance they gave me in laboratory procedures and technical support as I did my research.

I would also like to thank the Eastman Clinical Investigation Centre team. I specifically would like to express my appreciation to Victoria Hoskins, Rachel Knight, Beata Szczepkowska and Kasia Niziolek for their generous help throughout the research process. I

additionally would like to thank my colleagues in both the Microbial Diseases and Oral Medicine Departments for making this difficult educational journey easier for me. Special thanks go to Arley, Khadijah, Sophia, Paswach, and Felix.

Closer to my home, my appreciation also goes to King Saud University for their financial support, which enabled me to pursue this education and research. I also want to thank my friends Dr Roaa Alsulaiman, Dr Kunanya Pimolbutr, and Dr Shatha Al-Harthi for providing full support abroad and locally as I pursued my dream. You added some necessary inspiration, laughter, and motivation during my journey.

Finally, I will turn to my family, on whom I rely as my support system, and who really deserves the credit for supporting me during this period of my life. I especially want to thank my parents, Hessah and Abdulmohsen, for being the rock that held our family together while I was busy studying. They opened their loving home to the wonderful gift of my son, Salman, who came along during my studies, while my older daughters could be with me. Fahda and Amal, you are my heart and soul for accompanying me throughout this period of my education, always supporting me, helping me, and trying to cheer me up with your laughter. You took care of me as I had taken care of you. I'm also eternally grateful for my husband, Badr. You were so supportive and understanding with all the unforeseen events, and I cannot tell you enough how much that means to me.

Table of Contents

Abbreviations

Abbreviation	Definition
AF	Attributable fraction
ALCHEMIST	Adjuvant Lung Cancer Enrichment Marker Identification and Sequencing Trial)
ANOVA	Analysis of variance
AUC	Area under the curve
ASA	American Statistical Association
BMI	Body mass index
CI	Confidence interval
CTC	Circulating tumour cell
DGGE	Denaturing gradient gel electrophoresis
DHHS	Department of Health and Human Services
DNA	Deoxyribonucleic acid
EBM	Evidence-based medicine
EBV	Epstein-Barr virus
ECM	Extracellular matrix
EDH	Eastman Dental Hospital
EPIC	European Prospective Investigation into Cancer and Nutrition study
EV	Extracellular vesicle
FEP	Fibro-epithelial polyps
GCO	Global Cancer Observatory
GP	General practitioner
GS	Google Scholar
GSCC	Gingival squamous cell carcinoma

Abbreviation	Definition
HDI	Human Development Index
HHV-4	Human herpes virus 4
HMP	Human Microbiome Project
HNC	Head and neck cancer
HOMD	Human Oral Microbiome Database
HPV	Human papillomavirus
HSV	Herpes simplex virus
IARC	International Agency for Research on Cancer
IQR	Interquartile range
IRAS	Integrated Research Application System
LPS	Lipopolysaccharide
ML	Machine learning
MPE	Molecular pathological epidemiology
MSA	Multiple sequence alignment
NGS	Next generation sequencing
NHANES	National Health and Nutrition Examination Survey
NIH	National Institutes of Health
NSCLC	Non-small-cell lung cancer
OED	Oral epithelial dysplasia
OPMD	Oral Potentially Malignant Disorders
OR	Odds ratio
OSCC	Oral squamous cell carcinoma
OSF	Oral submucous fibrosis
OUT	Operational Taxonomic Unit
PAH	Polycyclic aromatic hydrocarbons
PCR	Polymerase chain reaction

Abbreviation	Definition
PERMANOVA	Permutational multivariate analysis of variance
RCT	Randomized controlled trial
RNA	Ribonucleic acid
ROC	Receiver-operator curve
SCC	Squamous cell carcinoma
SES	Socio-economic status
SRA	Sequence Read Archive
TIM	Tumor immune microenvironment
TSNA	Tobacco-specific nitrosamines
UCL	University College London
UCLH	University College of London Hospital
UF	UniFrac
UK	United Kingdom
US	United States

List of Tables

List of Figures

Chapter 1: Introduction

A dissertation is presented as part of doctoral degree program at University College London (UCL). It begins with this chapter, introducing the problem.

Globally, rates of head and neck cancer (HNC) continue to rise, and oral cancer of particular concern is oral squamous cell carcinoma (OSCC) (Gormley et al., 2022; Salehiniya & Raei, 2020). OSCC is prevalent globally, but in countries with higher rates of risk factors, such as tobacco smoking and chewing of the betel nut, the rates are even higher (Sung et al., 2021). Regions in the United Kingdom (UK) have been identified as having high rates of OSCC; between 2010 and 2015, the OSCC death rate in the UK went from 2.92 to 3.26 per 100,000 in men (for an increase of 11.6%), and from 1.13 to 1.21 in women (for an increase of 7.1%) (Bosetti et al., 2020). OSCC is typically diagnosed in later stages, and is associated with high mortality rates and low five-year survival rates (Huber & Tantiwongkosi, 2014; Salehiniya & Raei, 2020).

Epigenetic studies have provided candidate biomarkers of OSCC that can be targeted for further study, especially with regard to association of tobacco with cancer induction (Ambatipudi et al., 2016; de la Iglesia et al., 2020). Biomarkers are biologic molecules, proteins, or metabolites derived from the analysis of body fluids or tissues for diagnostic purposes (Hu & Dignam, 2019). Although studies of biomarkers in OSCC began as early as 1998, more recently, the invention of 16s rRNA gene next-generation sequencing (NGS) has facilitated a vast expansion in the ability to identify members of the oral

microbiome (Hooper et al., 2006, 2007; Nagy et al., 1998; Pushalkar et al., 2011).

Although 16s rRNA gene NGS provided a greater capacity for identifying members of the oral microbiome that may serve as biomarker targets in OSCC, to date, little foundation has been added to this potential evidence base, largely due to issues with study design and biostatistics (Goossens et al., 2015; Hu & Dignam, 2019; Kers & Saccenti, 2022; Ou et al., 2021). Although biomarker measurements such as α diversity, β diversity, and relative differential abundance have been operationalized, and study design and biostatistical guidance is available, it is generally not heeded in the studies published (Gloor et al., 2017; Goossens et al., 2015; Hu & Dignam, 2019; Kers & Saccenti, 2022; Ou et al., 2021). Hence, the general aim of this dissertation project, was to explore the oral microbiome of OSCC using 16s rRNA gene NGS sequencing.

Rationale

This dissertation seeks to add to the growing literature investigating the oral microbiome to support the identification of biomarkers of OSCC that can be useful for prevention, diagnosis, and/or treatment (Hu & Dignam, 2019). Existing literature is largely exploratory, and is not intended for clinical application (Gloor et al., 2017; Goossens et al., 2015; Hu & Dignam, 2019; Kers & Saccenti, 2022; Ou et al., 2021). The substudies conducted in this dissertation are aimed at applying evidence-based methods to studying the oral microbiome that could potentially lead to clinical applications (Hu & Dignam, 2019).

The following section will summarize the structure of this dissertation project.

Literature Review

Chapter 2 provides a background about OSCC followed by a comprehensive literature review with the following objectives: 1) To identify applicable scientific evidence from studies investigating high-throughput, NGS targeting the 16s rRNA gene in studying the oral bacteriome of OSCC, and 2) to characterize findings related to the oral bacteriome in OSCC from these studies.

Substudy 1: Diversity and Differential Relative Abundance in the Bacteriome of Oral Squamous Cell Carcinoma and Oral Epithelial Dysplasia Lesions: Results from the United Kingdom

Chapter 3 presents substudy 1, which was aimed to compare the oral microbiome in swapped tissue from OSCC and OED lesions in a small, heterogeneous sample of patients in the UK diagnosed with either condition using 16S rRNA NGS sequencing.

Substudy 2: Study to Characterize How the Oral Microbiome Differs in OSCC and OED from Healthy Tissue in the Same Individual

Chapter 4 presents substudy 2, which sought to build on the findings from substudy 1. When the swabbed tissue samples were obtained from the OSCC and OED lesions of UK participants in substudy 1, a swab sample of anatomically-matched non-lesional healthy tissue was also collected for 16s rRNA gene sequencing. Whether similar results would be seen when comparing lesional with healthy tissue in the same individual was the basis of inquiry for substudy 2.

Substudy 3: A Reanalysis of Oral Microbiome Repository Data Shows Consistent Results with Current Study Data

Chapter 5 presents substudy 3, which is a reanalysis of publicly available OSCC microbiome sequence data of a previously published study (National Institutes of Health, n.d.; Zhang et al., 2020). The original study by Zhang et al. (2020) from China, had a similar study design (contributing both lesional and healthy tissue samples) as substudy 2 but was analysed differently. Hence, the aim of the reanalysis was to allow a direct comparison between the results generated in this chapter to that of substudy 2.

Discussion and Conclusion

Chapter 6 provides a final discussion and conclusion to the dissertation. It synthesizes the results from the three substudies, and reflects on their consistency with prior studies described in Chapter 2. This section also describes the strengths and limitations of the three substudies, and recommends future research directions.

Chapter 2: Literature Review

The ability to perform next-generation sequencing (NGS) targeting the 16s rRNA gene in the study of the oral bacteriome in squamous cell carcinoma (OSCC) became technologically possible within the past decade due to both scientific and technological advances (T. Chen et al., 2010; Pushalkar et al., 2011). My PhD project intended to study the OSCC bacteriome (bacterial members of the microbiome) using 16s rRNA gene NGS sequencing technology. Prior to designing novel studies using 16s rRNA gene sequencing for the bacteriome in OSCC, it is necessary to undertake a comprehensive review of the scientific literature arising from this innovation. It is important to recognise that exploratory or hypothesis-generating research is important, and that not every study needs to involve an *a priori hypothesis.*

This chapter begins with a background section that is separated into two parts. In the first part, a background about OSCC is presented, including epidemiology, risk factors, and study design considerations behind cancer biomarker studies. In the second part, types of literature reviews are described. A comprehensive review is undertaken in this thesis that includes the background, the objectives of the review and the research question to be answered in the review. A section providing an overview of the methodologic approach selected and used for the comprehensive review, including details of article search and selection, and how articles were classified, and their results synthesized is included. Finally, the results of the comprehensive review are presented,

with a discussion synthesizing the findings from the articles included in the review.

2.1. Background

In this doctoral project, three substudies were conducted using 16s rRNA gene NGS of the oral bacteriome in OSCC (presented in Chapters 3, 4 and 5). 16s rRNA gene NGS represents a method of measuring the oral bacteriome with greater clarity than was available previously (Pushalkar et al., 2011). As 16s rRNA gene NGS is a decades-old approach, in order to design and execute the most scientifically rigorous sub studies on the topic, it is necessary to first analyse the body of peer-reviewed scientific literature in the area. However, in order to obtain the information necessary for optimal study design and execution, the correct type of review must be selected. Therefore, this section begins by reviewing the literature on OSCC, and then proceeds to review the literature on how to do a comprehensive review to ensure standard recommendations are met.

<u>OSCC Background</u>

Why Head and Neck Cancer Must Be Described When Studying OSCC

In order to understand the current status of the scientific literature that centring around using 16s rRNA gene sequencing to study the oral bacteriome in OSCC, it is first necessary to understand the epidemiology of OSCC. OSCC is epidemiologically classified as being a member of the category “squamous cell carcinoma (SCC) of the head and neck” (referred to as “head and neck cancer”, abbreviated HNC), which is a

non-homogenous group of cancers (Salehiniya & Raei, 2020). The anatomical locations of cancers in the HNC category include the oral cavity, tongue and lip, as well as nasal and other structures (Salehiniya & Raei, 2020).

Although there is an attempt in this literature to focus squarely on OSCC, this focus cannot be so narrow. This is because many studies do not make clear distinctions between whether they are studying HNC, OSCC, or a group of HNCs (Gormley et al., 2022; Salehiniya & Raei, 2020). Oral cancer is a type of HNC, and OSCC is a type of oral cancer (Gormley et al., 2022; Salehiniya & Raei, 2020). Studies focusing on clinical topics acknowledge that these cancers may have similar risk factors, and therefore, they are often discussed together (Gormley et al., 2022; Salehiniya & Raei, 2020).

Global Epidemiology of OSCC

Each of the HNCs has been found to have different risk factor profiles and prognoses (Salehiniya & Raei, 2020). It is estimated that each year, globally, there will be more than 450,000 new HNC cases and 350,000 HNC deaths will occur (Salehiniya & Raei, 2020). The oral cavity is the most common location for cancer in the head and neck region, although lip cancer is particularly challenging, as it is typically diagnosed at a malignant stage, and accounts for between 25 and 30% of all mouth cancers (Huber & Tantiwongkosi, 2014; Salehiniya & Raei, 2020). HNC is a significant cause of mortality and morbidity worldwide, and represents a serious global public health issue (Gormley et al., 2022; Salehiniya & Raei, 2020).

Because of the lack of homogeneity within HNCs as diseases, the genetic diversity of underlying worldwide populations, and the differential exposures experienced by populations worldwide, it is not possible to generalize risk factors and prognoses for HNCs across global populations (Gormley et al., 2022; Salehiniya & Raei, 2020). Gormley and colleagues (2022) recently summarized global risk factors for HNCs, which are tobacco smoking, tobacco used in combination with alcohol, and low socio-economic status (SES), which confers added risk unaccounted for by tobacco or alcohol use. The authors also pointed out that human papillomavirus (HPV) is a major risk factor for oropharyngeal cancer (Gormley et al., 2022).

Sung and colleagues (Sung et al., 2021) analysed data from the Global Cancer Observatory (GCO) to compare global rates of 36 groups of cancers, including cancers of the lip and oral cavity. Among those 36 cancer groups, lip and oral cavity cancer ranked 18th in terms of new cases in 2020 (377,713, 2% of cancers analysed), and were responsible for 177,757 deaths in 2020 (1.8% of cancers analysed) (Sung et al., 2021). Among men in 2020, of the 36 cancer groups analysed, lip and oral cavity cancers had the highest incidence in four countries, including India and Pakistan (Sung et al., 2021). In this analysis, countries were classified into a four-tier Human Development Index (HDI) (Sung et al., 2021). Among the low HDI countries, oral and lip cancer ranked as the third highest incidence (which was 10.2 per 100,000 in 2020) - largely due to India's influence (Sung et al., 2021).

Ferlay et al. (2021) conducted their own analysis of the same GCO data. In their analysis, they developed estimates of aged standardized rates per 100,000 of 38 different cancer groups (Ferlay et al., 2021). Like the previous authors, these authors grouped lip and oral cavity cancer together in one category (Ferlay et al., 2021; Sung et al., 2021). They estimated that among both sexes, the number of new cases in both sexes would be 377.7 (95% confidence interval [CI] 362.4 to 393.7) per 100,000, with the male-specific estimate being 264.2 (95% CI 251.2 to 277.9) per 100,000, and the female-specific estimate being much lower, at 113.5 (95% CI 105.6 to 122.0) per 100,000 (Ferlay et al., 2021). This analysis was consistent with the previous one in finding a high incidence of lip and oral cavity cancer among males in south-central Asia (Ferlay et al., 2021; Sung et al., 2021). In another analysis of the 2018 GCO data focusing on cancers of the lip, tongue, and mouth, researchers found that the highest rates of lip cancer were in Australia, the highest rates of mouth and oral tongue cancer incidence were in India (Miranda-Filho & Bray, 2020).

United Kingdom Epidemiology of Oral and Pharyngeal Cancer

In a large study of global trends in the incidence and mortality of oral and pharyngeal cancer, Bosetti and colleagues (2020) calculated estimates for several countries and regions, including the entire United Kingdom (UK). For comparison, they also calculated these rates in three different ways: 1) England, and Wales (compared to the rest of the UK); 2) for Northern Ireland (compared to the rest of the UK), and 3) for Scotland (compared to the rest of the UK) (Bosetti et al., 2020). For the

UK overall, they estimated the death rate from oral and pharyngeal cancer in men in 2010 to be 2.92 per 100,000, which went up to 3.26 per 100,000 in 2015, with a percent change of 11% (Bosetti et al., 2020). For women in the UK, the overall death rate was 1.13 per 100,000 in 2010, and 1.21 per 100,000 in 2015, a percentage change of 7.1% (Bosetti et al., 2020). Regionally, for England and Wales (compared to the rest of the UK), the male estimate was lower, but the percent change higher (2010: 2.79 per 100,000, 2015: 3.13 per 100,000, percent change 12.2%), which was a pattern also seen in females (2010: 1.07 per 100,000, 2015: 1.16 per 100,000, percent change 8.4%) (Bosetti et al., 2020). The trend was when comparing Northern Ireland to the rest of the UK, both estimates increased to 5.31 and 1.50 per 100,000 in men and women respectively in 2015, and when comparing Scotland with the rest of the UK, both estimates were the highest, with 3.97 and 1.59 per 100,000 in 2015 for men and women respectively (Bosetti et al., 2020).

Aetiology of Head and Neck Cancer

Noting the risk for HNC was 10 times higher in smokers than others, Jethwa and Khariwala (2017) speculated on tobacco-related carcinogenesis of HNC. They provided a history of tobacco promotion and use globally starting in the early 1900s, and summarized the carcinogenic compounds identified in tobacco since the US Surgeon General's report in 1964, which provided a new, expansive effort to understand the carcinogenic pathways initiated by tobacco use (Jethwa & Khariwala, 2017). In Table 3 of their paper, they listed the components that have been evaluated in either laboratory animals or humans by the

International Agency for Research on Cancer (IARC), which included polycyclic aromatic hydrocarbons (n=19), nitrosamines (n=8), aromatic amines (n=13), aldehydes (n=2), and various hydrocarbons (n=6) (Jethwa & Khariwala, 2017). They found a total of 70 tobacco-related compounds that are carcinogenic, including tobacco-specific nitrosamines (TSNA) and polycyclic aromatic hydrocarbons (PAH), which have been the most intensely studied to date (Jethwa & Khariwala, 2017). These findings are consistent with a study of 47 histologically proven OSCC cases, where researchers found a high level of mutations and a heterogenous mutational spectrum among OSCC, and multiple mutational events (Batta & Pandey, 2019). However, without showing a causal link between exposure to tobacco and OSCC cases, a causal association cannot be claimed.

Bugshan and Farooq (2020) reviewed the state of the literature with respect to OSCC on the topics of metastasis, potentially associated malignant disorders, aetiology, and recent diagnostic advancements. In their review, they found that OSCC can be responsible for both regional and distant metastases, and other malignant diseases can transform into OSCC depending upon the aetiologic factors present (Bugshan & Farooq, 2020). Ernani and Saba (2015) noted in their paper on oral cavity cancer risk factors, pathology and management that the possibility of a second primary malignancy should be considered in diagnosing all HNCs, but especially OSCC. Generally, the prognoses for HNCs are not good, and the goals of treatment are to improve survival and organ function (Ernani & Saba, 2015). Treatments include a combination of

surgery, radiation therapy, and chemotherapy, and patients are supervised by a multidisciplinary team (Ernani & Saba, 2015). Currently, there is no straightforward aetiologic pathway discerned for OSCC to develop, although as early as 2013, there were calls to standardize the case definition of oral cavity cancer to support such an investigation (Radoï & Luce, 2013).

Specific Risk Factors for Head and Neck Cancers

In their review, Gormley and colleagues (2022) included a map indicating the relative age-standardized incidence rates of HNC globally. According to the map, the highest rates were seen in Australia, India and surrounding countries in Southeast Asia, and the across all countries in northern Europe (Gormley et al., 2022). This reflects diversity in both genetic and environmental exposures to risk factors for HNCs globally (Gormley et al., 2022; Salehiniya & Raei, 2020; Toporcov et al., 2015). This section will summarize proposed causes for widely varying HNC rates country to country by examining evidence behind risk factors for HNCs.

Table 2.1 presents a very high-level summary of the known risk factors for HNCs (Gormley et al., 2022; Salehiniya & Raei, 2020).

Table 2.1 Summary of known risk factors for head and neck cancer.

Risk Factor	Oral Cavity Cancer	Lip Cancer	Other Head and Neck
Tobacco	Smoked tobacco has a positive relationship impacted mostly by strength of tobacco. Both smoke-free tobacco and tobacco smoke in environment are shown to be risk factors. Environmental risk factor is reversable after 15 years of clean environment (Gormley et al., 2022; Salehiniya & Raei, 2020).	Smoked tobacco increases principal risk factor for both lips, and occurs in the region where the smoking implement is placed (Gormley et al., 2022; Salehiniya & Raei, 2020).	Tobacco smoking confers a greater risk for laryngeal cancer (Gormley et al., 2022; Salehiniya & Raei, 2020).
Paan (betel quid chewing)	In Asia, *paan* with or without tobacco increases oral cavity cancer risk between 1.5 and 3 times that of tobacco use (Gormley et al., 2022; Salehiniya & Raei, 2020).		
Human papillomavirus (HPV) and other viral infections	HPV is a principal cause of squamous cell carcinoma of the head and neck, making it important to prevent sexual	HPV and herpes simplex type I are risk factors (Gormley et al., 2022; Salehiniya & Raei, 2020).	HPV infection is thought to be an increasingly important risk factor for head and

Risk Factor	Oral Cavity Cancer	Lip Cancer	Other Head and Neck
	transmission (Gormley et al., 2022; Salehiniya & Raei, 2020).		neck cancer (Gormley et al., 2022; Salehiniya & Raei, 2020).
Family history of cancer	Family history studies suggests that genetic risk factors are amplified by exposure to tobacco, alcohol, and diet, suggesting that genetic instability is a risk factor for oral cell carcinoma (Gormley et al., 2022; Salehiniya & Raei, 2020).	Those with family history of lip cancer have a higher risk of developing it compared to the background population, and this risk is increased through environmental exposures (Gormley et al., 2022; Salehiniya & Raei, 2020).	
Social inequity	Globally, head and neck cancers are more prevalent among low socio-economic statuses (Gormley et al., 2022; Salehiniya & Raei, 2020).		
Alcohol consumption	Higher consumption patterns linked to risk, but unclear if this is confounded by tobacco use (Gormley et al., 2022; Salehiniya & Raei, 2020).		Greater risk for oral cavity and oropharyn-geal cancer (Gormley et al., 2022; Salehiniya & Raei, 2020)..

Risk Factor	Oral Cavity Cancer	Lip Cancer	Other Head and Neck
Marijuana smoke	Increases risk for oral and lip cancer, but may be confounded by association with tobacco and alcohol use (Gormley et al., 2022; Salehiniya & Raei, 2020).		
Oral mucus disease	Potentially malignant oral disorders such as leukoplakia may increase the risk of SCC (Salehiniya & Raei, 2020).		
Occupational exposure	Increased risk due to occupational exposures (such as to perchloroe-thylene) has been demonstrated (Salehiniya & Raei, 2020).	Increased risk due to occupational exposures (such as to sunlight, agriculture, and greenhouse work) has been demonstrated (Salehiniya & Raei, 2020).	
Exposure to sunlight		Exposure to sunlight (from living or working outside) increases the incidence of lip cancer (Salehiniya & Raei, 2020). Those living in rural areas are at higher risk likely due to increased sunlight exposure	

Risk Factor	Oral Cavity Cancer	Lip Cancer	Other Head and Neck
		(Gormley et al., 2022).	
Light skin colour		Almost half of all lip cancers occur in Northern European countries, and lip cancer is almost negligible among those with darker skin due to natural melanin in skin pigmentation that is protective (Salehiniya & Raei, 2020).	
Immunos-uppression and immune-deficiency		Studies in the transplant population show that the risk of lip cancer increases the longer the patient is immunosuppressed (Salehiniya & Raei, 2020).	
Diet	Low vegetable consumption, high meat consumption, and hot tea increase risk (Salehiniya & Raei, 2020). Diet considered a minor risk factor for head and neck		Diet considered a minor risk factor for head and neck cancers (Gormley et al., 2022).

Risk Factor	Oral Cavity Cancer	Lip Cancer	Other Head and Neck
	cancers (Gormley et al., 2022).		
Body mass index (BMI) and physical activity status	Studies suggest weight gain is a protective factor and weight loss is a risk factor (Salehiniya & Raei, 2020). Physical activity status considered a minor risk factor for head and neck cancers (Gormley et al., 2022).		Physical activity status considered a minor risk factor for head and neck cancers (Gormley et al., 2022).
Oral and dental health	Lack of oral health behaviours is a risk factor. Some studies have shown mouthwash as a risk factor for malignancy but this remains debated (Salehiniya & Raei, 2020). Oral hygiene considered a minor risk factor for head and neck cancers (Gormley et al., 2022).		Oral hygiene considered a minor risk factor for head and neck cancers (Gormley et al., 2022)..

As shown in Table 2.1, although there is some debate as to the relative importance of each of the risk factors to different specific HNCs,

a particularly dominating causal exposure is tobacco (Gormley et al., 2022; Huber & Tantiwongkosi, 2014; Salehiniya & Raei, 2020). Tobacco can play the role of risk factor for HNC through the patient smoking it (in various ways) or chewing it, or from being environmentally exposed to tobacco smoke occupationally or within the residence (Gormley et al., 2022; Salehiniya & Raei, 2020).

Though much evidence exists to show that tobacco is a strong risk factor for HNCs, the Asian behaviour of *paan* (betel quid chewing) has been found to increase the risk of oral cavity cancer between one-and-a-half and three times the risk seen with tobacco (Hernandez et al., 2017; Huber & Tantiwongkosi, 2014; Salehiniya & Raei, 2020). Betel nut comes from the *areca catechu* palm tree, and betel chewing is practiced by between 10% and 20% of the world's population, mainly concentrated in South and Southeastern Asia and the Pacific (Hernandez et al., 2017). Although chewing betel nut has been linked to oral carcinogenesis, the underlying mechanisms are not well understood (Hernandez et al., 2017). A review of betel chewing without tobacco found that betel chewing is an independent risk factor for cancer from tobacco (Liu et al., 2015). In China, public health officials face unique challenges, because betel chewing, tobacco smoking (in the form of cigarette smoking), and alcohol consumption is a common triple exposure (Liu et al., 2015). One group conducted a case-control study of OSCC cases compared to controls with oral submucous fibrosis (OSF), and found that there was a positive dose-response association between OSCC case status and

odds of betel use, and the odds were higher among those who also used cigarettes (Liu et al., 2015).

As described in Table 2.1, attention has been paid more recently to viral origins of HNC (Gormley et al., 2022; Huber & Tantiwongkosi, 2014; Jiang et al., 2019; Salehiniya & Raei, 2020). HPV has been implicated in OSCC, so public health measures to prevent sexual transmission of HPV are recommended as a preventive strategy (Marur & Forastiere, 2016; Salehiniya & Raei, 2020). Additionally, there is evidence that infection with herpes simplex virus (HSV) type I is a risk factor for lip cancer (Gormley et al., 2022). Interaction of tobacco with Epstein-Barr virus (EBV), which is also known as human herpes virus 4 (HHV-4), has also been implicated as a potential cause of oral cancer (Jiang et al., 2019).

As shown in Table 2.1, family history of cancer has been consistently found to be a risk factor for HNCs, but the exact mechanism is unclear (Salehiniya & Raei, 2020). The genetic predisposition to HNC is thought to be amplified by exposure to tobacco, alcohol, and diet, which suggests that genetic instability itself is a risk factor for oral cell carcinoma (Salehiniya & Raei, 2020). Alcohol consumption has been linked to HNCs, but there is some question as to the level of confounding with tobacco use, as these exposures are also often linked with each other (Gormley et al., 2022; Huber & Tantiwongkosi, 2014; Salehiniya & Raei, 2020). “Epigenetics” is a term used to refer to mechanisms that impact the expressed phenotype without changing the underlying DNA sequence being expressed (Ghantous et al., 2018). In their literature

review, Ghantous and colleagues (2018) found that many epigenetic alterations have been described in the progression of OSCC that are induced by alcohol or tobacco. These mechanisms include hypermethylation in genes that act to suppress tumours, genome-wide hypomethylation, changes in patterns of methylation throughout genes, modification in histones, and alterations in non-coding RNA (Ghantous et al., 2018). The findings from that review provide potential biomarkers for further study (Ghantous et al., 2018). These findings are also consistent with a systematic review and meta-analysis of the synergistic impact of exposure to both alcohol and tobacco as a risk factor for OSCC (Mello et al., 2019). Their meta-analysis results estimated that the synergistic consumption was associated with an odds ratio (OR) of 5.97 (95% CI 3.54 to 8.14) compared to no consumption (Mello et al., 2019). When stratified by type of tobacco in the synergistic relationship, they found an OR of 4.74 (95% CI 3.51 to 6.40) for smoked tobacco, an OR of 7.78 (95% CI 2.18 to 21.14) for smokeless tobacco, and an OR of 16.17 (95% CI 7.97 to 32.79) for alcohol, smoked tobacco, and smokeless tobacco combined compared to no consumption (Mello et al., 2019). In their pooled analysis of case control studies from younger adults with HNC, Toporocov and colleagues (2015) found the same risk factors as for adults, although they noted that the proportion of never-smokers and never-drinkers is higher in younger age groups, and the attributable fractions (AFs) are lower.

The fact that globally, there is a strong association of HNCs with low SES even after adjusting for tobacco exposures suggests that the

impact of genetics is relatively minor, and there are many environmental causal parameters that are not yet identified (Salehiniya & Raei, 2020). Other risk factors that have been studied included oral and dental health, body mass index (BMI) and physical activity patterns, and marijuana smoke, all of which may impact the risk of HNCs, but none of these are thought to elevate risk much compared to the other exposures listed in Table 2.1 (Gormley et al., 2022).

Epigenetic Studies

Jiang and colleagues (2019) conducted a review of the tobacco-induced carcinogenic pathways in OSCC. Overall, their findings were consistent with that of Ghantous et al. (2018), in that they found that tobacco can cause epigenetic alteration of oral epithelial cells. Generally, these changes inhibit several different systemic immune functions in the host which is believed to lead to tumour formation (Jiang et al., 2019). In addition, tobacco causes oxidative stress on tissues through toxic metabolites which induce OSCC (Jiang et al., 2019). Admittedly, it is not clear how exactly these changes result in tumour formation, nor is it clear how that process would be impacted by the presence of bacteria.

Results of other epigenetic studies can be considered to search for candidate biomarkers for study, especially with respect to tobacco-induced cancer (Ambatipudi et al., 2016; de la Iglesia et al., 2020). In one such study, analysts took samples from the European Prospective Investigation into Cancer and Nutrition (EPIC) study, which is a large epidemiologic observational study conducted across ten European countries (Ambatipudi et al., 2016). These researchers examined the

potential of DNA methylation changes as a marker of exposure to tobacco smoke using 910 peripheral blood samples from undiagnosed participants (Ambatipudi et al., 2016). They identified 748 CpG sites that differed between smokers and non-smokers in terms of methylation changes (Ambatipudi et al., 2016). Importantly, they found that some methylation changes were reversable after smoking cessation, although specific genes that were differentially-methylated remained that way even 22 years after cessation (Ambatipudi et al., 2016).

Other epigenetic studies were designed differently to focus on various specific biomarkers. In one study, researchers studied tumour specimens from 177 patients with various tobacco-smoking histories with SCC of the head and neck (de la Iglesia et al., 2020). In their study, they analysed CD3, CD8, FOXP3, PD-1, PD-L1, and pancytokeratin, and found that current smokers have lower CD8+ cytotoxic T-cells and PD-L1+ cells in the tumour immune microenvironment (TIM) (de la Iglesia et al., 2020).

Study Design Considerations in Cancer Biomarker Clinical Trials

Hu and Dignam (2019) developed guidance for researchers looking to include biomarker studies in oncology clinical trials. In their guidance, they define the term “biomarker” to mean the characterization of biologic molecules (including deoxyribonucleic acid (DNA) and ribonucleic acid (RNA)), proteins, or metabolites from the analysis of body fluids or tissues for diagnostic purposes (Hu & Dignam, 2019). As per this definition, oral microbiome biomarkers – specifically molecules produced by the microbiome as well as microbes of the microbiome - in

OSCC should be included. Their paper was in response to the rise in “precision oncology”, referring to the ability to measure many biomarkers previously unmeasurable through the advent of various omics-based technologies, especially NGS (Hu & Dignam, 2019). As laboratory studies involving potential cancer biomarkers were increasingly being reported, those conducting clinical trials were increasingly pressured to include biomarker measurements in their study designs (Hu & Dignam, 2019). Hu and Dignam (2019) provide a basic outline of epidemiologic study designs that can be utilized to study such biomarkers.

Before presenting their study designs, the authors described important aspects that will need to be taken into account when doing epidemiology with cancer biomarkers (Hu & Dignam, 2019). First, they pointed out that even if a certain treatment is believed to be more effective in a patient who is positive for a particular biomarker, that does not necessarily mean that a biomarker-negative patient will not benefit (Hu & Dignam, 2019). This issue leads to the importance of identifying “companion biomarkers”, or other biomarkers that are correlated with the target biomarker, and can help establish a partition or cut-off value to make biomarker-positive vs. biomarker-negative determinations for the target biomarker (Hu & Dignam, 2019).

The authors also pointed out that not all biomarkers may have enough evidence to support their inclusion in a clinical trial (Hu & Dignam, 2019). First, biomarkers selected for inclusion in a clinical trial will need to be shown to be able to be measured accurately, and measurements should be demonstrated to be replicable (Hu & Dignam,

2019). Next, a determination will need to be made as to whether the biomarker is “prognostic” (associated with disease prognosis, regardless of treatment) or “predictive” (meaning that its prognostic influence differs depending upon treatment) (Hu & Dignam, 2019). Another consideration will need to be made as to whether the biomarker will be “integral” to the study (meaning part of the inherent study design) or “integrated” in the study (meaning the subject of specific testable hypotheses) (Hu & Dignam, 2019). The authors also recommend several new, adaptive designs that can be used for such biomarker studies (Hu & Dignam, 2019).

Hu and Dignam (2019) also introduced new terminology in their paper. They define a “basket trial” as a trial of an agent tested among multiple disease types that share a common molecular feature or target that can be identified by the biomarker (Hu & Dignam, 2019). They contrast this to an “umbrella trial” design, where for a common disease entity, multiple agents are investigated along with specific molecular targets and biomarkers (Hu & Dignam, 2019). The article goes on to describe the features of these study designs, and how they will produce knowledge about cancer prognosis and treatment (Hu & Dignam, 2019).

Study Design Considerations for Cancer Biomarker Laboratory Studies

This guidance from Hu and Dignam (2019) provides insight into the design of research about therapeutic agents for cancer, in that it implies that certain features could be included in cancer biomarker laboratory study designs to provide more definitive information for those

who design clinical trials of potential therapeutic agents. For example, most study designs recommended in the guidance to only include one biomarker, and the authors recommend that this biomarker should be well-characterized in terms of its known companion biomarkers (Hu & Dignam, 2019). They also pointed out that the measurement for this biomarker should be shown to be accurate, reliable, and replicable (Hu & Dignam, 2019). Although this may seem obvious, experimental biomarkers might be measured using different methods, and these methods could introduce significant variability in the results (Hu & Dignam, 2019). Determining a specific measurement of a biomarker is accurate and replicable should take place before clinical research commences (Hu & Dignam, 2019). Biomarkers that are the subject of human trials should also have evidence-based cut points derived to determine positive vs. negative biomarker status (Hu & Dignam, 2019). The authors included a table describing different recommended biomarker study designs, including enrichment, biomarker-stratified, biomarker strategy, biomarker-directed, umbrella and basket (Hu & Dignam, 2019). In most study designs proposed in the table, the biomarker is integral to the study, underscoring why it would need to be able to be measured accurately using a reliable measurement approach that is easily replicated (Hu & Dignam, 2019).

Also, most clinical trial designs recommended by the authors would be for confirmatory intent (Hu & Dignam, 2019). At this time, OSCC microbiome biomarkers would not be at this stage. A systematic review and meta-analysis of the OSCC microbiome indicated that the

current microbiome biomarker studies have important limitations in basic methods of data collection, processing, and reporting (Peter et al., 2022). Another articles proposed OSCC microbiome biomarkers that could be used for screening, but these are not able to be measured accurately, reliably and in a replicable manner at this time (Doddawad et al., 2022). This is partly because the study design proposed by Hu & Dignam (Hu & Dignam, 2019) must rely on a body of literature where appropriate statistical considerations were made, and currently, this does not apply to the OSCC microbiome biomarker literature (Gloor et al., 2017; Goossens et al., 2015; Ou et al., 2021).

As described earlier, in a table presented in their paper, Hu and Dignam (Hu & Dignam, 2019) recommended different study designs for cancer biomarkers, and include a column stating the objective of "confirmatory intent". This indicates that for the cancer biomarker laboratory literature to be most useful, it should encourage a transition from discovery designs early in the investigation of a biomarker to confirmatory designs as evidence accumulates. As an example, in their paper, Hu and Dignam (2019) cite the case of vemurafenib being approved for the treatment of *BRAF*V600 mutation-positive metastatic melanoma. Advancing to this point in translational medicine suggests that the laboratory evidence behind *BRAF*V600 mutation-positive metastatic melanoma and the ability to treat it with vemurafenib was rigorous enough for these elements to have been included in a clinical trial.

Also, since cancer has so many different risk factors, for laboratory biomarker studies to provide evidence useful for epidemiology, it is important that participants – especially in small studies – are subject to highly-specific and well-measured case definitions, and that studies utilize qualification criteria that filter in an extremely homogenous sample, especially with respect to known risk factors (e.g., tobacco use). As an example cited in the paper, the ALCHEMIST (Adjuvant Lung Cancer Enrichment Marker Identification and Sequencing Trial) was designed to study biomarker-specific cohorts of patients with resectable non-small-cell lung cancer (NSCLC) (Hu & Dignam, 2019). Without evidence from rigorous laboratory studies of biomarkers in well-characterized NSCLC patients, it would not be possible to design a study like ALCHEMIST, which emphasizes the importance of laboratory study design in the process of translational cancer research (Hu & Dignam, 2019).

The Oral Bacteriome in OSCC

This section will provide a history of the exploration of the oral bacteriome in OSCC.

Early Studies

The link between the oral bacteriome and OSCC was first explored to better understand the observed link between poor oral hygiene and morbidity associated with OSCC post-treatment infections (Nagy et al., 1998). Authors examined the microbial composition in specimen collected from patients with OSCC (Nagy et al., 1998). In their study, researchers swabbed the OSCC lesion surface and compared it

to an adjacent normal area using culture-based techniques in order to determine bacterial relevance to patient morbidity (Nagy et al., 1998). Although no inferential statistics were done, the total number of species within swabs (both aerobes and anaerobes) were found to be higher in cancer samples as compared to the normal controls (Nagy et al., 1998). They also found that *Fusobacterium* species, among others, were present in higher amounts in tumour as compared to non-tumour areas (Nagy et al., 1998). Despite the limitations of the culture methods, this study was the first to provide evidence of differences in bacterial composition between oral cancerous and healthy sites (Nagy et al., 1998).

About a decade later, Mager et al (2005) studied salivary microbiota as a diagnostic indicator of oral cancer in both OSCC and healthy patients using DNA-DNA hybridization. In that study, the methodology used to analyse the salivary bacterial community was restricted to 40 oral bacterial species, hence comprehensive knowledge of the tumour microbiota cannot be inferred from such design (Mager et al., 2005). In their study, they reported that they found three oral bacteria (*Prevotella malenginogenica*, *Streptococcus mitis* and *Capnocytophaga gingivalis*) to be significantly increased in OSCC salivary samples ($p < 0.001$) (Mager et al., 2005). However, upon careful reading, this p-value interpretation is likely inaccurate, because no a priori hypothesis was posed (Mager et al., 2005). Instead, the p-value is likely the result of a *post hoc* analysis that is interpreted without an adjustment to the p-value, hence the difference reported might be due to chance (Goossens et al.,

2015; Mager et al., 2005). Even so, when interpreted as descriptive analysis, it is notable that these oral bacteria were the most prevalent (Mager et al., 2005). Author proposed that if reliable results could be obtained, oral bacteria could be used as biomarkers for the initiation of oral cancer (Mager et al., 2005).

16s rRNA Gene NGS and the Oral Bacteriome in OSCC

As described earlier, this comprehensive review is focused on the use of 16s rRNA gene NGS of the oral bacteriome in the study of OSCC. While OSCC is the specific topic of this inquiry, it is acknowledged that OSCC falls within the larger category of HNCs; therefore, biomarker studies including 16s rRNA gene NGS in the research of OSCC may also include other groups of participants (healthy comparisons, or those with other disease entities, including other HNCs) (Salehiniya & Raei, 2020). Because of this, it might be difficult to compare results from study to study, because different populations will be sampled. Biomarkers may be measured in different types of sample – such as saliva vs. tissue – and this could produce results that might be difficult to compare from study to study (Hu & Dignam, 2019).

Further, while biomarker studies of the oral bacteriome using 16s rRNA gene NGS in the investigation of OSCC will likely use similar measurement methods, they will likely use a variety of study designs, in that some will be designed to be exploratory (i.e., intended for discovery), and others will be confirmatory (i.e., based on a presupposed hypothesis) (Hu & Dignam, 2019). Results from well-designed

confirmatory studies could be used to inform an epidemiologic study design like the ones described by Hu and Dignam (2019).

As a general rule, evidence-based medicine (EBM) follows a framework where the lowest level of evidence is reserved for case studies and expert opinion, with other observational designs (such as cross-sectional and case-control studies) in the middle, and randomized controlled trials (RCTs) as the highest-quality individual study design (Semrau et al., 2023). Systematic reviews and meta-analyses appear at the top of the pyramid as the highest quality of scientific evidence (Semrau et al., 2023). While it is unreasonable to expect that laboratory studies using 16s rRNA gene NGS in the investigation of the oral bacteriome in OSCC would use advanced or complex study designs, it is not unreasonable to expect them to take into account important aspects of basic study design (e.g., determining whether a control group is necessary, and if so, how to select one with the least bias) (Semrau et al., 2023; Zaura et al., 2021).

In further consideration of oral bacteriome studies in OSCC, Hooper and colleagues (2006) conducted a similar study as Mager and colleagues (2005) with the aim of identifying viable bacteria present in OSCC tissue, but were able to use 16s rRNA gene (without the NGS technology – they sequence individual isolates using standard low throughput technology). These authors cultured the microbiota present within tissues of excised OSCC tumours (as well as corresponding superficial sections of these tissues) from male and female OSCC patients in South Wales and analysed them by 16s rRNA gene

sequencing (Hooper et al., 2006). Their study revealed the existence of heterogeneous viable species (mainly bacteria) within these tissues with a tendency for saccharolytic and acid tolerating species in tumours (Hooper et al., 2006). The same author group also analysed the bacteriome of OSCC tissue in a similar small sample by applying a non-culture based approach using 16s rRNA sequencing (Hooper et al., 2007). The authors interpreted both sets of results to support the conclusion that a higher affinity of saccharolytic and aciduric bacteria reside in tumour tissues (Hooper et al., 2006, 2007).

Pushalkar et al (2011) were the first to utilize NGS technology to analyse the differences in bacterial composition between OSCC patients and normal controls. In this pilot study, which did not use inferential statistics (that apparently took place in the US, although it was not clearly stated), 67% of the total sequences were either from uncultivable bacteria or unassigned bacteria indicating a significant gap in the then existing knowledge that required further exploration (Pushalkar et al., 2011). These authors were also able to identify fifteen bacteria that were exclusive to OSCC patients, among which two species were found to be in agreement with the previous finding by Mager et al (2005) (i.e., *Prevotella melaninogenica and Porphyromonas gingivalis*) (Pushalkar et al., 2011). The same author group used Sanger sequencing (16s rRNA gene) of bacteria residing in OSCC tissues in another study (Pushalkar et al., 2012). The results of this study, which again did not use inferential statistics, revealed a shift in the composition of the microbiota from Gram-negative to Gram-positive by 19% in cancer tissues as compared

to the controls (Pushalkar et al., 2012). Also, apart from a few species, most of the bacteria identified were common between the two groups, with differences in their distribution in each group (Pushalkar et al., 2012). These findings of altered bacterial composition within cancerous samples suggested a link that warrants additional exploration using more rigorous study designs and inferential statistics.

Schmidt and colleagues (2014) studied samples of cancer and healthy tissue from five patients with oral cancer as well as samples from pre-cancer patients in the US, and did not use inferential statistics. In their analysis, the authors found that the abundance of the *Streptococcus* genus was decreased in both precancerous and cancerous samples and thus might indicate these as a transitional change in the microbiome (Schmidt et al., 2014). It was also observed that the genus *Fusobacterium* was increased in abundance in cancer samples as compared to the matching non-cancerous samples (Schmidt et al., 2014). The authors interpreted this finding to indicate its later association in the microbial shift as compared to *Streptococcus* (Schmidt et al., 2014).

Al-Hebshi et al (2015) applied a new approach to analyse the sequencing data which they tried on three OSCC samples. The new algorithm used enabled the researchers to successfully assign almost all of the reads to species level (Al-Hebshi et al., 2015). This revealed the presence of thirty-five species in all the samples, some of which were present in high abundance (Al-Hebshi et al., 2015). Unfortunately, the relevance of this is unknown due to the lack of a normal reference - as

the detected bacteria are common oral microbiota (Al-Hebshi et al., 2015). Of interest*, Bacteroides fragilis* was detected in two out of the three samples used, which is interesting due to the fact that this bacterium is a rare finding in the oral microbiome (Al-Hebshi et al., 2015).

In a similar approach, Mok et al. (2017) studied the oral microbiome in normal (n=9), those with oral potentially malignant disorders (OPMD, n=9), and those with oral cancer (n=9) from the Malaysian population. The authors detected taxa that tend to cluster into three groups (normal, dysplasia and cancer) where taxa of dysplasia group overlapped between the other two (Mok et al., 2017). This provides more evidence for the transitional shift of the microbiome from healthy to disease. As with earlier studies, this study did not use inferential statistics, and quantified the overlap of taxa identified in the dysplasia group with the other groups (Mok et al., 2017). The authors contend that their results provide more evidence for the transitional shift of the microbiome from healthy to disease (Mok et al., 2017).

Recently, Al-Hebshi et al. (2017) further applied their prioritised BLASTN-based algorithm, which previously enabled a species-level taxonomy assignment, to a larger study of 20 OSCC samples and 20 normal controls. The samples came from anonymized leftover DNA extracts from fresh OSCC biopsies obtained from a previous study in Yemen, and the healthy controls were recruited using gender- and age-matching from a Faculty of Dentistry in Jazan University in Saudi Arabia (Al-Hebshi et al., 2017). These authors found *Fusobacterium nucleatum* subspecies *polymorphum* and *vincentii* to be associated with OSCC

which was consistent with their previous study (Al-Hebshi et al., 2017). In addition to *Fusobacterium nucleatum*, *Pseudomonas aeruginosa* was found to be associated with OSCC, a finding that has not been reported before (although no inferential statistics were used) (Al-Hebshi et al., 2017). The authors also examined the potential function of the bacteria present, and genes enriched in OSCC were predicted to be involved in pathways similar to that of bacterial genes in chronic periodontitis (e.g. bacterial motility, lipopolysaccharide (LPS) biosynthesis and flagellar assembly) (Al-Hebshi et al., 2017). Their predictions suggest that there is a bacterially-induced inflammatory background in OSCC (Al-Hebshi et al., 2017).

In another pilot study, Wolf et al (2017) analysed the oral salivary microbiome in 11 patients with oral and oropharyngeal SCC compared to 11 healthy controls from a European country. These authors examined the microbiome of OSCC saliva using NGS, and found the genus-level abundance of bacteria associated with OSCC was similar to some previous studies, but different from other, as reported in their Table 1.1 (Wolf et al., 2017). These differences in the findings are perhaps due to the lack of lower-level taxonomy assignment (i.e., species level), thus overlooking a true bacterial association with OSCC, but also likely suffered from a lack of inferential statistics (Goossens et al., 2015; Wolf et al., 2017). However, they were in overall agreement with the fact that the microbiome associated with the OSCC and healthy groups are distinct (Wolf et al., 2017). They also looked at the potential functional role of the microbiome in each group (Wolf et al., 2017). They postulated

that the OSCC microbiome would be potentially more involved in sugar and carbohydrate metabolism as well as stress response mechanism compared to the healthy group, while that of the healthy group would be more involved in antioxidant activity and metabolism of lipids compared to OSCC (Wolf et al., 2017). Interestingly, the microbiota of HPV-positive cancers in this study were found to be more closely related to the normal microbiota rather than that of the HPV-negative OSCC samples (Wolf et al., 2017). This led the authors to speculate that HPV-positive and HPV-negative OSCCs are two distinct diseases (Wolf et al., 2017).

In a similar study, Zhao et al (2017) conducted a case series study where they took 40 swabs of OSCC lesions and compared them to anatomically matched normal sites in Chinese patients, and analysed them using 16s rRNA gene sequencing. In their study design, they assigned detected taxa to species level (Zhao et al., 2017). The authors found different species of *Fusobacterium*, amongst other bacteria, to be increased in the OSCC samples as compared to the normal samples, although the authors did not use inferential statistics (Zhao et al., 2017). One aim of the study was to predict ecological relationships across the different bacterial communities detected (Zhao et al., 2017). Seven members of the *Fusobacterium* genus, which were shown to be ecologically connected in OSCC samples in this study, were used in a simulation to predict diagnosis (Zhao et al., 2017). Specifically, they were entered into a prediction model and graphed in terms of diagnostic power (Zhao et al., 2017). In addition, functional prediction of bacterial genes in OSCC samples suggested a decrease in amino acid and lipid

metabolism, membrane transport and signal transduction, bacterial functions that were also predicted to be reduced in periodontitis (Zhao et al., 2017). The authors speculate that common functional prediction findings might also indicate an inflammatory background in OSCC (Zhao et al., 2017).

Measurement and Statistical Approaches in OSCC Gene-targeted Bacteriome Studies

As described earlier, an important consideration in cancer biomarker studies is the quality of the measurement as biomarker measurements must be reliable, replicable, and interpretable to be utilized as an integral or integrated part of an cancer clinical trial (Hu & Dignam, 2019). This section will provide an overview of measurement, epidemiologic, and statistical approaches in cancer biomarker studies, and specifically studies of the human microbiome using 16s rRNA gene NGS.

Early Measurement Approaches in Oral Microbiome Studies

The earliest article in the review was published in 2011 (Pushalkar et al., 2011). Although these authors were not the first to try this, they used culture-dependent methodology previously, and followed up with a manuscript on 16s rRNA gene NGS because it is less biased (Pushalkar et al., 2011, 2012). Scientific writing on this topic in this early period provides a window into the measurement philosophies over a decade ago related to studying microbiota in OSSC, and how they might apply to hypotheses derived at the time.

Diversity in Oral Microbiome Studies

In this early article, Pushalkar and colleagues (2011) made it clear that they were using the methods in their study to evaluate whether the application of “diversity” and “relative abundance” measures could be useful in the pursuit of trying to identify a member of the oral bacteriome in OSCC that might be responsible for most of the cancerous tumours (Pushalkar et al., 2011). Specifically, they pointed out the situation where *Helicobacter pylori* was found to be responsible for 60 to 90% of gastric cancers, and suggested that if the entire diversity of OSCC microbiota can be measured, perhaps it could be possible to identify a specific bacterial infection responsible for a large percentage of OSCC cases (Pushalkar et al., 2011). Their contention was that up until the availability of NGS, much of the oral bacteriome went unmeasured, so it would have been impossible to identify a bacterial infection responsible for a large proportion of OSCC cases if there was one (Pushalkar et al., 2011). The obvious counterargument is that biomarkers for either OSCC or healthy tissue or saliva need not to be diverse to be accurate. According to their contention, For example, if an OSCC lesion is quantitatively more diverse in terms of the oral bacteriome in diseased samples compared to healthy samples, it would suggest that there are members of the bacteriome worth identifying and studying further to see if they have a causal relationship with the tumour (Pushalkar et al., 2011). Hence, it is not the actual diversity of the members of the oral bacteriome which is the intention of the measurement; it is instead more of an approach to identifying the members of the oral bacteriome as biomarkers that may

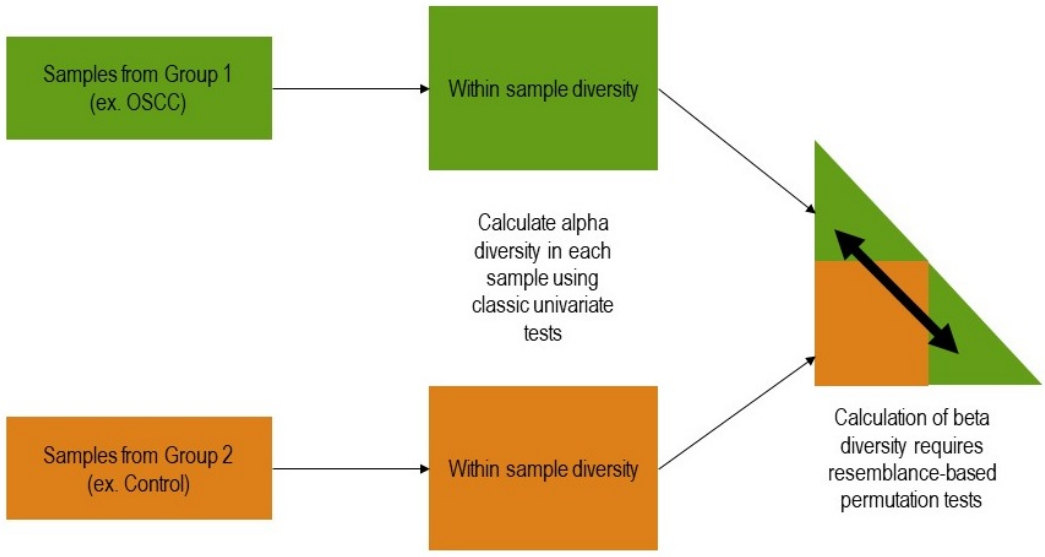
have been unmeasured before, and might be revealed to be important targets for study (Pushalkar et al., 2011).

In the study of the human oral microbiome, as with study of other microbiomes, the concept of diversity was extended to define α (alpha) diversity vs. β (beta) diversity, which are ideas lifted directly from statistics and computation, and not evolutionary biology (Kers & Saccenti, 2022). Laboratory biostatisticians Kers and Saccenti (2022) provided guidance on choosing measures of α diversity and β diversity *a priori* when designing laboratory studies. This is because in order to reduce Type I error (false-positive results), it is necessary to state hypotheses *a priori* – before collecting the data in the study – and to operationalize those hypotheses to specific measurements (Kers & Saccenti, 2022). So if there is a hypothesis that there is statistically significant α diversity in samples from Group 1, or that there is a statistically significant difference between the β diversity of the samples in Group 1 vs. Group 2 compared to the samples in Group 2 vs. Group 3, the selection of the diversity metric of record is necessary when formulating the hypothesis *a priori* to preserve statistical power (Kers & Saccenti, 2022). If the researcher fails to select their operationalized diversity metrics when stating their hypotheses *a priori*, and instead simply runs multiple *a posteriori* tests using different α and β diversity metrics hunting for statistical significance, it is considered “fishing” for a p-value, and not statistically legitimate (Kers & Saccenti, 2022). In addition to being seen as fraudulent (“p-hacking”), the practice does not provide guidance on how to interpret the results of the study, because

certainly all measures of the same type of diversity on the same group of samples will be correlated, but may not produce identical interpretations (Kers & Saccenti, 2022).

The intention of the paper was to provide guidance to laboratory researchers in the design of their hypotheses and the operationalization of diversity metrics when stating hypotheses *a priori* that will be used to analyse results of their microbiome studies after the samples are measured (Kers & Saccenti, 2022). As described by Kers and Saccenti (2022), theoretically, α diversity is intended to be a measure of "within-sample diversity", while β diversity represents "between sample" diversity (see Figure 2.1).

Figure 2.1. Explanation of α vs. β diversity calculations in microbiome analysis.



Note: Adapted from Kers and Saccenti (2022)

As shown in Figure 2.1 and described by Kers and Saccenti (2022), to statistically test if there is within-sample diversity (α diversity) in any group of samples, a particular metric can be selected for use, and subjected to a statistical test. The α diversity metrics described in the paper that may be chosen include richness (defined by the number of taxa identified), phylogenetic diversity (a phylogenetically weighted measure of richness), and the Chao1 index (an abundance-based non-parametric estimator of taxa richness) (Kers & Saccenti, 2022). They also recommend two indices which are different estimators of taxa diversity that combine richness and evenness (also known as relative abundance), including the *H* estimator from Shannon's Index, and the *D* estimator from Simpson's Index (Kers & Saccenti, 2022). Abundance and relative abundance measures are discussed in the next section.

As demonstrated by experimental case studies included in the paper by Kers and Saccenti (2022), these authors recommend selecting one α diversity metric, then using a non-parametric Kruskal-Wallis test, with the grouping variable (e.g., taxa labels) as the independent variable, and the calculated α diversity metric as the dependent variable. Interpreting the p-value derived from this test will determine whether or not the researcher rejects the null, and finds the sample or groups of samples statistically significantly positive for α diversity (Kers & Saccenti, 2022).

In terms of β diversity metrics, the authors describe the Bray-Curtis Dissimilarity index (which measures the compositional dissimilarity between the microbial communities of two samples), and

the Jaccard index (another way of calculating dissimilarity), both of which are computations from statistics and not specific to microbiome studies (Kers & Saccenti, 2022). The authors also describe using UniFrac (UF) distances to compare β diversity, which in both their weighted and unweighted forms, take into account the phylogenetic tree, and distances between community members (Kers & Saccenti, 2022).

For statistical testing of β diversity, the authors recommend a PERMANOVA approach, which is an analysis of variance- (ANOVA-) like test that uses random permutations to compare matrix similarities and dissimilarities (Kers & Saccenti, 2022). The null hypothesis for the PERMANOVA is the calculated space represented by each sample's diversity is the same for all groups tested; hence, rejecting the null would be interpreted as statistically significant β diversity among groups in the test (Kers & Saccenti, 2022). It is important to emphasize here that since PERMANOVA is ANOVA-like, it only produces one p-value; therefore, if multiple participant groups are being tested in the PERMANOVA, if any one group in the analysis has statistically significantly different β diversity with respect to one other group, this statistic will be statistically significant, and will not be able to differentiate between the individual groups without further *post hoc* analysis.

Abundance and Relative Abundance in Oral Microbiome Studies

In the earliest article reviewed by Pushalkar and colleagues (2011), the authors go on to imply that not just diversity, but relative abundance of the bacteria in the oral microbiome might be an important clue as to which members of the oral bacteriome might be responsible

for tumour progression. As mentioned earlier, the term “relative abundance” is also called “evenness”, a theoretical estimate of the prevalence of the bacteria in the sample, which in practice only estimates the numerator (Mandal et al., 2015). Hence, reporting absolute abundance, or the measurement of the level of taxa detected in the sample, is not accurate, there is some debate over the best representation for a relative abundance metric (Mandal et al., 2015). Nevertheless, the point of comparing relative abundance in samples is to estimate if a sample is dominated by the presence of a particular taxa compared to another sample, which cannot be estimated with diversity measures, although some indices attempt to account for abundance to some degree (Kers & Saccenti, 2022; Mandal et al., 2015).

Given this understanding, quantifying the relative abundance of different members of the oral bacteriome at different stages in the natural history of the OSCC tumour may shed light on whether there are certain members associated with progression, as those exerting a greater impact at any given time will be likely to also have a relative abundance that is higher than the other members of the bacteriome which are also measurable (Pushalkar et al., 2011). While this is reasonable, the unfortunate assumption underlying the hypothesis put forth by Pushalkar et al. (2011) in this early study is that is a member of the bacteriome that has the same relationship with OSSC that *H. pylori* has with gastric cancer, In other word, the underlying hypothesis is that there is a single biomarker that has an outsized role in promoting OSCC (Pushalkar et al., 2011). While this hypothesis may be true, the study designs that have

been used subsequently in this line of research were not designed to identify such a member, and did not claim to be designed around such a hypothesis (J.-W. Chen et al., 2021; Gaziano et al., 2021; Gopinath, Kunnath Menon, et al., 2021; Gopinath, Menon, et al., 2021). The authors also speculate that instead of a dominant member of the bacteriome, there may be multiple members that initiate inflammation that leads to OSCC (Pushalkar et al., 2011).

The Human Microbiome Project and the Human Oral Microbiome Database

In 2010, researchers from the Forsyth Institute in Boston announced in a published paper that they had been awarded a large grant from the National Institutes of Health (NIH) to amass a database about the human oral microbiome called Human Oral Microbiome Database (HOMD), which is part of the larger Human Microbiome Project (HMP) (T. Chen et al., 2010; NIH Human Microbiome Portfolio Analysis Team, 2019). The intention of the HMP was to leverage findings from studies of the human microbiome to facilitate the design of human health interventions as described earlier with immunotherapy approaches (Hu & Dignam, 2019; NIH Human Microbiome Portfolio Analysis Team, 2019). These anticipated human health interventions include the immunotherapy approaches described earlier (Hu & Dignam, 2019), and may result from models, integrated datasets, and computational tools all provided under the HMP (NIH Human Microbiome Portfolio Analysis Team, 2019).

The NIH Human Microbiome Portfolio Analysis Team reported that in fiscal years 2012 to 2016, the project spent a total of $376 million on microbiome projects using human cohorts, and of that, $40 million focused specifically on the oral microbiome (NIH Human Microbiome Portfolio Analysis Team, 2019). It is likely that this larger allocation supported the smaller grant to develop the HOMD (T. Chen et al., 2010; NIH Human Microbiome Portfolio Analysis Team, 2019). In their article, the researchers awarded the HOMD contract report that the primary goals of their contract with NIH is to establish a stable taxonomic structure for emerging taxa that will be identified as part of the project and are currently unnamed, and to provide tools for analysing 16s rRNA gene sequence data and other oral genomic data (T. Chen et al., 2010).

Curiously, the article announcing the creation of the HOMD does not explain its philosophy on basic epidemiologic and biostatistical issues, even though it is essentially creating an epidemiologic registry (T. Chen et al., 2010; Hu & Dignam, 2019; Vogtmann et al., 2023). One foundational epidemiologic issue with the HOMD approach that is not discussed in the article is how the genomics of the underlying populations represented in the HOMD could potentially bias the function of the entire database (T. Chen et al., 2010; Hu & Dignam, 2019; Vogtmann et al., 2023). For example, if tumours from high SES countries (as in northern Europe) are overrepresented in the database, then scientific inferences about biomarkers will not be replicable across populations (e.g., with respect to lip cancer) (Hu & Dignam, 2019; Salehiniya & Raei, 2020; Vogtmann et al., 2023). Further, a fundamental

biostatistical consideration not addressed in the article is how the authors propose researchers should use the HOMD data to execute studies with large amounts of correlated biomarkers without encountering Type I error (T. Chen et al., 2010; Mandal et al., 2015). This statistical consideration is important regardless of the type of investigation being conducted; it is not just an issue with therapeutic studies (Gloor et al., 2017). The intention of the HOMD maybe to catalogue the existence of oral microbiomes, but it is not possible to apply epidemiologic or statistical inference to the data without the necessary metadata. Further, the lack of quantification of bias makes utilizing such a database for training machine learning (ML) algorithms questionable. The article only describes the mechanical and operational function of their taxonomic and digital process, and does not address any epidemiologic or biostatistical issues (T. Chen et al., 2010). Of course, it is possible that providing such high quality of data was not the intention of this database. In that case, these data may be useful for other purposes, but might not be very accurate when used in ML algorithms (such as those run in QIIME 2) (Estaki et al., 2020).

Today, the HOMD is a set of web-enabled tools that provide many online functions related to the analysis of these sequences, including a reference sequence lookup (*HOMD :: Human Oral Microbiome Database*, n.d.). However, even today, there is no clear documentation throughout this online set of tools related to the epidemiologic and biostatistical considerations behind this registry described above, especially with respect to the bias that could impact ML models trained

on the contents of the database (*HOMD :: Human Oral Microbiome Database*, n.d.). An argument can be made that this was not the purpose of the database. In their article, the authors stated that, “the goal of the contract [to set up the HOMD] was to create a stable taxonomic structure for the unnamed oral taxa and to provide tools for analysing 16S rRNA sequence data and oral genome data” (T. Chen et al., 2010). It is unclear how the goal of the HOMD relates to the overarching objectives of the HMD, which is aimed at understanding human disease through epidemiologic and biostatistical approaches.

As described earlier, the HMD and the HOMD are funded by the NIH which stands for the National Institutes of Health in the US (T. Chen et al., 2010; NIH Human Microbiome Portfolio Analysis Team, 2019). NIH functions as a national medical research agency, and is part of the Department of Health and Human Services (DHHS), which functions as the US version of a ministry of health (Vogtmann et al., 2023). Other population-level epidemiologic initiatives are funded through the US government, including an annual surveillance effort called the National Health and Nutrition Examination Survey (NHANES) (Vogtmann et al., 2023). Vogtmann and colleagues (2023) found they needed to create their own sample that was representative of the oral microbiome data for the US population utilizing stored samples from the US National Health and Nutrition Examination Survey (NHANES) (available from the authors on request). The lack of epidemiologic documentation precluded the use of data from the HOMD. Because the HOMD is being used as the reference database in many oral microbiome studies, and it has known

biases that to this date have been undiscussed and uncharacterized, results based on data from the HOMD must be interpreted with extreme caution, and not considered generalizable to the human condition (Hu & Dignam, 2019; Vogtmann et al., 2023).

Using Operational Taxonomic Unit (OTU) Clustering in 16s rRNA Studies

Although the advent of 16S rRNA NGS provided the ability to measure sequences previously unmeasurable, this issue brought about challenges in operationalizing results output from the sequencing approach that could be used functionally in biostatistics to answer presupposed hypotheses (Gopinath, Kunnath Menon, et al., 2021; Nguyen et al., 2016). For example, for the diversity metrics, what actual measures are being used to represent the values of the dependent variables, and how are they derived from the samples analysed? The most common way to solve this problem was to cluster the sequences output as “Operational Taxonomic Units” (OTUs) (Nguyen et al., 2016). Although the OTU approach has been used throughout 16s rRNA sequencing research, it has some important limitations that need to be considered (Nguyen et al., 2016).

First, in the process of computing OTUs from sequences present in a laboratory sample, a similarity threshold is set for sequence similarity (typically at 97%) (Nguyen et al., 2016). Next, these sequences are clustered together into an OTU, and then from each OTU, a single sequence is selected and annotated, and this annotation is used throughout the rest of the remaining sequences (Nguyen et al., 2016).

These sets of operations are called “pipelines” which are operationally built into software to execute the pipeline; two commonly-used pipelines are QIIME and MOTHUR (Nguyen et al., 2016). Nguyen and colleagues (2016) speculated that OTUs have been popular because they reduce the output from millions of reads from 16s amplicon analysis into thousands, and this makes downstream analysis, such as multiple sequence alignment (MSA) analysis and phylogeny estimation, much easier because the dataset is smaller and representative. Using this smaller, more manageable data also speeds up the whole analysis (Nguyen et al., 2016).

However, this ease and functionality comes with serious downsides (Nguyen et al., 2016). First, the “percent sequence similarity” calculation can overestimate evolutionary similarity between pairs, so using evolutionarily-corrected distances based upon a MSA is recommended (Nguyen et al., 2016). Next, it can be difficult to honestly interpret the results from this type of an analysis (Nguyen et al., 2016). As an example, two different species may have 99% similar 16s sequences, erroneously placing them in the same OTU (e.g., *Bacillus globisporus* and *B. psychophilus*), or the same strain may have multiple copies of the 16s rRNA gene that differ by as much as 5% in some regions that then are erroneously classified into multiple OTUs (e.g., *Escherichia coli K12*) (Nguyen et al., 2016).

While the perspective of Nguyen and colleagues (2016) does not reject the utility of OTUs in 16s rRNA analysis, it does advocate for more responsible approaches to handling OTUs in analysis so as to minimize

the measurement error inherent in the method. The main point of their paper was to recommend dissimilarity metrics for quantifying the evolutionary distance between pairs of sequences that can be applied to help researchers more honestly interpret the results from the OTU analysis (Nguyen et al., 2016). Combining this advice with the advice from the Kers and Saccenti (2022) to carefully select diversity of metrics of record and operationalize them when defining research hypotheses *a priori* will lead to an optimally-designed, evidence-based study of the oral microbiome in OSCC using 16s rRNA NGS technology (Nguyen et al., 2016).

Oral Potentially Malignant Disorders and Oral Epithelial Dysplasia

Oral potentially malignant disorders (OPMDs) represent a heterogeneous group of clinical disorders that carry an elevated risk of oral malignant development (Warnakulasuriya et al., 2021). Recently, the World Health Organization (WHO) Collaborating Centre for Oral Cancer in the UK updated the list of diseases based on the current evidence to include the following OPMDs: oral leukoplakia, oral erythroplakia, proliferative verrucous leukoplakia, oral submucous fibrosis, oral lichen planus, oral lichenoid lesion and oral graft versus host disease (Warnakulasuriya et al., 2021). Each disorder is associated with different risk of malignant transformation (Warnakulasuriya et al., 2021). While risk of malignancy in OPMDs is associated with clinically abnormal mucosa, malignancy may also develop in clinically normal mucosa (Warnakulasuriya et al., 2021).

The risk of malignancy development in OPMDs is assessed by the histopathological alterations observed in the oral epithelium architecture/cytology as defined by the grade of epithelial dysplasia (Odell et al., 2021). These architectural alterations may include loss of either epithelial stratifications or/and epithelial cell cohesion; abnormal superficial mitosis; irregular stratification; drop-shaped rete ridges or/and the presence of keratin pearls within these ridges; and higher number of mitosis or/and single cell premature keratin formation (Odell et al., 2021). Alterations in cytology may include abnormal variations in cellular or nuclear size or/and shape, a higher nucleus to cytoplasm ratio, larger number and size of the nucleoli, mitotic figure atypia, and hyperchromasia (Odell et al., 2021). To date, the existence of oral epithelial dysplasia (OED) is associated with a three-tier grading system that has prognostic value for malignant transformation and is considered the gold standard (Sperandio et al., 2023). This system classifies OED into mild, moderate, and severe (Warnakulasuriya et al., 2021). Mild is characterized by dysplastic changes confined within the basal one third of the oral epithelium; moderate is where these changes are confined within the basal two thirds of the oral epithelium; and severe, where these changes extend beyond the basal two thirds of the oral epithelium (Warnakulasuriya et al., 2021).

Comprehensive Literature Review Background

Literature Review Types

Sutton and colleagues (2019) recently summarized the types of reviews in the scientific literature in an effort to provide guidance to

scientific authors. The authors acknowledged that there are many terms for the different types of reviews in the scientific literature, and grouped these into review "families" (Sutton et al., 2019). They defined the systematic review family as determining whether studies would be included in the review by way of their study design (Sutton et al., 2019). This inclusion criterion could be defined through a broad category of study design (e.g., any observational studies), or a very specific one (e.g., case series) (Sutton et al., 2019). The term "systematic" implies that once these inclusion and exclusion criteria are set up to filter in only certain study designs, the literature is learned comprehensively for all studies meeting the study design criteria on the topic of the review (Sutton et al., 2019).

Selection of Literature Review Approach

Because the intention of all three substudies (presented in Chapters 3, 4, and 5) was to utilize NGS targeting the 16s rRNA gene in the study of OSCC, a systematic review should not be attempted, because all study designs on this topic should be included. In choosing to focus on studies of OSCC using the 16s rRNA approach, a narrative review was selected, and "purposive sampling" of the literature was applied (Sutton et al., 2019). The narrative review is a type of traditional review (Sutton et al., 2019). In the narrative review, purposive sampling can be used to – for example – focus on one particular topic, which is becoming more popular (Sutton et al., 2019). Purposive sampling was achieved through the application of inclusion and exclusion criteria to the articles selected for review (Sutton et al., 2019). It may be important to

point out that while narrative review with purposive sampling is part of the traditional review family, there are other families of reviews that were not employed in this dissertation, including the rapid review family, and the purpose-specific review family, which includes scoping reviews (Sutton et al., 2019).

2.2. Research Question and Objectives

Research Question

The comprehensive review objectives were designed to answer the following research question: Based on evidence from current metagenomic studies utilizing the NGS technique to target 16s rRNA gene in OSCC, how should the oral bacteriome be characterised using descriptive analyses?

Objectives

This comprehensive literature review had the following objectives:

1. To identify scientific evidence from studies investigating high-throughput, NGS targeting the 16s rRNA gene in studying the oral bacteriome of OSCC, and
2. To characterize findings related to the oral bacteriome in OSCC from these studies.

2.3. Methods

In the background to this comprehensive literature review, it was explained why a traditional review with purposive selection was chosen for the format of this comprehensive review (Sutton et al., 2019). This section will describe the methods used to identify studies included in this review, and how the results from the studies were synthesized.

Approach to Article Selection

This section will explain the search and article selection strategy for this comprehensive review. It will also describe the article inclusion and exclusion criteria.

Search Strategy

The objective of the article selection strategy was to identify all published peer-reviewed studies that met the inclusion and exclusion criteria listed below. Several landmark studies have demonstrated that Google Scholar (GS) is superior to all other scientific databases in the identification of published scientific works (Falagas et al., 2008; Gehanno et al., 2013). Gehanno and colleagues (2013) conducted a study using 29 systematic reviews published in the Cochrane Database of Systematic Reviews. GS was used to replicate each search, and it was found that GS was able to retrieve 100% of the original studies (n=738). (This article should not be confused with an article about the unreliability of the GS report of “number of citations” by Martin-Martin and colleagues (2018).) The comprehensive coverage by GS was enabled because GS is “journal agnostic” in how it builds its search engine, enabling it to include a wider array of publications compared to database indexes, such as PubMed, Scopus, and Web of Science, which may place limitations on what content is included (Falagas et al., 2008; Gehanno et al., 2013). In other words, the way GS works, keywords that would be entered into other databases to identify articles will also work in GS (even if they will not work in another database). As proof of this, all the articles identified by GS for this review were also the

only articles identified when the other databases were searched as well (including PubMed, Web of Knowledge, and EMBASE, as described in the next paragraph).

Therefore, to design an evidence-based search strategy, first GS was searched to identify all potential articles. Next, article selection procedures were applied (see next section). Finally, after this took place, the results from the GS were confirmed through replicating the same in other independent scientific indexes, including PubMed, Web of Knowledge, and EMBASE (searched through the OVID interface). These additional searches failed to identify any additional articles not identified through the GS search.

Search Terms Used

Given that NGS is emerging technology, no limits on publication year were included in searches. In the search in GS and all the other scientific databases, the following search terms were used: OSCC, oral squamous cell carcinoma, oral cancer, cancer of the oral cavity, mouth cancer, cancer of the mouth, microbiome, microbiota, flora, bacteria. In GS, Boolean operators are assumed, but in the other databases, Boolean operators (AND) and (OR) were used to refine and produce more relevant results. Consistent with the scientific literature, even with this extensive search approach, no additional articles were identified in the other databases that were not identified using GS (Falagas et al., 2008; Gehanno et al., 2013). Please note that the results of this search were confirmed in October of 2021 to allow for the review to be conducted.

Article Selection

Results from the initial search were screened by their titles and abstracts. Duplicate abstracts were removed. Studies not on the topic of the oral microbiome using 16s rRNA gene analysis were excluded. The full text for all potentially relevant studies was identified and further assessed for eligibility criteria (described in the next section). Studies that did not fit eligibility criteria were excluded.

Article Inclusion and Exclusion Criteria

Articles returned in the search and later identified in full-text were considered eligible for inclusion if they met the following criteria:

1. They reported on original studies.
2. The authors reported using the NGS technique in the article.
3. The NGS technique was used to target 16s rRNA gene.
4. The article studied OSCC.
5. The article was in English.
6. Articles must include at least one participant group other than OSCC (e.g., control groups or other patient groups)

Articles that fit the above criteria were excluded if they also fit these criteria:

1. The article stated it was a review.
2. The article studied the association of a specific member or members of the bacteriome and others (was hypothesis-driven rather than exploratory).
3. The article reported results of non-human studies.

4. The article did not study OSCC specifically.

Assessment of Article Quality

The methods that can be used to evaluate the quality of an article typically depend upon the study design (Moskalewicz & Oremus, 2020; Semrau et al., 2023). However, after identifying the articles, it was realized that all of them had very low quality study designs (Moskalewicz & Oremus, 2020; Semrau et al., 2023). For that reason, applying structured assessments to the articles identified would have to rely on a framework tailored for exploratory laboratory studies, which was the primary study design of the studies identified.

Frameworks for evaluating other types of studies are available. A framework for this specific method of assessing study quality was not available. Quality assessment frameworks exist for synthesis reports (Bezerra et al., 2022), systematic reviews and mixed-methods studies (Harrison et al., 2021) prevalence studies (Migliavaca et al., 2020), as well as observational study designs in evidence-based medicine (Moskalewicz & Oremus, 2020). Because an appropriate framework could not be identified, articles were categorized by the type of evidence they provided based on the evidence-based medicine pyramid, which has case studies and case series as the lowest level of evidence, with cross-sectional studies being a higher level of evidence (Semrau et al., 2023).

Identifying Outcome Measures

The studies reviewed were not treatment studies, so they did not have defined treatment outcomes. The primary outcomes identified for

each study was whether or not any members of the oral microbiome were identified to be associated with OSCC. The secondary outcomes identified were level of study quality, so that they could be classified and reviewed by class.

Article Classification, Analysis, and Synthesis of Findings

Once identified, articles were classified and grouped as to their study design. The results for each group were synthesized separately. Next, articles were considered in terms of their research aims (discovery or confirmatory), participant groups included, measurement and statistical approaches, and findings reported. As is typical in traditional reviews, the results are presented in subheadings thematically (Sutton et al., 2019).

Data Synthesis Methods

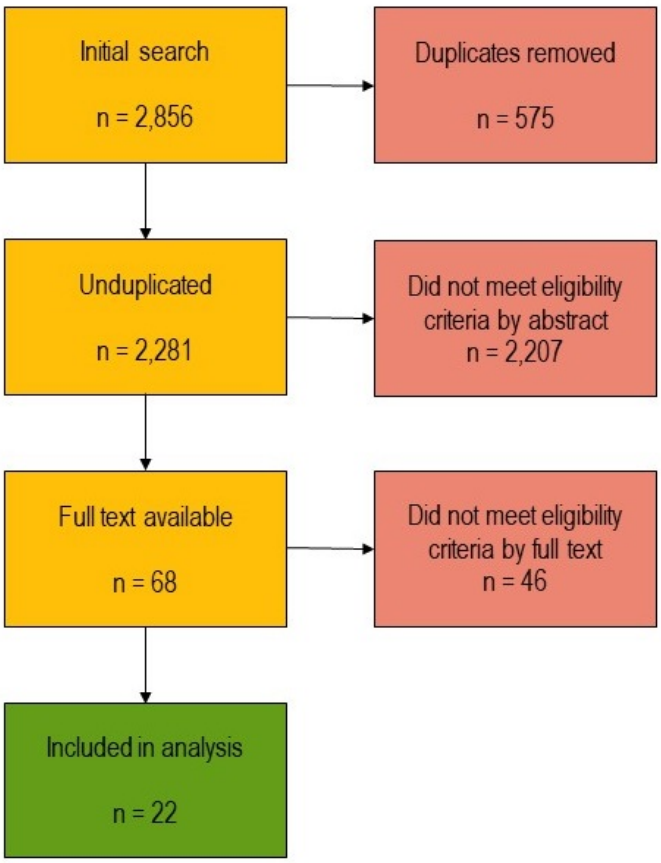
As described previously, articles were grouped by study design, with the lowest quality study design being synthesized and presented first. “Small sample studies” was considered the lowest quality study design, because these studies did not include enough sample to make statistical inferences. For these studies, the study aim and critical limitations were identified and synthesized. The second lowest quality study design was a classification titled “natural history studies”, which were also studies with too small of a sample to provide statistical inference, but with characteristics that may provide insight into the natural history of the oral microbiome in OSCC. As only two studies fell in this category, they were simply described, and their study designs and results compared.

The next highest level of quality was a case series study design, and several were included that compared healthy tissue to diseased tissue in the same individual. Therefore, these were titled “comparative case series” studies, and were synthesized in terms of their study aims, the metrics specified in the studies, whether or not they used the HOMD database as part of a machine learning step, and what was ultimately reported as findings. The next highest level of quality were cross-sectional studies, but since all of them identified had serious study design issues, they were considered “low quality cross-sectional studies”. These were compared in terms of study aim, statistical approach, and findings. The most rigorous category of articles identified includes two “case-control” studies which also had serious study design issues. Since these there were only two, they were described and compared to each other.

2.4. Results

The article selection results are documented in Figure 2.2

Figure 2.2. Article selection results



As described in Figure 2.2, the initial search revealed 2,856 results, 575 of which were eliminated because they were duplicates. Of the 2,281 unduplicated screened abstracts, 2,207 of them were excluded because according to their abstract, they did not meet eligibility criteria. The full-text was examined for the remaining 68. From this review, it was found that 46 did not meet eligibility criteria, leaving 22 articles for the analysis in this comprehensive review.

Table 2.2 summarizes the basic study design considerations, including study quality and findings, for all 22 articles. For ease of interpretation, they were placed in chronological order of year of

publication. Please note that epidemiologic studies with an n < 30 are unlikely to detect even large effects, and when studying compositional data as in the oral microbiome, such small sample sizes are likely to have extremely unstable results (Kers & Saccenti, 2022; Serdar et al., 2021).

Table 2.2. Article summary

Article	Country Participants are From	Study Design	Samples	Study Quality	Findings
Pushalkar, et al. (2011) "Microbial diversity…"	Unclear, but it appears to be United States - New York. It was approved by the Institutional Review Board of New York University.	Cases of OSCC (n=3 over age 50 males smokers and drinkers) and "matched controls" (n=2)	Stimulated saliva	Low quality study due to extremely low sample size and unclear study aim.	Findings: “Members of 8 phyla (divisions) of bacteria were detected. The majority of classified sequences belonged to phyla, Firmicutes (45%) and Bacteroidetes (25%).“
Zhao, et al. (2017) "Variations in oral microbiota…"	China	Case series of OSCC patients in Shanghai, China (n=40, each gave 1 sample of lesion and 1 sample of normal site, so total analysis n=80 samples)	Swabs of oral lesions and normal sites	High quality comparative case series.	The authors claim in their study, they found different taxa than in previous studies.

Article	Country Participants are From	Study Design	Samples	Study Quality	Findings
Mukherjee, et al. (2017) "Bacteriome and mycobiome…"	Unknown, but used a Cleveland IRB, so possibly Ohio in the US	Case series - n= 40: "Fifty-three unrelated patients with mobile tongue cancer undergoing resection were prospectively enrolled (2003-2014). Of these 53, 40 had adequate fresh-frozen specimens (30-50 mg) of matched tumor and adjacent normal tissues…"	Tongue cancer tissue vs. normal tissue	High quality comparative case series - but applicable to tongue cancer only.	Although many statistical analyses using multiple software packages are described, there are no hypotheses posed *a priori*, and all appear to be run as *post hoc* analyses. Nevertheless, the authors claim, "Our results demonstrate differences in bacteriome and mycobiome between oral (mobile) tongue squamous cell carcinoma, and their matched normal oral epithelium, and their association with T-stage."
Al-hebshi, et al. (2017) "Inflammatory bacteriome…"	Yemen/Saudi Arabia - but cases and controls collected at different times and sites	Biopsies from cases of OSCC (n=20) and deep-epithelium swabs from matched controls (n=20)	DNA extracts from previous tissue biopsies vs. deep-epithelial swabs	Low quality study due to low sample size and unclear study aim.	The authors pose no statistical hypotheses, and conduct many descriptive analyses. On that basis, they conclude that genes involved in certain functions are enriched in tumours, while those responsible for other functions are "significantly associated with controls".

Article	Country Participants are From	Study Design	Samples	Study Quality	Findings
Lee, et al. (2017) "Bacterial alterations…"	Taiwan	Cross-sectional study of Taiwanese individuals all evaluated for OSCC, and in 3 groups: Normal (n=127), epithelial precursor lesion (n=124), and cancer (n=125)	Saliva only (no tissue)	High quality cross-sectional natural history study	To evaluate statistical significance, they used multi-response permutation procedures (MRPPs) on measurements of biodiversity, and used statistical tests to compare the three groups. They found that even if there was a statistically-significant difference between groups, the difference was small.
Perera, et al. (2018) "Inflammatory bacteriome…" - based on reanalysis of data from Perera, et al. (2017) "A dysbiotic mycobiome…"	Sri Lanka	Case-control study, where cases were male Sri Lankan patients with histologically-confirmed OSCC (n=25) and the control group was male Sri Lankan patients with a clinical diagnosis of intra-oral fibro-epithelial polyps (FEP, n=27)	Biopsy tissue in OSCC, and FEP tissue in FEP	Medium quality case-control study, as the sample size is low, and findings only applicable to OSCC vs FEP.	This was a reanalysis of data from a previous article (Perera 2017), which focused on the mycobiome, and found "a dysbiotic mycobiome dominated by *C. albicans* was found in association with OSCC". In this study, no hypotheses were posed *a priori,* and many analyses were conducted. On the basis of a largely descriptive analysis, the authors conclude that proinflammatory bacterial attributes were enriched in the OSCC tissues.

Article	Country Participants are From	Study Design	Samples	Study Quality	Findings
Yang, et al. (2018) "Oral microbiota…"	Taiwan	Cross-sectional: Healthy individuals (n=51), OSCC patients (total n = 197, stage 1 n=41, stage 2&3 n = 66, stage 4 n = 90)	Oral rinse	High quality cross-sectional natural history study	The authors speculate on how their findings may relate to the natural history of OSCC.
Takahashi, et al. (2019) "Analysis of oral microbiota…"	Japan	Cross-sectional: OSCC patients (n=60), non-cancer individuals (n=80).	Salivary samples	Cross-sectional analysis lacking epidemiologic context	Authors posed hypotheses about differences in α diversity, β diversity, and abundance between groups overall, but instead engaged in *post hoc* testing, and interpreted the results without correction. They summarize their findings as having observed increased bacterial diversity in OSCC and having found distributional changes for some bacteria.
Chang, et al. (2019) "The prevalence rate…"	China	Cross-sectional study involving OSCC patients (n=61) and healthy controls (n=35)	Tissue samples	Cross-sectional analysis lacking epidemiologic context	Although authors describe their statistical approach, they do not pose any *a priori* hypotheses, and do not describe how they approached a *post hoc* analysis. Hence, any findings are likely due to Type I error. The authors conclude by saying there may be a close relationship between oral microorganisms and OSCC which might enrich its pathogenesis.

Article	Country Participants are From	Study Design	Samples	Study Quality	Findings
Hashimoto, et in. (2019). "Changes in oral…"	Japan	Cross-sectional: OSCC patients (n=6), oral leukoplakia patients (n=6) and healthy controls (n=4).	Unstimulated saliva	Low quality study due to extremely low sample size and unclear study aim.	Authors report differences in abundance of certain microbes, but they do not overinterpret their findings. They suggest that changes in the salivary microbiome may have a potential application as a novel diagnostic tool.
Li, et al. (2020) "Composition and function of oral microbiota…"	China	Cross-sectional: Gingival squamous cell carcinoma (GSCC) patients (n=10), periodontitis patients (n=15), healthy (n=15)	Subgingival plaque and saliva from all, and tongue dorsum, buccal mucosa, cancerous, and para-cancerous tissues from GSCC.	Low quality study due to extremely low sample size and unclear study aim.	The authors claim that because a high level of periodontal pathogens were found in GSCC, there is a need to explore a potential causal relationship.

Article	Country Participants are From	Study Design	Samples	Study Quality	Findings
Zhang, et al. (2020) "The oral microbiota…"	China	Case series of OSCC patients in Shanghai, China (n=50, each gave 1 swab sample of lesion and 1 swab sample of normal site, so total analysis n=100 samples)	Bilateral buccal mucosal tissue swabs	High quality comparative case series	No *a priori* hypotheses were posed, and many statistical analyses were conducted, greatly increasing the risk of Type I error. Although the authors claim that oral bacterial profiles showed a "significant difference" between cancer sites and normal tissue, they likely do not mean it in the statistical sense due to the limitations of their analysis. They recommend further research on the oral bacterial profiles they found as potential diagnostic markers and treatment targets.
Namburi, et al. (2020) "Low abundance of capnophiles…"	India	Cross-sectional: OSCC patients (n=5) and "healthy" patients (n=5)	Unstimulated whole saliva	Low quality study due to extremely low sample size and unclear study aim.	The authors of this study with very low sample, no clear research aim, and no testable hypotheses claim that *Capnophilic Capnocytophaga* species were found in the healthy samples so they are suggested to be associated with health.
Gopinath, et al. (2021) "Differences in the bacteriome…"	India	Case-control: OSCC patients (n=48), those evaluated and found not to have OSCC (n=46)	Tissue, swabs, and saliva	High quality case-control study	Although the sampling was done correctly for this case-control study, no hypotheses were posed *a priori*, and no statistical tests were described. Authors do not draw conclusions about the differences between the

Article	Country Participants are From	Study Design	Samples	Study Quality	Findings
					groups, and instead comment on their methods and within-group descriptive findings.
Gopinath, et al. (2021) "Salvatory bacterial shifts…"	India	Cross-sectional with 3 groups: Oral cancer (n=31), leukoplakia patients (n=20), healthy "molar removal" patients (n=23)	Whole mouth fluid	Low-quality cross-sectional study due to lack of clear study aim and appropriate statistics.	No hypotheses were posed a priori, and the statistical methods are unclear. Authors reported finding both overlapping and discriminative bacterial genera between leukoplakia and oral cancer, and suggest their findings have implications for cancer prevention strategies.
Granato, et al. (2021) "Meta-omics analysis…"	Brazil	Cross-sectional with 3 groups: Healthy controls (n=8), OSCC patients w/o "active lesion" (n=8), OSCC patients w/"active lesion" (n=8)	Saliva	Low quality study due to extremely low sample size and unclear study aim.	Findings are unclear: "In summary, the present study has characterized and compared the microbiome of control individuals and OSCC patients with and without active tumorus. Moreover, it was capable of showing that both OTUs and proteins, from bacteria and from hosts, are associated with clinical characteristics, highlighting once again the dynamics of microbiome and biological components in different conditions."

Article	Country Participants are From	Study Design	Samples	Study Quality	Findings
Sarkar, et al. (2021) "Dysbiosis of oral microbiota…"	India	Case series: OSCC patients (n=50).	Tissue sampled from lesion and uninvolved areas and compared	High quality comparative case series	No *a priori* hypotheses were posed, and many statistical analyses were conducted, greatly increasing the risk of Type I error. Although the authors report that in contrast to previous studies, there was significantly lower bacterial diversity observed in malignant vs. normal samples, they likely do not mean statistical significance due to their lack of formal statistical testing.
Su, et al. (2021) "Oral microbial dysbiosis…"	Taiwan	Case series with two cohorts of OSCC patients - regular and (n=74) "validation" (n=42).	Oral swabs of tumor lesions and their contralateral, normal regions were compared	High quality comparative case series	Authors report statistically significant results for bacterial diversity and relative abundance of specific oral microbiota.
Rai, et al. (2021) "Dysbiosis of salivary microbiome…"	India	Cross-sectional: included OSCC patients (n=11) and "healthy controls" (n=10) for metagenomic analysis	Unstimulated saliva	Low-quality cross-sectional study due to lack of homogeneity in the sample combined with small sample.	Authors report significantly higher abundance of some oral microbiome signatures, and significant elevations of certain cytokines. However, the study likely suffers from Type I error due to interpreting these results as a *post hoc* analysis rather than stating hypotheses *a priori*.

Article	Country Participants are From	Study Design	Samples	Study Quality	Findings
Sawant, et al. (2021) "Identification and correlation…"	India	Although authors call this a case-control study, it is actually cross-sectional in design: Tobacco chewers without OSCC (n=10), OSCC patients (n=10), healthy controls without OSCC who do not chew tobacco (n=10)	Oral rinse samples	Low-quality cross-sectional study due to lack of homogeneity in the sample combined with small sample.	Authors claim that study results provide evidence of oral bacterial dysbiosis due to tobacco chewing habits, but the sample is too small to make a causal inference.
Zhou, et al. (2021) "The clinical potential…"	China	Cross-sectional: OSCC patients (n=47) and "healthy controls" (n=48)	Salivary samples, subgingival plaque samples, tumor surface samples, healthy mucosa samples, and tumor tissue samples.	Low-quality cross-sectional study due to lack of clear study aim and appropriate statistics.	Author propose to use microbiota as a screening tool for OSCC, but do not propose a testable screening approach. No statistical tests are performed, and authors suggest that the high diagnostic accuracy rates found in their study could be replicated in practice.
Chen, et al. (2021) "Taxonomic and functional…"	Taiwan	Cross-sectional (all male): "Healthy controls" (n=27), non-recurrent oral verrucous hyperplasia (OVH) patients with no OSCC at follow-up (n=21), and OVH patients with OSCC at follow-up (n=27).	Saliva samples	Low-quality cross-sectional study due to lack of clear study aim and appropriate statistics.	Authors do not describe an approach to statistical testing that can be supported with such low sample. The authors promote the use of predicted functional profiles over using taxonomic data to make epidemiologic inferences with patient characteristics, but do not provide strong evidence for this.

As shown in Table 2.2, many of the studies reviewed had study quality issues. These will be discussed in detail later in this chapter. Studies were classified by quality levels, and will be reviewed in order. Eight of the studies had such a small sample size for each group that the study results could not be interpreted (Granato et al., 2021; Hashimoto et al., 2019; Li et al., 2020; Namburi et al., 2020; Pushalkar et al., 2011; Rai et al., 2021; Sawant et al., 2021). As described earlier, epidemiologic studies with $n < 30$ are unlikely to detect effects unless they are large, and when considering compositional data as with the oral microbiome, the sampling variability will be higher, and the likelihood of replicating findings with this amount of sample would be very low (Serdar et al., 2021). These studies will be reviewed together. Next, two of the studies had designs that implied they were exploring the natural history of OSCC; these studies will be reviewed together (Lee et al., 2017; C.-Y. Yang et al., 2018). A third category containing five studies could be considered comparative case series studies, where healthy tissue is compared with diseased tissue in each participant in each group, and results interpreted (Mukherjee et al., 2017; Sarkar et al., 2021; Su et al., 2021; Zhang et al., 2020; Zhao et al., 2017). The articles in this category will be examined together.

A fourth category included five articles reporting the results of cross-sectional studies including OSCC and other groups (Chang et al., 2019; J.-W. Chen et al., 2021; Gopinath, Kunnath Menon, et al., 2021; Takahashi et al., 2019; Zhou et al., 2021). These studies will be reviewed together and their findings compared. Finally, at the highest level of

study quality, two case-control studies were identified, and will be reviewed together (Gopinath, Menon, et al., 2021; Perera et al., 2018).

Small Sample Studies

As described in Table 2.2, of the 22 studies included in this review, eight were identified as having too low of a sample size to be able to provide conclusions about biomarkers (Al-Hebshi et al., 2017; Granato et al., 2021; Hashimoto et al., 2019; Li et al., 2020; Namburi et al., 2020; Pushalkar et al., 2011; Rai et al., 2021; Sawant et al., 2021). These studies were conducted between 2011 and 2021, and appear to have involved populations from the United States (US), Yemen and southern of Saudi Arabia, Japan, China, India and Brazil (Al-Hebshi et al., 2017; Granato et al., 2021; Hashimoto et al., 2019; Li et al., 2020; Namburi et al., 2020; Pushalkar et al., 2011; Rai et al., 2021; Sawant et al., 2021).

All of these studies included multiple groups, which unfortunately increased the participant-contributed variation in their sample, and limited their statistical power. When Considering the largest studies in this group, it is observed that there is one by Rai and colleagues (2021), which included 25 OSCC patients (where only 11 participants of this group were included for metagenomic analysis) and 24 “healthy controls” (where only 10 of the participants were included as controls for the metagenomic analysis) in India, and one by Al-hebshi et al. (2017), which included biopsies from 20 cases of OSCC in Yemen that were matched to 20 controls recruited from a different environment than the cases. Both of these studies included groups of non-homogenous

participants, and neither put forth *a priori* hypotheses before conducting biomarker analyses (Al-Hebshi et al., 2017; Rai et al., 2021). In Rai et al. (2021), “healthy controls” were not evaluated for OSCC to confirm their health, and Al-hebshi and colleagues (2017) described using “matched” controls, which would likely introduce bias from the matching process.

Table 2.3 presents a summary of these small sample studies.

Table 2.3. Summary of aims and critical limitations of small sample studies

Article	Study Aim	Critical Limitation
Pushalkar, et al. (2011) "Microbial diversity…"	Evaluate diversity and relative abundance	This study included a total n=5, which is not enough to make inferences about diversity or relative abundance measures.
Al-hebshi, et al. (2017) "Inflammatory bacteriome…"	Characterize the species composition and functional potential of bacteriome	With only 20 samples in each group (OSCC and comparison), it would not be possible to accurately characterize the species composition and functional potential of bacteriome, as even the samples taken would be subject to sampling error within the tissue sampled.
Hashimoto, et in. (2019). "Changes in oral…"	Identify specific oral microbial profiles associated with OSCC	With this sample size, it would be difficult to draw conclusions about one biomarker testing an *a priori* hypothesis. To determine an entire biomarker profile would require a large observational study.

Article	Study Aim	Critical Limitation
Li, et al. (2020) "Composition and function of oral microbiota…"	Exploratory study of microbial composition and functions in periodontitis and gingival squamous cell carcinoma	Difficult to draw conclusions due to the diversity of disease processes operating in the participants included in the study
Namburi, et al. (2020) "Low abundance of capnophiles…"	Assess alterations in the microbiome linked to OSCC	This study included a total of n=10 divided into two groups, those with OSCC (n=5) and so-called "healthy patients". Due to utilizing this small of a sample and not using statistical hypotheses, it is likely these authors' conclusion is merely a reflection of Type I error.
Granato, et al. (2021) "Meta-omics analysis…"	Investigate the capability of using metagenomic and metaproteomic saliva profiles to distinguish between healthy patients and those at various OSCC stages	While the study aim is reasonable, including only 8 participants in each of three groups will not provide a large enough sample to enable the researchers to draw conclusions without a presupposed statistical hypothesis.
Rai, et al. (2021) "Dysbiosis of salivary microbiome…"	Develop saliva-based oral microbiome and cytokine biomarker panel for screening OSCC patients	To develop a screening approach, each group would need much more than n=25 (so as to enable multivariate statistics), and each group would need to be more homogenous.
Sawant, et al. (2021) "Identification and correlation…"	Identify and correlate bacterial diversity in the oral cavity of tobacco chewers, patients with oral cancer, and	To achieve this research aim, many more than 10 participants are needed in each group, and *a priori* hypotheses need to be presupposed.

Article	Study Aim	Critical Limitation
	healthy subjects in the Indian population.	

As can be seen in Table 2.3, these eight studies had a variety of aims. Two studies had aims that were focused on evaluating measures of diversity and relative abundance specifically (Pushalkar et al., 2011; Sawant et al., 2021), while three studies proposed descriptive aims, such as characterizing microbial composition, functional potential, and alterations in the bacteriome (Al-Hebshi et al., 2017; Li et al., 2020; Namburi et al., 2020). The remaining three studies proposed identifying entire profiles of biomarkers related to the oral bacteriome that could be used for diagnostic purposes (Granato et al., 2021; Hashimoto et al., 2019; Rai et al., 2021). These studies will be reviewed in terms of their aims, findings and critical limitations in study design as listed in Table 2.3.

Small Sample Studies Evaluating Diversity and Abundance

Among the small sample studies, only two included the measurements of diversity and/or abundance in their research aims – the earliest article included in the review authored by Pushalkar and colleagues (2011), and a study by Sawant et al. (2021) published more recently.

In the first article, which included only five participants (in two groups, see Table 2.3), the authors went on to describe how bacterial

DNA extraction took place from the saliva samples, followed by all the steps they took to develop OTUs (Pushalkar et al., 2011). It was not clear from the methods section as to how the authors intended to operationalize the diversity and relative abundance measurements. The authors suggest that the abundance-based coverage estimators ACE and Chao1 were used, but this was not made explicit, and how differential abundance bacteria was measured was also not made explicit (Pushalkar et al., 2011). In the results and discussion section, it appears that the number of detectable bands identified through the denaturing gradient gel electrophoresis (DGGE) was used as a diversity measure (Pushalkar et al., 2011). The authors present visualizations and tables of their results as exploratory, and do not attempt to draw conclusions on a statistical basis (Pushalkar et al., 2011). They provide raw output from their analysis, identifying clusters by number (e.g., I and II) (Pushalkar et al., 2011).

In the results and discussion sections, the authors claim to be pursuing an oral microbiome “profile” that would theoretically differentiate healthy patients from those with OSCC (Pushalkar et al., 2011). To be clear, although mentioned in their research aims, measurements of diversity and abundance are not explicitly addressed in the analysis; instead, visualizations of their results are presented more as an argument to support the profiles they identified in their analysis (Pushalkar et al., 2011). To support their identification of profiles, they present a bar chart comparing the relative distribution of both phyla and genera in the OSCC samples compared to control (Pushalkar et al.,

2011). The bars exhibit no more than 20% difference between the groups, , and no statistical tests were used (Pushalkar et al., 2011). Further, a Venn diagram of the profiles was presented, but it also lacks statistical rigor and does not visually illustrate the obvious emergence of two differentiated profiles (Pushalkar et al., 2011). Due to this study lacking enough sample and authors not using appropriate statistical approaches, the design of the study was insufficient to make such claims which were likely due to Type I error. (Gloor et al., 2017; Leung, 2011; Mandal et al., 2015; Pushalkar et al., 2011).

The second small sample study included in this review aimed to research the diversity of the oral microbiome in patients with OSSC, long-term tobacco chewers, and non-tobacco-chewing healthy patients, but they only included samples from ten people in each group (Sawant et al., 2021). To the credit of the authors, they describe it accurately as a “pilot” study, but the only way such an underpowered study could ever show usable results on which to base the design of a larger study would be if *a priori* hypotheses were posed, and diversity and abundance metrics were clearly operationalized as part of the study design (Kers & Saccenti, 2022; Leung, 2011; Mandal et al., 2015; Sawant et al., 2021). Instead, the authors posed no hypotheses, and described running multiple α and β diversity indexes without choosing a metric of record – essentially admitting they were “fishing” for a significant p-value (Kers & Saccenti, 2022; Leung, 2011). The authors also fail to choose one metric for abundance or relative abundance, and choose to run both a Kruskal-

Wallis and a Wilcoxon test on LDA effect size (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011; Sawant et al., 2021).

The previous paragraphs described studies lacking *a priori* statistical testing that also exhibit challenges with sampling and issues with applying statistical approaches. This combination of these study design features, including the lack of selection of metrics and statistical tests *a priori*, lack of stated *a priori* hypotheses, multiple heterogeneous participant groups with extremely small sample, and the appearance of "fishing" make it so that results cannot be interpreted according to the scientific method (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011; Sawant et al., 2021). As a result, the authors are forced to cherry-pick and report certain particularly interesting results from their analyses, such as their finding that the genus *Streptococcus* dominated the control group, with less abundance in the tobacco and OSCC groups (Sawant et al., 2021). With lack of rigorous study design, it is not possible to tell if this is a real finding, or a chance finding (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011; Sawant et al., 2021).

Small Sample Studies with Descriptive Aims

Three of the small sample studies in this review included descriptive aims (Al-Hebshi et al., 2017; Li et al., 2020; Namburi et al., 2020). The aim of the first of these studies was to characterize the species composition and functional potential of the bacteriome associated with OSCC, but to investigate this, the authors only included 20 OSCC biopsies, and 20 deep-epithelium swabs from "matched controls" (Al-Hebshi et al., 2017). While the aim is reasonable, in order

to characterize the species composition and functional potential of the bacteriome associated with OSCC, all of the participants in the study should have been OSCC patients in order to reduce heterogeneity and increase the likelihood of seeing an effect (Hu & Dignam, 2019). Further, all of them should have had a well-characterized diagnostic entity – meaning all of them should have undergone a diagnostic evaluation for OSCC according to a research protocol. Also, the data collected should be promoted as providing the basis for extrapolation about a particular population, the way the NHANES analysis can provide a characterization of the oral microbiome in US residents (Hu & Dignam, 2019; Vogtmann et al., 2023). That is because government agencies fund oncology studies for the purposes of improving the health of the population of their countries, so it is incumbent upon researchers receiving this funding should steward these funds for greater public health impact (NIH Human Microbiome Portfolio Analysis Team, 2019).

While descriptive analyses present no obligations to engage in statistical testing, without a clear description of how the authors intended to characterize the species composition and functional potential of the oral bacteriome in patients in their study with OSCC, it is difficult to understand what the authors found (Al-Hebshi et al., 2017; Leung, 2011). While the authors presented many visualizations of their results, these images do not tell a compelling story when presented together, and the authors are reduced to simply listing the abundant taxa found in their analysis as their findings (Al-Hebshi et al., 2017; Gloor et al., 2017; Leung, 2011). This lack of scientific rigor in study design and reporting

suggests that multiple studies done with the same study design on the same background population would render widely different findings, in which case, the reported results would be due to Type I error (Al-Hebshi et al., 2017; Gloor et al., 2017; Hu & Dignam, 2019; Kers & Saccenti, 2022; Leung, 2011). Mainly, this is due to the authors not accommodating the compositional nature of the data (Gloor et al., 2017).

The second of the three small sample studies with descriptive aims included in this review suffers from many of the same issues as the first one (Li et al., 2020). As stated in the article, the aim of the study was to explore the microbial composition and functions in periodontitis and gingival squamous cell carcinoma (GSCC) (Li et al., 2020). Even from the study aim, it is apparent that in order to characterize the oral microbiome in two completely different disease entities, a very large sample of at least 30 if not 100 homogenous patients must be drawn (Hu & Dignam, 2019; Serdar et al., 2021). Yet, this study included three groups of patients, with only ten participants in one group, and 15 in each of the two other groups (Hu & Dignam, 2019; Li et al., 2020). Further, the authors describe applying multiple *post hoc* tests of diversity and other indices without a stated *a priori* hypothesis, rendering their results uninterpretable (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011; Li et al., 2020). As with the last article, the authors are reduced to cherry-picking and reporting results that they find interesting, such as their observation of *Atopobium* (a gram-positive bacteria that was found to colonize periodontal abscess and plaque) which was the most abundant in GSCC saliva and plaque as compared to periodontitis followed by

controls.(Kers & Saccenti, 2022; Leung, 2011; Li et al., 2020). While these observations are interesting, they should be treated with caution as the lack of rigorous study design and the high likelihood of Type I error would likely lead to unreproducible findings (Kers & Saccenti, 2022; Leung, 2011; Li et al., 2020).

The third small sample study with descriptive aims reviewed in this comprehensive review actually exemplifies and illustrates the danger of not using statistical rigor when designing studies of the microbiome (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011; Namburi et al., 2020). While this study appears to be the first to use MinION nanopore in sequencing 16s rRNA gene of OSCC – a technology allowing the researcher to sequence longer reads as compared to Illumina Miseq, it is still expected to follow a scientific rigorous design (Namburi et al., 2020). The authors in this study confidently reported their findings – that there was a low abundance of *Capnophilic Capnocytophaga* species in OSCC patients compared to healthy patients – based on the findings from ten patients placed in two groups, with no statistical hypotheses posed, nor any metrics operationalized (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011; Namburi et al., 2020). It is completely unclear if this finding was random due to Type I error, or a finding that accurately characterizes the OSCC microbiome in the background population of individuals the authors were studying, which are Indian OSCC patients (Gloor et al., 2017; Leung, 2011; Namburi et al., 2020). Nevertheless, due to severe issues with sampling, these study results do not provide any usable evidence related

to the human oral microbiome in general relating to OSCC (Gloor et al., 2017; Leung, 2011; Namburi et al., 2020).

Small Sample Studies Aiming to Develop Diagnostic Biomarker Profiles

Three of the eight small sample studies included in this review sought to develop diagnostic biomarker profiles (Granato et al., 2021; Hashimoto et al., 2019; Rai et al., 2021). In the first article, Granato and colleagues (2021) conducted a study aiming to determine how a unique biomarker profile might differentiate between OSCC patients without an active lesion, OSCC patients with an active lesion, and healthy controls. This would be challenging, as described earlier, since epidemiologic studies to address cancer (such OSCC) utilize the concepts of “target biomarker” and “companion biomarkers” (Hemminki et al., 2020; Hu & Dignam, 2019). Therefore, seeking simply a profile would not be useful for cancer research, and from a statistical standpoint, it would be hard to demonstrate that a unique profile that differentiates one patient group from another could be identified, as so many biomarkers are correlated (Hemminki et al., 2020; Hu & Dignam, 2019).

Further, the authors only included eight participants in each group, did not state any *a priori* hypotheses, utilized the HOMD for data profiling, and employed unusual and undefended statistical approaches, severely limiting the utility of the data they collected (Gloor et al., 2017; Granato et al., 2021; Kerr, 2016; Leung, 2011). For example, one of the analysis they did includes a linear regression equation which contained covariates as independent variables at two hierarchical levels of order:

Lesion type and tumour site (level one) and participant age (level two) (Granato et al., 2021). Such a regression analysis would not be appropriate without hierarchical modelling (Granato et al., 2021; Grantham et al., 2020; Kerr, 2016). Similarly, other analyses were described without any hypotheses, and as with the linear regression model described earlier, many of these statistics would be considered inappropriate to use on such small samples of compositional data (Gloor et al., 2017; Granato et al., 2021; Kerr, 2016; Leung, 2011). In summary, the study design and statistical approach in this study had many errors, and this precludes interpretation (Gloor et al., 2017; Granato et al., 2021; Kerr, 2016; Leung, 2011). Although the authors proposed that genus *Alloscordovia* was highly abundant in active OSCC group as compared to inactive OSCC samples while *Veillonella* was lowered in the active OSCC group as compared to the inactive group. In addition the authors suggested that *Centipeda*, *Veillonella* and *Gamella* to be associated the activity, size, and the clinical stage of the tumour, these findings should be interpreted carefully due to the estimates likely being unstable (Gloor et al., 2017; Granato et al., 2021; Leung, 2011).

The second of the small sample biomarker profile studies included in this comprehensive review had a very similar design and approach as the one just described, so it carries with it a lot of the same methodologic errors (Gloor et al., 2017; Hashimoto et al., 2019; Leung, 2011). As with the previous study, these authors included three groups and only a total of 16 individuals: Six having oral leukoplakia, six having OSCC, and four healthy controls (Hashimoto et al., 2019). With groups

with sample sizes this small, it would have been unlikely for the authors to find statistically significant differences between groups on any metric, and the statistical power needed to build an entire evidence-based replicable biomarker profile would have definitely have been out-of-reach to the authors (Gloor et al., 2017; Hashimoto et al., 2019; Hu & Dignam, 2019; Kerr, 2016; Leung, 2011). Consequently, the results they report – such as the abundance of the genus *Solobacterium* as being higher in OSCC compared to the oral leukoplakia group – requires further confirmation as the result might be due to Type I error (Gloor et al., 2017; Hashimoto et al., 2019; Kerr, 2016; Leung, 2011). It would be extremely unlikely for this author group to be able replicate this study and achieve the same results, so the lack of rigor in the study design and statistical approaches preclude its utility for targeting cancer biomarkers or developing profiles (Gloor et al., 2017; Hashimoto et al., 2019; Kerr, 2016; Leung, 2011).

The third small sample study seeking to define biomarker profiles included in this review had a relatively larger sample in each group, with 25 OSCC patients and 24 healthy controls (Rai et al., 2021). However, of these patients, only 11 patients of the OSCC groups were recruited for the metagenomic analysis. This level of sample is still much too small to support the development of an entire unique diagnostic profile that differentiates the groups, especially without defining target and companion biomarkers of interest *a priori* (Gloor et al., 2017; Hu & Dignam, 2019; Leung, 2011; Rai et al., 2021). These authors took a similar approach as the others, and reported interesting observations

they made resulting from their studies of microbiome signatures and bacterial abundance, such as higher abundance of certain taxa (Table 2.2) (Rai et al., 2021). Again, because these results are not replicable, a study with these design and statistical flaws cannot contribute to the understanding of the microbiome in OSCC (Gloor et al., 2017; Hu & Dignam, 2019; Kerr, 2016; Leung, 2011; Rai et al., 2021).

Natural History Studies

Although the small sample studies reviewed in the last section lacked rigor leading to results that were not robust, using small samples in studying the natural history of a lesion, especially in one patient or a homogenous group of patients, can yield useful results (Sepich-Poore et al., 2021). In the first of these two studies, Lee and colleagues (2017) used a cross-sectional design of participants who all had been evaluated for OSCC, and placed them in three groups: normal, epithelial precursor lesion, and OSCC. The authors applied appropriate statistical techniques, declaring they were only using species richness and the Shannon Index to evaluate α diversity, and they used the UniFrac distance as a metric for β diversity (evaluated using the Wilcoxon rank-sum test) (Lee et al., 2017).

Even with the high level of statistical rigor in this study, the results were complex reflecting the complexity of the oral microbiome (Lee et al., 2017). Visualizations of results demonstrated that biomarker profiles in the three different groups each contained a complex network of microbiomes (Lee et al., 2017). The authors were able to conclude that five genera - *Bacillus, Enterococcus, Parvimonas, Peptostreptococcus*,

and *Slackia* – had statistically significantly different metrics when compared between the epithelial precursor lesion group and the OSCC group (Lee et al., 2017). These results are useful, in that they could form the basis of future studies of target and companion biomarkers that might be prognostic or predictive (Hu & Dignam, 2019). However, it is important to note that these findings at this point only apply to the background cohort from which the sample was drawn, which was in Taiwan (Lee et al., 2017). Even so, by utilizing rigorous research approaches, the results of this study can contribute useful information to those studying biomarker-related therapies for OSCC (Hu & Dignam, 2019; Lee et al., 2017).

The second of the two articles included in this review with study designs aimed to elucidate the natural history of the OSCC microbiome had many similarities in design to the first article (C.-Y. Yang et al., 2018). In their study, the authors included a group of healthy individuals, and a group of OSCC patients who had been evaluated and classified as Stage I, Stage 2 and 3, or Stage 4 (C.-Y. Yang et al., 2018). However, unlike with the previous study, these authors did not use an appropriate statistical approach where metrics for comparison were defined *a priori,* and instead, just compared all the metrics without regard to a hypothesis testing framework (C.-Y. Yang et al., 2018). The authors also used the LEfSe analysis comparing differential abundant taxa which does not take the compositionality of the data into account. In addition, the authors developed receiver-operator curves (ROC) and reported area under the curve (AUC) without apparently realizing that since these biomarkers are

all correlated, such an approach will not provide results that can be interpreted (C.-Y. Yang et al., 2018). The reason using such tightly correlated biomarkers will not produce a usable predictive model is due to multi- collinearity (Ogoke, 2023). Logistic regression was also performed without a description or rationale for the equation (C.-Y. Yang et al., 2018).

In the results section, the authors found *Fusobacterium periodonticum*, *Parvimonas micra*, *Streptococcus constellatus*, *Haemophilus influenza*, and *Filifactor alocis* to be associated with OSCC, and they were observed to progressively increase in abundance from stage 1 to stage 4 (Granato et al., 2021; C.-Y. Yang et al., 2018). As with many other author groups making similar study design and biostatistical errors, the results from the current article cannot be used as conclusive scientific evidence (Gloor et al., 2017; Granato et al., 2021; Leung, 2011; C.-Y. Yang et al., 2018). This can immediately be contrasted with the utility and clarity of the other paper reviewed in this section, which identified five genera that were statistically different between groups (Lee et al., 2017). Due to the poor selections made in the study design and statistical approaches, the results from the current article cannot be used on a scientific basis (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011; C.-Y. Yang et al., 2018).

Comparative Case Series Studies

Five of the studies that will be reviewed here were classified as a comparative case series in design, because they included analyses where healthy tissue and diseased tissue were compared within the

same individual (Mukherjee et al., 2017; Sarkar et al., 2021; Su et al., 2021; Zhang et al., 2020; Zhao et al., 2017). These study designs are compelling, because they can shed light on the natural history of the tissue within the individual; however, they are limited in making inferences to background population groups (Semrau et al., 2023).

Table 2.4 summarizes the studies in terms of study aims, metrics specified, use of the HOMD, and findings. “Metrics specified” refers to biostatistics, where the metrics that will be tested are stated prior to data collection (*a priori*, according to the scientific method). If metrics to be tested are not specified, and a descriptive analysis is completed, there are no issues. However, if the scientific method is applied, a test statistic calculated, and a *p* value declared and interpreted. Therefore, the reporting of *p* values implies the scientific method has been used. If metrics are not pre-specified and *p* values are being calculated and presented, it is evidence that there has been an error in the application of the scientific method. In the case of multiple post hoc analyses being conducted without adjustment of the *p* value before interpretation, this is called “fishing”. This should be contrasted with merely reporting descriptive statistics without *p* values. Using the HOMD in ML models is also included in the table because it is risky as described earlier. The reason it is risky is that the HOMD lacks the necessary metadata to support its utilization in ML models, because it is unclear if it is appropriate training data. Only one of them used a discernible biostatistical approach, so findings in four of the five studies are unclear.

Table 2.4. Characteristics of comparative case series studies

Article	Study Aim	Metrics Specified	Use of HOMD in ML	Findings
Zhao, et al. (2017) "Variations in oral microbiota …"	To unravel the connections underlying oral bacterial dysbiosis and OSCC.	None, and evidence of fishing	Yes	Unclear
Mukherjee, et al. (2017) "Bacteriome and mycobiome …"	To explore the bacteriome and mycobiome in mobile tongue cancers.	None, and evidence of fishing	No	Unclear
Zhang, et al. (2020) "The oral microbiota …"	To determine the characteristics of oral microflora on OSCC tumour sites.	None, and evidence of fishing	No	Unclear
Sarkar, et al. (2021) "Dysbiosis of oral microbiota …"	To establish the association of bacterial dysbiosis and OSCC among the Indian population	Either there are no metrics specified, or this was intended as a descriptive analysis	No	Unclear
Su, et al. (2021) "Oral microbial dysbiosis… "	To characterize the disturbances in the oral microbial population mainly due to oral tumorigenicity	Shannon Index for alpha diversity, UniFrac for beta diversity.	No	"Significant alterations in the bacterial diversity and relative abundance of specific oral microbiota (most profoundly, an enrichment for genus Fusobacterium and the loss of genus

Article	Study Aim	Metrics Specified	Use of HOMD in ML	Findings
				Streptococcus in the tumour sites) were identified."

As described in Table 2.4, due to a lack of rigor in study design or the application of biostatistics, the results for four of these five studies was unclear. In all of these five studies, because individual participants are contributing more than one sample to the study, statistical methods need to account for and adjust for this individual variation (Grantham et al., 2020). Without taking this into account statistics, it would be very hard to make inferences about how healthy tissue might have evolved into diseased tissue in these individuals.

Studies Published in 2017

The earliest of the two studies were done by Zhao and colleagues (2017) in a Chinese cohort, and Mukherjee et al. (2017) in individuals from an unnamed country (speculated to be the US). Because of lack of scientific rigor, it is difficult to interpret the results of these two studies. In the Chinese study, the aim was to “unravel the connections underlying dysbiosis and OSCC”, which seems to presume that dysbiosis in OSCC is considered accepted science (Zhao et al., 2017). The authors describe many analyses, acknowledge the limitations of using the HOMD for

labelling, and say they calculated many statistics (Zhao et al., 2017). They also present many visualizations, but without using statistical inference or posing any *a priori* analysis, it would be difficult to address a research aim like this (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011; Zhao et al., 2017). The authors appear to be intending to study the degree to which the oral bacteriome becomes dysbiotic when transitioning from healthy tissue to OSCC, but the study design is not set up to answer such as research question (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011; Zhao et al., 2017). The authors reported that they identified many different taxa than previous studies (see Table 2.4), and this will likely keep happening as long as studies refrain from using appropriate statistical inference methods for answering research questions (Gloor et al., 2017; Leung, 2011; Zhao et al., 2017).

In the Mukherjee et al. (2017) study, the aim was to "explore the bacteriome and mycobiome" in mobile tongue cancers (see Table 2.4). Because the authors took pairs of samples of tumour and adjacent normal tissue from their participants who all had tongue cancers, the authors seem to intend to study how the bacteriome and mycobiome change when transitioning from healthy to disease tissue in the same individual (Mukherjee et al., 2017). The authors include a "statistics and bioinformatics" section, where they explain the metrics they intended to use to represent various constructs (such as diversity and relative abundance) (Mukherjee et al., 2017). However, for each of these metrics they test, they failed to state *a priori* hypothesis (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011; Mukherjee et al., 2017). While the

authors reported findings that might be of interest (such as the existence of *Rothia mucilaginosa* at a higher rate in the tumour group as compared to controls), the direct comparison of taxa proportions is not appropriate when dealing with microbiome data (Gloor et al., 2017). Because of the previously mentioned issues, the results are not considered robust and might be attributable to Type I error (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011; Mukherjee et al., 2017).

Studies of Microbial Imbalance/Dysbiosis

The next study listed in Table 2.4 after these two early studies was similar in design to the first one, in that bilateral buccal mucosal swab samples were analysed in 50 Chinese participants, with each participant contributing a sample of diseased tissue and a sample of healthy tissue (Zhang et al., 2020). The authors explained that they believed "imbalances between microbes and their hosts could lead to OSSC", which could be interpreted as another way of saying that the microbiome may be critical to the development of OSCC lesions, and this is the impetus for the study (Zhang et al., 2020). In their reporting in the "statistics and bioinformatics" section, the authors describe running many statistical tests on many different measurements, which suggests "fishing" and cannot provide interpretable results (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011; Zhang et al., 2020). The authors reported identifying abundances of *Fusobacterium Nucleatum*, *Prevotella intermedia*, *Aggregatibacter segnis*, *Capnocytophaga leadbetteri*, *Peptostreptococcus stomatitis*, in addition to another five species to be significantly increased in OSCC, suggesting their potential

association with OSCC. While the authors reported differential abundant taxa between lesional and non-lesional swabs, and they mentioned using the STAMP software package, it was not clear which statistical test was used. Hence, without rigorous study design and using statistical inference, it is not clear if these findings are due to chance or are actually statistically significant (Gloor et al., 2017; Hu & Dignam, 2019; Leung, 2011; Zhang et al., 2020).

The next study in this category, by Sarkar and colleagues (2021), aimed to "establish the association of bacterial dysbiosis and OSCC" among the Indian population. This aim is similarly-worded to the one posed by Zhao and colleagues (2017), which implied that bacterial dysbiosis in OSCC is accepted science (Sarkar et al., 2021). However, from the wording of the aim, it is not clear what the authors intended as a study design or analytic approach (Sarkar et al., 2021). The design of this study was similar to the others in this category, where malignant lesions and adjacent normal tissues were sampled within the same individual (Sarkar et al., 2021). The statistical reporting was also similar, where the authors did not pose any *a priori* hypotheses, yet described the generation and comparison of multiple metrics with many statistical tests, suggesting "fishing" (Gloor et al., 2017; Leung, 2011; Sarkar et al., 2021).

Finally, the last of the five studies listed in Table 2.4 is different, in that while the study design was similar to the other four studies, the statistical approach was acceptable, and the results can be interpreted (Su et al., 2021). Like the previous two studies, the study aim had to do

with understanding disturbances in the oral microbial population due to oral tumorigenicity (Su et al., 2021). First, it is important to note that this study included a comparatively large sample of participants (Su et al., 2021). Authors first sampled oral swabs of tumour lesions and their contralateral normal regions in 74 OSCC patients as one cohort, then completed the same data collection effort subsequently in what they termed a “validation cohort” of 42 participants (Su et al., 2021). The statistical analysis section of this article explains exactly how the authors approached this analysis, including the metrics they used, and the comparisons they conducted (Su et al., 2021). Further, they indicated that they used the knowledge they gleaned from analysing the data from the initial cohort to test predictions they made in the validation cohort about the microbiome (Su et al., 2021). This suggests that this is a rigorous study, and its results should be interpreted (Gloor et al., 2017; Leung, 2011).

The background population for this study was in Taiwan, so the results that the authors report should apply directly to this population (Hu & Dignam, 2019; Su et al., 2021). Comparing OSSC tissue to normal, the authors found significant enrichment of genus *Fusobacterium* and loss of genus *Streptococcus* in tumour sites (Su et al., 2021). From their functional predictions, they found that microbial genes related to the metabolism of terpenoids and polyketides were significantly enriched differently in tumour vs. control (Su et al., 2021). Although the statistical test used for the previous comparison was criticized in the literature, because it does not take into account the compositional nature of the

data, and therefore increases the risk of false positive results when applied with microbiome data, all these findings were observed to be replicated in the validation cohort, suggesting they were robust (Gloor et al., 2017; Su et al., 2021).

To summarize this section, as a generalization, all five studies in this category were aiming to better understand how the microbiome is disturbed as healthy tissue transitions to diseased tissue (Mukherjee et al., 2017; Sarkar et al., 2021; Su et al., 2021; Zhang et al., 2020; Zhao et al., 2017). Positive features of all of these studies is that they were conducted in specific populations, had a large enough sample in each participant group to provide for statistical inferences, and were careful, deliberate and transparent in their measurements (Mukherjee et al., 2017; Sarkar et al., 2021; Su et al., 2021; Zhang et al., 2020; Zhao et al., 2017).

However, all but one study in this section failed to provide usable scientific evidence because of the study's lack of statistical rigor (Gloor et al., 2017; Leung, 2011; Mukherjee et al., 2017; Sarkar et al., 2021; Su et al., 2021; Zhang et al., 2020; Zhao et al., 2017). While any data collection effort will produce tabulatable and graphable results, without the use of statistical inference, it is not possible to know which results to trust as likely correct, and which ones to conclude are likely due to chance (Gloor et al., 2017; Kerr, 2016; Leung, 2011; Mandal et al., 2015).

Low Quality Cross-Sectional Studies

The five low quality cross-sectional studies included in this review took place between 2019 and 2021, long after 16s rRNA gene NGS was well-adopted in the oral health research field (Chang et al., 2019; J.-W. Chen et al., 2021; Gopinath, Kunnath Menon, et al., 2021; Takahashi et al., 2019; Zhou et al., 2021). The groups included in these cross-sectional studies had large enough sample sizes to enable the authors to make statistical inferences (Chang et al., 2019; J.-W. Chen et al., 2021; Gopinath, Kunnath Menon, et al., 2021; Takahashi et al., 2019; Zhou et al., 2021). Nevertheless, many had significant study design flaws (see Table 2.5).

Table 2.5 Cross-sectional studies

Article	Study Aim	Statistical Approach	Findings
Takahashi, et al. (2019) "Analysis of oral microbiota…"	To clarify the relationship between oral cancer and oral microbiota in Japanese patients.	Evidence of "fishing"	Authors selectively present statistical findings in the abstract
Chang, et al. (2019) "The prevalence rate…"	To determine whether periodontal pathogens may have a role in oral cancer development	Evidence of "fishing"	Authors report relative abundance of periodontal pathogens, but unclear if this is due to chance.
Gopinath, et al. (2021) "Salvatory bacterial shifts…"	Study aim was not stated clearly	Evidence of "fishing"	Authors implicate 14 taxa belonging to 8 genera, but it is unclear if this is due to chance.
Zhou, et al. (2021) "The clinical potential…"	To develop an early diagnostic model based on the relationship between OSCC	Evidence of "fishing"	The authors report results that seem difficult to replicate

Article	Study Aim	Statistical Approach	Findings
	and oral microbiota.		
Chen, et al. (2021) "Taxonomic and functional…"	To investigate the role of ecological patterns in healthy and diseased oral microbiomes	Evidence of "fishing"	Unclear

As shown in Table 2.5, the study aims for four of these studies were diverse, and one did not have a discernible study aim (Chang et al., 2019; J.-W. Chen et al., 2021; Gopinath, Kunnath Menon, et al., 2021; Takahashi et al., 2019; Zhou et al., 2021). One study sought to “clarify the relationship” between oral microbiota among Japanese oral cancer patients (Takahashi et al., 2019), while Chen and colleagues (2021) intended to investigate the role of “ecological patterns” in both healthy and diseased microbiomes, and another author group intended to develop an “early diagnostic model” based on OSCC microbiota (Zhou et al., 2021). The way these aims were stated, it was unclear how the authors would approach their study designs, in terms of what comparisons they would make between what metrics (Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011). One study aimed to determine whether periodontal pathogens specifically had a role in oral cancer development (Chang et al., 2019), while another study lacked stated aims altogether (Gopinath, Kunnath Menon, et al., 2021).

As described in Table 2.5, there is evidence of statistical fishing in all of the papers in this category in this review (Chang et al., 2019; J.-W. Chen et al., 2021; Gloor et al., 2017; Gopinath, Kunnath Menon, et

al., 2021; Kers & Saccenti, 2022; Leung, 2011; Takahashi et al., 2019; Zhou et al., 2021). Many of the authors reported that they calculated different diversity, relative abundance, and other metrics, but when they do not pair the metrics with an *a priori* hypothesis, then they are essentially admitting that all of their analyses will be *a posteriori*, or conducted after the data are gathered without a presupposed hypothesis (Chang et al., 2019; J.-W. Chen et al., 2021; Gloor et al., 2017; Gopinath, Kunnath Menon, et al., 2021; Leung, 2011; Takahashi et al., 2019; Zhou et al., 2021). Doing this greatly complicates the understanding of the true scientific results of these studies (Leung, 2011).

In each of the cases in Table 2.5, the studies engaged in *post hoc* analysis, yet presented the results as if they were derived from an *a priori* analysis (Chang et al., 2019; J.-W. Chen et al., 2021; Gopinath, Kunnath Menon, et al., 2021; Leung, 2011; Takahashi et al., 2019; Zhou et al., 2021). The reason this is inappropriate is that when conducting a true *post hoc* analysis, adjustments are made to the cut point p-value to account for multiple comparisons, and after the application of this adjustment, interpretations are made (Leung, 2011). Without doing this, authors are likely to report spurious results – some of which appear to be obviously incorrect. For example, in the abstract of Takahashi et al. (2019) it is reported that "*Peptostreptococcus*, *Fusobacterium*, *Alloprevotella*, and *Capnocytophaga* were more abundant in the cancer group compared to the control, whereas *Rothia* and *Haemophilus* were less abundant ($p < 0.01$)" (p. 120). Because this p-value was derived from an unadjusted *post hoc* analysis, the results should not be reported

without proper adjustment, as it is likely after adjustment, the significance might disappear (Leung, 2011). With a similar scientifically incorrect approach, Zhou and colleagues (2021) reported that they developed a diagnostic model for OSCC microbiota with a greater than 95% accuracy and a 0% false positive rate, which on the face of it seems very optimistic (Leung, 2011).

The American Statistical Association (ASA) has established "Ethical Guidelines for Statistical Practice" that they keep updated (American Statistical Association, 2022). These are used to guide the professional development of statisticians, including biostatisticians. These guidelines include "Principal B: Integrity of Data and Methods" (American Statistical Association, 2022). In terms of the statistical issues described with these studies, had a professional statistician been consulted and had recommended the analyses presented in these papers, it would have been unethical. However, these ethical guidelines only apply to professional statisticians. Technically, anyone who is able to use statistical software will be able to produce results and report them. Currently, the peer review process is expected to conduct oversight of such issues, but this is an optimistic position. (Makin & Orban de Xivry, 2019) describe ten common statistical errors made in scientific papers and conducting post hoc analysis without adjustment of *p* value is among them.

In one of the articles in this category, the authors mentioned they wanted to see whether periodontal pathogens have a role in OSCC and reported their findings that *P. gingivalis* and *F. nucleatum* had a higher

relative abundance in cancer vs. normal tissue (Chang et al., 2019). This seems to be a classic example of what can happen without applying appropriate statistical methods, where the authors “confirm” their hypothesis without using statistics according to the ethical guidelines provided by the ASA (American Statistical Association, 2022; Gloor et al., 2017; Kers & Saccenti, 2022; Leung, 2011). In these cases, if a statistician conducted the analysis, it would be seen as unethical, whereas if a statistician was not responsible for the analysis, it would likely be seen as an honest mistake.

Nevertheless, when apparently biased results like these are reported as arising from rigorous scientific testing, it creates two primary issues (Leung, 2011). First, it is seen by the ASA as unethical, because it presents scientific findings in the literature as if they had been subjected to rigorous statistical testing when they have not, creating misinformation (Leung, 2011). But besides being misleading, publishing findings as if they had undergone rigorous statistical testing when they have not confuses evolution of scientific thinking on the topic (Leung, 2011). While this is a problem in many scientific domains, including both management and medicine, whenever it occurs, it seriously impedes scientific progress in that domain (Leung, 2011).

Up to this point, three of the five studies in this category have been discussed (Chang et al., 2019; Takahashi et al., 2019; Zhou et al., 2021). The first article engaged in *post hoc* analyses, but reported and interpreted findings with an unadjusted p-value (Takahashi et al., 2019); the second article appeared biased toward the authors’ hypothesis due

to the lack of adjustment for multiple comparisons in their *post hoc* analysis (Chang et al., 2019); and the third reported that the authors were indeed conducting a *post hoc* analysis, but then did not include an adjustment for multiple comparisons, and interpreted the findings anyway (Zhou et al., 2021). The fourth article in this category did not include a discernible research aim (Gopinath, Kunnath Menon, et al., 2021), and the fifth article intended to investigate the "role of ecological patterns" in healthy tissue and OSCC which is unclear (J.-W. Chen et al., 2021). In the article without a clear research aim, the authors apparently intended to make inferences about how the oral bacteriome is different when comparing leukoplakia patients, OSCC patients, and healthy controls, but without a specific study aim and rigorous statistical testing described, it is not possible to understand this article (Gopinath, Kunnath Menon, et al., 2021). The article intending to investigate the role of ecological patterns describes many methods for deriving data from sample, and presents many different visualizations (J.-W. Chen et al., 2021). Without a clear focus and rigorous study design and statistical approaches, it is not possible to understand what evidence this article contributes to the scientific literature (J.-W. Chen et al., 2021).

Case-Control Studies

There are two case-control studies included in this comprehensive review of NGS of 16s rRNA gene oral microbiome results from studies of OSCC (Gopinath, Menon, et al., 2021; Perera et al., 2018). Case-control control study designs can be extremely useful for building evidence even though they are observational in design

(Brooks, 2016; Gibbons et al., 2018; Hu & Dignam, 2019; Schifano, 2019; Semrau et al., 2023). The downside is that case-control designs have a lot of inherent bias, so careful steps need to be taken at the study design stage to minimize bias, and these are often challenging in microbiome studies (Brooks, 2016).

Brooks (2016) recommends that hypotheses involving microbiome measurements in case-control studies either seek to see if there is a difference in α diversity between cases and controls, or to test if relative abundance follows a different multivariate probability distribution in cases vs. controls (to provide evidence for *post hoc* testing of specific microbiota). How the hypotheses are formulated and posed in a case-control study provides the setting for the sources of bias that can creep into any case-control design, but in microbiome studies, the propensity for bias can be elevated due to the exploratory nature of the hypothesized causal parameters being tested (Brooks, 2016; Gibbons et al., 2018; Schifano, 2019). When authors do not clearly state the theoretical exposure (i.e., hypothesized cause, such as particular biomarker profile) being analysed in the case-control design, and follow a traditional case-control statistical analysis, then even if the data were collected according to a case-control framework, the analysed data cannot be interpreted (Schifano, 2019).

In the first case-control study in this review, the authors were actually conducting a follow-up analysis of data gathered in a case-control study reported in a different article (see Table 2.2) (Perera et al., 2017, 2018). As reported in the original study, the cases were made up

of 25 Sinhala males (aged 40 and over) with histologically-confirmed OSCC, and the controls were comprised of 27 Sinhala males diagnosed with fibro-epithelial polyps (FEP) (Perera et al., 2017). However, no case-control hypotheses were proposed in the original report, and upon further inspection, the study did not have all the necessary case-control features, and did not follow the traditional statistical approach for case-control studies (Brooks, 2016; Gibbons et al., 2018; Perera et al., 2017, 2018; Schifano, 2019). In the article included in this review, the authors said their research aims were to “corroborate the findings” from another study of similar design, but did not say what exact findings from the previous study they were seeking to corroborate (Perera et al., 2018). The article does not attempt statistics, and contains many visualizations of data (Perera et al., 2018). In the abstract, the authors list the different members of the microbiome they found in the different groups (Perera et al., 2018).

The final of the two case-control studies included in this review did not have a clear reported study aim (Gopinath, Menon, et al., 2021), and it was observed that the first author of this article is the same first author of the article included in the cross-sectional category in this review that also did not include a clear study aim (Gopinath, Menon, et al., 2021). This lack of a study aim precluded a case-control study design framework and statistical investigation in this paper (Brooks, 2016; Gibbons et al., 2018; Schifano, 2019). The authors reported different taxa found in the samples analysed from 48 OSCC patients and 46 healthy controls who were evaluated for OSCC, but without using

statistical testing and the correct case-control analytic framework, these results are likely due to chance (Brooks, 2016; Gibbons et al., 2018; Gopinath, Menon, et al., 2021; Schifano, 2019).

2.5. Discussion and Conclusion

Primary Findings

Per the objectives of this comprehensive review, through purposive sampling, 22 studies were identified with applicable scientific evidence to inform the design of research investigating high-throughput, NGS technology targeting the 16s rRNA gene in the study of the oral bacteriome OSCC. Of those 22 studies, only two used rigorous enough study methods to allow for interpretation, hence the findings related to the oral bacteriome in OSCC in only these two studies will be synthesized here.

In the first study, authors found that among Taiwanese individuals evaluated for OSCC and found to either be normal, have an epithelial precursor lesion, or have cancer, the compositions of five genera - *Bacillus, Enterococcus, Parvimonas, Peptostreptococcus,* and *Slackia* – significantly differed between the epithelial precursor lesion group to the cancer group (Lee et al., 2017). In the second study, in two separate Taiwanese cohorts of OSCC patients, it was found that there were significant alterations in the diversity and relative abundance between healthy and OSCC tissue, in that the genus *Fusobacterium* was relatively more abundant and genus *Streptococcus* was relatively less abundant in OSCC tumour sites compared to control (Su et al., 2021). Also, their study showed that microbial genes related to terpenoid and

polyketide metabolism played a functional role in the tumour microenvironment (Su et al., 2021). Because these authors used two separate cohorts, they were able to validate their results from the first cohort to the second, thus increasing the robustness of their findings (Su et al., 2021).

Unfortunately, summarizing the results of two studies cannot provide an answer to the question of how the oral bacteriome should be characterized in OSCC. The results of the two studies appear to be contradictory in some ways, but if all of the studies in this review had been conducted with the same high level of statistical rigor, their results could have been synthesized into this interpretation, and a clearer picture of the oral bacteriome in OSCC might have emerged.

Study Design and Statistical Issues in Microbiome Studies

It seems an overreach to consider these two articles “landmark” studies simply because they used adequate study design and statistical techniques, and were able to report scientific findings (Lee et al., 2017; Su et al., 2021). On the other hand, the results of this review expose serious issues in the field of microbiome analysis, especially with respect to the oral microbiome (Gloor et al., 2017; Gopinath, Kunnath Menon, et al., 2021; Gopinath, Menon, et al., 2021; Kers & Saccenti, 2022; Leung, 2011). Although guidance exists for the design and analysis of microbiome studies using NGS 16s rRNA gene technology, this guidance appears to go largely unheeded, especially in the study of the oral microbiome (Brooks, 2016; Gibbons et al., 2018; Gloor et al., 2017; Gopinath, Kunnath Menon, et al., 2021; Hu & Dignam, 2019; Kers &

Saccenti, 2022; Schifano, 2019; Semrau et al., 2023). Although many methodologic issues were raised in this review, only two primary ones will be addressed here: issues with scientific rigor in study design and statistical analysis, and issues with accounting for compositional data in analysis (Gloor et al., 2017; Gopinath, Kunnath Menon, et al., 2021; Hu & Dignam, 2019; Kers & Saccenti, 2022; Leung, 2011).

Issues of Scientific Rigor

It is unreasonable to expect that all microbiology researchers will be biostatisticians or epidemiologists. However, microbiology research requires an advanced understanding of biology, so certainly the need to utilize rigorous statistical methods when comparing biological cohorts should be appreciated by these researchers (Brooks, 2016; Kers & Saccenti, 2022; Leung, 2011). In several of the studies included in this review, the authors say in the methods section that they will calculate multiple measures of diversity and other metrics, but do not include an *a priori* hypothesis, and do not go on the report the numerical results they say they will calculate (Chang et al., 2019; Mukherjee et al., 2017; Sarkar et al., 2021; Takahashi et al., 2019; Zhang et al., 2020). They also do not conduct *post hoc* adjustments, and then they provide an interpretation that may even contain a p-value (Chang et al., 2019; Mukherjee et al., 2017; Sarkar et al., 2021; Takahashi et al., 2019; Zhang et al., 2020). These author groups appear to be describing that they are fishing in their methods sections, and this suggests that they are unaware that the practice is inappropriate (Leung, 2011). In two studies included in this review, the authors did not include a discernible research

aim (Gopinath, Kunnath Menon, et al., 2021; Gopinath, Menon, et al., 2021), and both case-control studies included did not apply the case-control methodology correctly (Gopinath, Menon, et al., 2021; Perera et al., 2018).

Very straightforward articles exist to give guidance to researchers using NGS 16s rRNA technology in terms of biostatistical and epidemiologic considerations necessary for such research, and these are not being cited in papers such as the ones in this review (Brooks, 2016; Gibbons et al., 2018; Hu & Dignam, 2019; Kers & Saccenti, 2022). This observation suggests that there may be an educational gap in the training of microbiome researchers. This review showed that authors continue to conduct, publish, and review research studies without having an adequate understanding of the study design and statistical considerations that need to be applied in their field. Many of the papers included in this review had long lists of authors from highly reputed labs, and all of the papers with these serious methodologic errors underwent peer-review and were published in prestigious journals.

On balance, it is important to consider the scientific challenges faced by the teams of scientific authors who produced these papers. By the nature of the questions posed, these studies are necessarily multi-disciplinary. This requires clinicians and microbiologists to be able to effectively interact with and communicate to bioinformaticians and the statisticians. As these domains do not have a common language, these study design and statistical errors likely represent a failure of these groups to communicate effectively. Clinicians may be seeking guidance

they can use in a clinical setting, or results from an exploration, while microbiologists by their profession only focus on analysing the microbiome, so they are well-versed in laboratory methods, but do not have a clinical perspective. Statistician and bioinformaticians may offer help after data collection, but if an *a priori* hypothesis is not set up at the study design stage, there is little that can be done. It is important to focus on how challenging the research process is, and ways that these challenges can be reduced.

As a remedy, Hamada and colleagues (2019) proposed the creation of a new field of knowledge called molecular pathological epidemiology (MPE). They described it as a transdisciplinary field designed to investigate hypothesized causes of disease (termed "exposures" in epidemiology) that may include microorganisms or utilize molecular pathological signatures of disease (Hamada et al., 2019). The vision would be to essentially aggregate all the existing guidance on how to do epidemiologic research that involves the microbiome and other biomarkers into one transdisciplinary field so that standards may be developed and applied (Hamada et al., 2019). Consequently, such reform could result in studies of the oral microbiome in OSCC with more robust designs that are comparable (Hamada et al., 2019).

Accounting for Compositional Data in Analysis

Gloor and colleagues (2017) accurately identify sources of bias inherent in the compositional data that arise from 16s rRNA gene technology. A main point of their writing (which is echoed in Mandal et al. (2015)) is that the 16s rRNA gene results do not have a denominator,

per se (Gloor et al., 2017). Imagine a situation where a sample is drawn from a human, split into three portions, and then analysed. Any individual microbiome signature may be detected at some level in each of the portions, but this does not provide a clear picture of the *proportion* of that particular microbiome in the whole sample portion, entire sample before it divided, an entire lesion that was the source of the sample, or entire human host *relative to all the other microbiomes* (Gloor et al., 2017; Mandal et al., 2015). By definition, samples are “cleaned” of material to enable the investigation of the microbiota of interest, and this necessarily introduces bias (Gloor et al., 2017; Mandal et al., 2015).

While addressing this analytic puzzle is certainly a topic that could be remedied through the creation of a new MPE field of knowledge, currently, many researchers who study the microbiome are well aware of this challenge, and have developed easy-to-use, practical solutions (Gloor et al., 2017; Hamada et al., 2019; Mandal et al., 2015). Hence, improved attention to sharing information within the microbiology research community could help elevate awareness of this particular issue, and highlight the many practical solutions available (Gloor et al., 2017; Mandal et al., 2015). In fact, both articles cited here were written with the intention of educating the microbial research community on this topic (Gloor et al., 2017; Mandal et al., 2015).

Building an Evidence Base Behind the Oral Microbiome

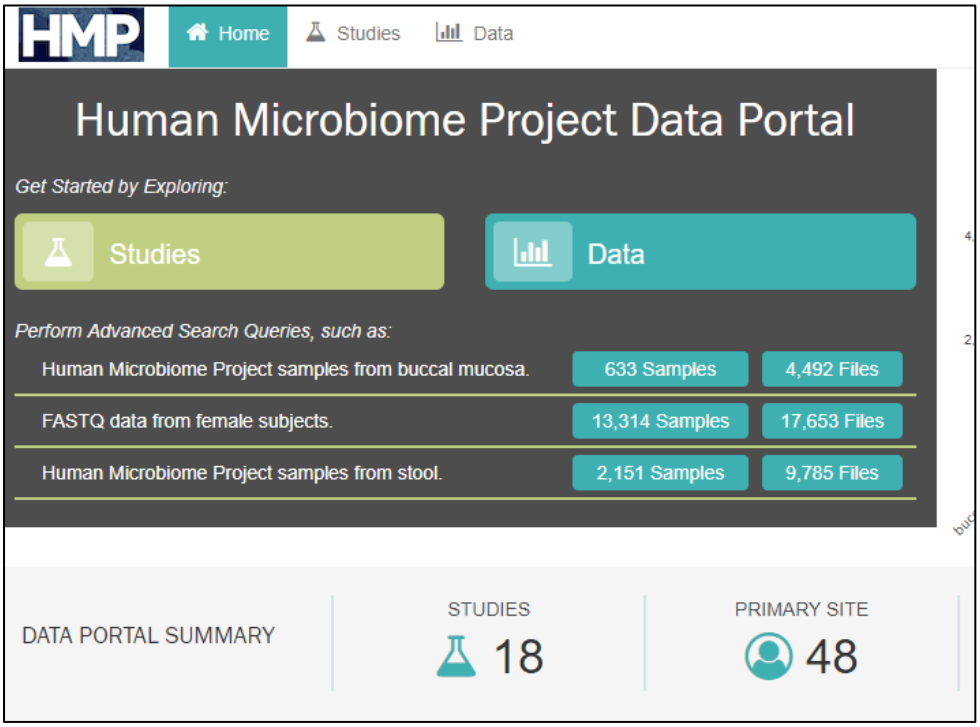
Although it is acknowledged that applying epidemiologic study designs and rigorous biostatistical methods may be challenging in microbiome research, this review suggests that those investigating the

oral microbiome for OSCC diagnostic and therapeutic targets are not progressing as quickly as researchers of biomarkers for other cancers (Hu & Dignam, 2019). As evidenced by only two in 22 articles reviewed having statistical methods adequate for interpretation, the state of the current research methods in the oral microbiome of OSCC are far inferior to the ones being used in lung cancer and for other neoplasms, where research has advanced to a therapeutic level (Hu & Dignam, 2019). Not only does this lack of using rigorous scientific research methods prevent the progression of the field of OSCC research toward more useful applications like in diagnostics and immunotherapies, it also consumes valuable resources. Each of the studies that was unable to report a rigorous finding represented unnecessary monetary expenditure as well as scientific delay. Even studies with very small sample could have provided some usable scientific evidence had they stated presupposed hypotheses; even in the absence of adequate sample, a test could have taken place, and scientific knowledge could be gleaned (Kers & Saccenti, 2022). Yet, peer-reviewers approved them without these features, and they were published.

This disparate inattention to epidemiologic and biostatistical considerations in the study of the oral microbiome in OSCC compared to other microbiomes is exemplified by the relative lack of epidemiological documentation available for the HOMD (T. Chen et al., 2010; *HOMD :: Human Oral Microbiome Database*, n.d.; Human Microbiome Project, n.d.). By contrast, the data portal for the HMP, the

parent project of the HOMD, presents epidemiologic documentation up front (see Figure 2.3).

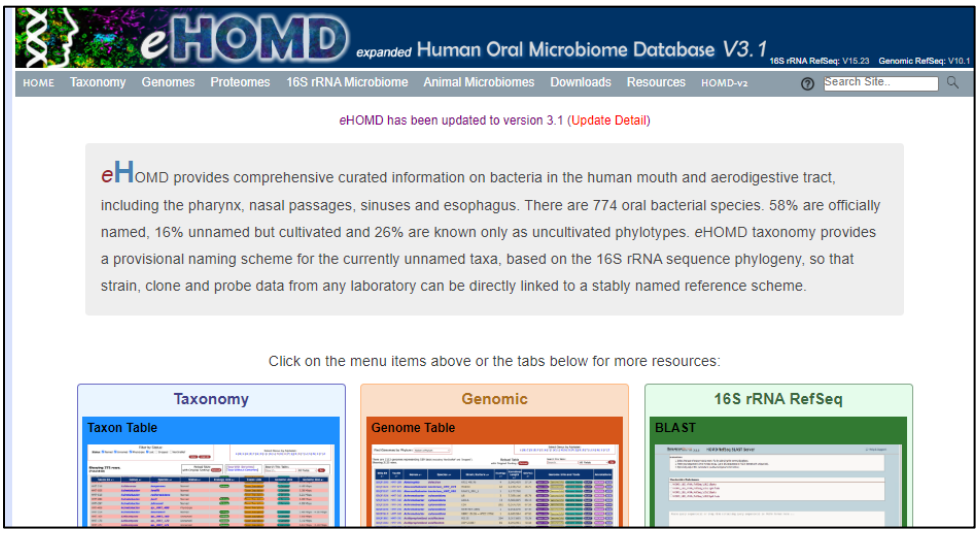
Figure 2.3. Landing page of the human microbiome project data portal



To be clear, the HMP involves a cohort study where well-characterized samples from oral sites are collected and analysed, with the results available on the HMP portal (Markowitz et al., 2012). However, these data are distinct from the HOMD's collection.

Notice that on the landing page for the HMP data portal in Figure 2.3, the epidemiologic information is immediately available. The number of studies hosted on the repository is displayed clearly, and functionality to search for the studies and review their relevant documentation is included. By contrast, the HOMD access page does not direct the user the epidemiologic information behind the studies included in the database (see Figure 2.4).

Figure 2.4. Upper part of the landing page for the HOMD



Observe that in Figure 2.4, there is not an obvious menu choice to obtain information about the epidemiologic studies contained in this database. The lower part of the page (not shown in Figure 2.4) provides access to various analytic tools, but does not provide a search page or any other portal to review the epidemiologic information from the source studies included in the HOMD. This database was designed for a different time and purpose.

That those who maintain the HOMD could continue to make such a serious oversight speaks to the issues specific to research on biomarkers for OSCC. Although it could be argued that the HOMD has nothing to do with clinical studies, Escapa and colleagues (2018), who host and maintain the database, proposed it as a valuable resource for both basic and clinical researchers. All 22 studies reviewed were essentially at the bottom of the evidence-based pyramid in terms of study design quality; the most rigorous study design reviewed was cross-sectional, as the case-control studies included did not in actuality have a case-control design (Semrau et al., 2023). It appears that there was

some confusion in the establishment of the HOMD, because theoretically, it should be as usable as the oral component of the HMP if it was indeed developed for research that could be applied in the clinical or epidemiologic setting (T. Chen et al., 2010; Escapa et al., 2018). If it was developed for another purpose, it is unclear exactly what that purpose is.

This situation further underscores the necessity of setting up a transdisciplinary field that will combine microbiome research with rigorous epidemiologic and statistical methods as proposed earlier (Hamada et al., 2019). It seems that this type of reform is needed in order to provide scientific checks and balances in studies using NGS 16s rRNA gene technology in general, and much needed oversight to investigations into the oral microbiome reported in the scientific literature (T. Chen et al., 2010; Hamada et al., 2019; Hu & Dignam, 2019). When projects like the HOMD and HMP are set up, it is necessary to be transparent about the goal of these projects: what they are intended to do, and what kind of science they are intending to serve. OSCC microbiome researchers should eagerly embrace such reform, as it will promote scientific rigor in their field while facilitating the progression of OSCC research to the level of lung and other cancers (T. Chen et al., 2010; Hamada et al., 2019; Hu & Dignam, 2019).

Strengths and Limitations

This comprehensive review has some strengths and limitations. In terms of its strengths, the use of purposive sampling ensured that the 22 articles reviewed were topically appropriate and should have been

able to inform investigations of the oral microbiome in OSCC. These articles were all published since 2010, and all focused on the technology of interest, which is NGS of 16s rRNA gene, and all analysed the oral microbiome in OSCC. However, this review eventually was limited by the poor quality of the articles available for selection. Only two of the 22 used fundamental study design and statistical methods, so only those could be interpreted. However, it is believed that had more studies been included, this issue would still exist, as there seems to be a persistent knowledge gap about basic research methods throughout the OSCC microbiome literature. While this lack of scientific knowledge about the oral microbiome in OSCC resulting from a lack of appropriate research methods provides an open field for investigation and new findings, it also reveals that the OSCC research field is not progressing meaningfully towards defining and developing therapeutic agents (Hamada et al., 2019; Hu & Dignam, 2019). Due to this unfortunate circumstance, it is unlikely that many of the findings of this review would be changed had other articles on OSCC and the oral microbiome been selected for review.

Conclusion

This comprehensive review used purposive sampling to identify 22 appropriate articles for review; of these, only two were of adequate scientific rigor to be interpreted. The first study found that five genera significantly differed between the epithelial precursor lesion group to the cancer group in a male Taiwanese cohort (Lee et al., 2017), and the second study found the genus *Fusobacterium* was relatively more

abundant and genus *Streptococcus* was relatively less abundant in OSCC tumour sites compared to healthy tissue in two cohorts of Taiwanese OSCC patients, and that microbial genes related to terpenoid and polyketide metabolism played a functional role in the tumour microenvironment (Su et al., 2021). Of course, making a conclusion about microbial genes playing a functional role is based on speculation, not rigorous science. Lack of rigorous studies on the oral microbiome in OSCC have led to a comparative lag in the progression toward defining biomarkers that may serve as therapeutic targets in OSCC (Hu & Dignam, 2019). While rigorous study design and biostatistical methods have been developed to analyse microbiome data, research groups specifically investigating the oral microbiome appear to be uniformly unaware of these approaches (Brooks, 2016; T. Chen et al., 2010; Gibbons et al., 2018; Gloor et al., 2017; Kers & Saccenti, 2022; Schifano, 2019). While acknowledging that microbiome studies and analyses are extremely challenging from an epidemiologic and statistical point-of-view, calls are being made to consolidate this research under one interdisciplinary umbrella field to ensure knowledge gaps like this one are addressed field-wide (Hamada et al., 2019). Such reform could greatly improve the quality of the scientific output related to investigations into the oral microbiome in OSCC.

Chapter 3: Substudy 1: Diversity and Differential Relative Abundance in the Bacteriome of Oral Squamous Cell Carcinoma and Oral Epithelial Dysplasia Lesions: Results from the United Kingdom

3.1. Background

Cancer of the oral cavity, specifically oral squamous cell carcinoma (OSCC), is a worldwide public health issue, contributing to significant global morbidity and mortality (Gormley et al., 2022; Salehiniya & Raei, 2020). Although primary prevention through the elimination of risk factors is the optimal public health approach to reducing rates of OSCC, as risk factors for OSCC remain prevalent worldwide, rates of OSCC continue to climb (Gormley et al., 2022; Salehiniya & Raei, 2020). Once oral cavity cancer is diagnosed, it is often in late stages, and there is poor prognosis (Gormley et al., 2022; Salehiniya & Raei, 2020). The advancement of biomarker analysis has expanded the capacity to study the oral microbiome, and using modern methods, members of the oral microbiome could be identified as diagnostic, prognostic, or treatment-related biomarkers (Bugshan & Farooq, 2020; Zaura et al., 2021). This section describes the epidemiology and risk factors for OSCC, and the advancement of biomarker studies of the oral bacteriome in OSCC.

<u>Epidemiology and Risk Factors for Oral Squamous Cell Carcinoma (OSCC)</u>

The purpose of this section is to summarize and highlight just the points provided in Chapter 2 that will later be relevant to my specific

study in this current chapter, Chapter 3. Currently, as described in Chapter 2, there is a global increase in incidence rates of cancer of the oral cavity, specifically OSCC, with a complementary rise in mortality rates (Gormley et al., 2022; Salehiniya & Raei, 2020). Tobacco use and betel nut chewing (*paan*) are two of the strongest risk factors for OSSC, but these behaviours are more common in certain countries, and among certain demographics (Gormley et al., 2022; Salehiniya & Raei, 2020; J. Yang et al., 2021). Globally, men are more likely to use tobacco, but trends as to how it is used (e.g., cigarette smoking, pipe smoking, smokeless) vary by country (Salehiniya & Raei, 2020). Chewing of the betel nut is generally practiced among Southeast Asia and Asia-Pacific regions, and is again more common among men (Gormley et al., 2022; Salehiniya & Raei, 2020; J. Yang et al., 2021). As men are more likely to use tobacco, chew betel nut, and use alcohol, they are more likely to have the risk factors in OSCC (Salehiniya & Raei, 2020). For this reason, globally, men consistently have higher incidence and mortality rates for OSCC than women (Salehiniya & Raei, 2020).

OSCC in the United Kingdom (UK)

This is a summation of findings specifically related to the UK. Northern Europe, and specifically the United Kingdom (UK), has a unique epidemiologic profile of head and neck cancer (HNC) (which includes OSCC) (Salehiniya & Raei, 2020). Salehiniya and Raei (2020) observed that between 1995 and 2011, oral cavity cancer rose 7.3% for men and 6.5% for women in England, while rising 2.8% in men and 3.0% in women in Scotland in the same period. UK Cancer Registry data

showed an increase of 34% in total oropharyngeal cases across all regions between 2011 and 2018, and incidence rates have been found to be highest in Scotland, where they have increased 85% between 2001 and 2012 (Gormley et al., 2022). Regional differences reflect socio-economic status (SES) patterns, as low SES is also a risk factor for HNC and OSCC (Gormley et al., 2022; Salehiniya & Raei, 2020).

In their analysis of more recent data, Bosetti and colleagues (2020) found that the age-standardized death rates from oral and pharyngeal cancer increased over the entire UK between 2010 and 2015 by 11.6% in men and 7.1% in women. When stratified regionally, rates increased between 2010 and 2015 among men by 12.2% and women by 8.4% in England and Wales, however when looking at Northern Ireland, among men, rates increased by 78.8%, and among women, rates dropped by 33.9% during the same time period (Bosetti et al., 2020). In Scotland, male rates dropped by 6.6%, while female rates increased by 15.2% during the same period (Bosetti et al., 2020). These mortality rates reflect that most cancers identified in the UK are at advanced stages, as 58.8% of HNCs diagnosed in the UK are at stage III or IV (Gormley et al., 2022).

Biomarkers in OSCC

Although survival from cancer of the lip, oral cavity, pharynx and larynx has increased 10% over the past few decades, still, just over half these patients survive beyond five years (Gormley et al., 2022; Hashim et al., 2019). For this reason, focus should be placed on primary prevention of OSCC through the elimination of established risk factors

(especially tobacco, betel, and alcohol use), and through early identification to provide the best prognosis (Hashim et al., 2019; Salehiniya & Raei, 2020).

From this perspective, reliable biomarkers relevant to the natural history of OSCC could be very useful to identify (Zaura et al., 2021). In cancer, biomarkers can be seen as either biologic molecules or diagnostic tests carried out on fluids or tissue for the purpose of diagnosis and/or treatment (Hu & Dignam, 2019). Exploratory studies have identified candidate biomarkers in OSCC that may indicate the degree of metastasis, impacts of etiologic factors like tobacco or alcohol consumption, and other diagnostic information (Bugshan & Farooq, 2020). The ability to reliably identify valid biomarkers of OSCC would create a pathway to the development of advanced diagnostic and prognostic capabilities, as well as biomarker-targeted immunotherapies that could expand treatment options and improve survival (Hu & Dignam, 2019; Zaura et al., 2021).

However, the current biomarker picture for OSCC is admittedly confusing (Cristaldi et al., 2019; Doddawad et al., 2022). Doddawad and colleagues (2022) summarized and classified the scientific literature to date regarding biomarkers and oral cancer, noting that sampling the tissue (healthy or lesional) for biomarker identification has many advantages over other types of specimens such as saliva, and should be used whenever possible. The authors described several historical classification schemes for oral cancer biomarkers, and explained that setting up classifications itself is challenging, and based on available

science (Doddawad et al., 2022). As described in the paper, one historical classification scheme looked at the type of biomarker in terms of its general function (e.g., cytokines and blood group antigens), whereas another system classified the biomarker in terms of what it indicated (e.g., tumour growth markers, and markers of tumour suppression and antitumor response) (Doddawad et al., 2022). Ultimately, the authors proposed four classifications of oral biomarkers to use going forward: prognostic biomarkers, biochemical markers, hormone receptors, and proliferation markers (Doddawad et al., 2022). This simple system is based on a clinical perspective, and promotes the progression of biomarker research towards translation into clinical science applicable to patients (Doddawad et al., 2022).

However, the development of this simple classification system does not solve the problem of accurately placing biomarkers already researched into the classification system (Cristaldi et al., 2019; Doddawad et al., 2022). Cristaldi and colleagues (2019) provided a comprehensive review of research results from OSCC biomarker studies in various patient samples. The authors presented the results from the perspective that regardless of the exact biomarker being measured, the resulting numerical value is simply a measurement of a larger process going on that may impact biomarker values across different types of biomarkers (Cristaldi et al., 2019). The article reviewed circulating tumour DNA, extracellular vesicles (EVs), and microRNAs as salivary OSCC biomarkers, and proposes looking into circulating tumour cells (CTCs) as another set of OSCC biomarkers (Cristaldi et al., 2019).

Although this review paper classifies oral cancer biomarkers differently than the previous one, both articles suggest that current research directions now should focus on determining which of the various biomarkers are involved in which processes, and classifying them into relationships of target biomarkers and companion biomarkers, as described in Chapter 2 (Cristaldi et al., 2019; Doddawad et al., 2022; Hu & Dignam, 2019). This step is necessary to enable biomarkers to be developed into valid and reliable measurements that can be used to facilitate clinical decision-making (Cristaldi et al., 2019; Hu & Dignam, 2019).

Oral Bacteriome Biomarkers in OSCC

Due to advancing technology, the use of the next-generation sequencing (NGS) of the 16s rRNA gene for biomarker identification in cancer has greatly expanded in the last decade, and this has generated advanced knowledge as to methodologic considerations necessary for biomarker studies (Hu & Dignam, 2019; Zaura et al., 2021). In OSCC, studies have been conducted on the oral bacteriome using 16s rRNA gene sequencing in an effort to identify members critical to the formation of the OSCC lesion as reliable biomarkers (Lee et al., 2017; Mukherjee et al., 2017; Pushalkar et al., 2011; Sarkar et al., 2021; Su et al., 2021; Zhang et al., 2020; Zhao et al., 2017). Many of these studies have included a case series of OSCC patients, along with one or more groups of patients or healthy individuals as a comparison group (Lee et al., 2017; Mukherjee et al., 2017; Pushalkar et al., 2011; Sarkar et al., 2021; Su et al., 2021; Zhang et al., 2020; Zhao et al., 2017). These studies

analysed lesional tissues, healthy tissues, and saliva from participants in an effort to identify members of the oral microbiome which may be responsible for OSCC lesion development (Lee et al., 2017; Mukherjee et al., 2017; Pushalkar et al., 2011; Sarkar et al., 2021; Su et al., 2021; Zhang et al., 2020; Zhao et al., 2017).

Unfortunately, these studies have not produced consistent results, and the most likely reason is lack of scientific rigor in study design, study conduct, or statistical testing (Hu & Dignam, 2019; Kers & Saccenti, 2022; Zaura et al., 2021). In terms of study design, researchers designing oral microbiome studies often fail to make important epidemiologic considerations when sampling individuals, and may include small sample sizes too heterogeneous to produce statistical results (Goossens et al., 2015; Hu & Dignam, 2019; Ou et al., 2021; Zaura et al., 2021; Zheng, 2018). In terms of study conduct, scientific authors often do not realize how much configuration is involved in a typical 16s rRNA gene sequencing pipeline, and therefore do not fully report their configurations or their rationale (Gloor et al., 2017; Goossens et al., 2015; Zaura et al., 2021). Finally, in terms of statistical testing, rarely are hypotheses posed *a priori* (Kers & Saccenti, 2022; Ou et al., 2021; Zaura et al., 2021). Instead, *post hoc* testing is routinely performed, and typically interpreted without adjustment for multiple comparisons (Kers & Saccenti, 2022; Ou et al., 2021; Zaura et al., 2021).

A few of these studies will be presented here to provide context. Zhao et al. (2017) took swabs of both lesional and healthy tissue sites from 40 OSCC patients in China, but due to the absence of an *a priori*

hypothesis, the authors were only able to conclude that the taxa they found in their study was different than found in previous studies. Using a similar study design, Mukherjee and colleagues (2017) compared tongue cancer tissue to normal tissue in a case series of tongue cancer patients (apparently in the US) undergoing resection. Like the previous author group, these authors did not pose any *a priori* hypotheses; for statistical analyses, they ran many *post hoc* tests, did not use an adjustment, and struggled to interpret the results, saying that they found "differences" in the microbiome between cancer and normal tissue (Mukherjee et al., 2017).

In oncology, epithelial dysplasia refers to anomalous growth of the epithelium resulting in a lesion exhibiting disturbed differentiation and maturation (Tilakaratne et al., 2019). Epithelial dysplasia is diagnosed microscopically based on individual features of the lesion, and has long been considered a pre-malignant disorder, although it is recognized that many cases of epithelial dysplasia do not proceed to a cancerous tumour (Tilakaratne et al., 2019). This is not just an artifact of grading or having a range of disease conditions with similar architectural changes labelled together (Tilakaratne et al., 2019). Some cases simply remain in a state of epithelial dysplasia but do not become cancerous, although the patient remains high risk (Tilakaratne et al., 2019). Oral epithelial dysplasia (OED) therefore is specifically considered a pre-malignant condition requiring careful management and follow-up (Tilakaratne et al., 2019). Shen and colleagues (2023) conducted a systematic review of oral dysbiosis as a risk factor for the onset and carcinogenesis of OED, and

overall, the authors concluded that when comparing the oral microbiome of OED to OSCC, the results are inconsistent. The authors attributed their lack of findings due to the heterogeneity of the type and size of sample in the underlying groups, reflecting criticisms echoed about lack of rigor in study design across oral microbiome studies (Hu & Dignam, 2019; Shen et al., 2023; Zaura et al., 2021).

In an attempt to focus on findings from only high-quality oral microbiome studies in OSCC, Peter and colleagues (2022) conducted a systematic review and meta-analysis of case-control studies of OSCC and the associated microbiome. Reflecting the lack of rigor in the previous works, the authors were only able to identify eleven articles of high enough quality to be included in the systematic review, and only five of those qualified for their meta-analysis (Peter et al., 2022). Five of the eleven studies used the patient as their own control, whereas the others used other participants as controls, and the eleven studies included represented samples from the following countries: China, US, Sri Lanka, Japan, and Taiwan (Peter et al., 2022). For the meta-analysis, the authors pooled the data from the five studies and reanalysed them with the intention of identifying specific pathogenic bacteria, genes, and functional pathways in OSCC (Peter et al., 2022). For the specific pathogenic bacteria, they identified the *Fusobacterium* genus and specifically implicated *F. nucleatum* (Peter et al., 2022). They also identified the gene K06147, an ABC transporter, as associated with OSCC, and discussed potential functional pathways (Peter et al., 2022).

The authors of the meta-analysis, as well as authors of other works reviewed here, repeatedly emphasized the challenges with determining the scientific significance of the results of oral microbiome studies without rigorous study design (Goossens et al., 2015; Hu & Dignam, 2019; Kers & Saccenti, 2022; Ou et al., 2021; Peter et al., 2022; Zheng, 2018). First, they recommended that swabs be used whenever possible, because they are a more direct measure of the tissue environment, as associated with OSCC, and discussed potential functional pathways (Doddawad et al., 2022; Peter et al., 2022). Potential functional pathways refer to how biomarkers behave in terms of protein function; several different potential functional pathways in OSCC are proposed by Doddawad and colleagues (2022) as well as Peter et al. (2022). Next, they described how they believe their meta-analysis was confounded through poor epidemiologic sampling and measurement in some of the studies contributing data to the meta-analysis, introducing bias (Peter et al., 2022). They explained how different approaches to amplicon sequence analysis (in terms of computer platforms and bioinformatics) may be the sources of important differences in results between studies, and emphasized that this issue is currently a serious shortcoming that needs to be taken into account when comparing the results of oral microbiome studies (Peter et al., 2022). Finally, they emphasized the necessity to use appropriate statistical methods and hypothesis testing, while keeping in mind that microbiome data are compositional in nature, and need to be analysed

using methods appropriate to this structure (Gloor et al., 2017; Peter et al., 2022).

As described earlier, OSCC is prevalent globally, but due to differential distributions of risk factors, rates are higher in certain countries, including the UK (Bosetti et al., 2020; Gormley et al., 2022; Salehiniya & Raei, 2020). The aim of my study was to test three hypotheses: Among UK cancer centre patients with either an OSCC or OED lesion, 1) the OSCC lesions would have on average more α diversity in the oral bacteriome than the OED lesions; 2) there would be significant β diversity between the oral bacteriome of the OSCC lesions compared the OED lesions; and 3) the OSCC lesions would have significantly different relative abundance for species/genera than the OED lesions.

3.2. Methods

This was a cross-sectional study of swab tissue from patients diagnosed with either OSCC or OED. In this study, participants were consented and enrolled, then data and tissue specimens were collected. The specimens underwent processing, and statistical tests were applied to answer the research aims. Details of these steps are included here.

Participants and Setting

All participants were recruited from the Head and Neck Cancer Centre, University College of London Hospital (UCLH) or the Oral Medicine clinics at Eastman Dental Hospital (EDH) in London, UK. UCLH Head and Neck Cancer Centre is a tertiary care centre which provides diagnosis and treatment of cancer identified through the

general practitioner (GP) through primary care and local hospitals (UCLH, n.d.). The EDH is a dental hospital that also provides diagnosis and treatment for oral cancer (UCL, 2018). Data collection took place between November 2018 and November 2019.

All participants had been referred to the study location to be evaluated for potential OSCC. Study participants needed to meet the following qualification criteria. Because of the unique participant group being targeted, it was difficult to reduce bias due to the study design. No sample size estimate was developed because recruitment was limited to one clinical centre within a limited time. Further, it was intended to be a pilot study, so it was expected to be underpowered.

Inclusion Criteria

Must be age 18 and older.

Must be diagnosed with either OSCC (case definition in Macey (2015)) or OED (case definition in Tilakaratne (2019)) as a result of the evaluation. Patients with OSCC were recently diagnosed, while patients with OED included both recently diagnosed as well as OED patients with ongoing active disease.

Must speak, understand and write English well enough to provide informed consent in writing.

Exclusion Criteria

Patients currently undergoing treatment for OSCC or OED were excluded.

Patients with a current diagnosis of active cancer of any type other than OSCC were excluded.

Patients reporting antibiotic intake in the previous two weeks, or previous long-term antibiotic use were excluded (per Goodrich (2014)).

Enrolment and Consent

Patients at both clinical locations diagnosed with OSCC or OED were approached for study participation. Records were reviewed for eligibility. Those who were interested underwent eligibility screening and consent.

Participant Data Collection

Those who consented were assigned an anonymous Study ID which was then used on all study materials. After consent and enrolment, the participants were asked a series of questions from a questionnaire, and recorded their answers on paper (see Appendix A for consent form and participant information sheet, and Appendix B for blank questionnaire). These data were later transferred to a Microsoft Excel spreadsheet for data analysis. After the participant visit, the gathered clinical records about the participant's biopsy information were added to the spreadsheet.

Specimen Sampling and Analysis

This section will describe sample collection, transport, storage, and processing. These processes were designed to limit bias in measurement.

Sample Collection

For swab samples, both lesion area (i.e., biopsied lesion diagnosed as OSCC or OED) and an anatomically matched non-lesional area (i.e., clinically intact) were swabbed by applying the swab at an

angle of approximately 20° using a gentle rubbing movement against the area (ten times). Then, the swab was rotated 180°, and the same process was followed for the other side of the swab (three rotations were performed if the patient could tolerate it). Then the swab was pressed against the wall of the DNeasy PowerSoil Kit collection tube (QIAGEN Inc., Germantown, MD, USA) for 20 seconds to ensure bacterial transfer. Then, the tip of the swab was detached from the handle and kept inside the tube.

Although healthy tissue was collected at this stage, it was not considered in the current analysis. Please see Chapter 4 (Substudy 2) where analysis of healthy tissue is included in the study design.

Sample Transport and Storage

All sample tubes were dated and labelled with the participant's Study ID. Sample tubes were placed immediately on ice and transported within two to four hours to the laboratory. Swab samples were placed in DNeasy PowerSoil Kit collection tubes and directly stored at -80 ºC.

Sample Processing

This section will describe DNA extraction and quantification, amplification, purification, library preparation, and final bioinformatics processing.

DNA Extraction

DNA extraction of the samples were performed using the DNeasy PowerSoil Kit (QIAGEN Inc., Germantown, MD, USA) following the manufacturer's instructions with slight modifications (see Appendix C). DNA extraction included a combination of mechanical (using bead-

beating) and chemical methods for cell lysis. All DNA samples were stored in tubes labelled according to the Study ID and stored at - 20 ºC until further processing.

DNA Extract Quantification

Quality and quantity of DNA extracts were measured using Nanopore (Oxford, Cambridge, UK).

Polymerase Chain Reaction (PCR) Amplification (V3-V4 16s rRNA gene)

PCR was performed to amplify the hypervariable regions V3 and V4 of 16s rRNA gene using NGS specific primers with overhang adaptors for metabarcoding (Appendix D) The reactions were set up to a final volume of 25 µl by preparing the following mixtures: 10 µl of 2.5 x Master Mix 16s/18s Basic (suitable for 100 reactions) (vh bio, Gateshead, UK) and 5.2 µl of sterile molecular grade water (Sigma-Aldrich, St. Louis, MO, USA). The Master Mix volumes were: 0.8 µl Moltaq; 2 µl of reverse primers; 2 µl of forward primers; and 5 µl of DNA template. The reactions were then carried out in a thermocycler under the following conditions: Initialization cycle at 95 ºC for 30 seconds, followed by 30 cycles of denaturing at 95 ºC for 30 seconds, 30 cycles of annealing at 58 ºC for 40 seconds, 30 cycles of elongation at 72 ºC for 1 minute, and a termination cycle at 72 ºC for 10 minutes. A negative control of “no DNA template” was used to ensure no contamination.

PCR Amplicon Purification

Agencourt AMPure XP (Beckman Coulter, Indianapolis, IN, USA) was used for amplicons purification.

Library Preparation and Sequencing

The following steps were performed to further enhance the quality of the library. Amplicons were quantified using Qubit™ dsDNA High Sensitivity Assay kit (ThermoFisher Scientific, Waltham, MA, USA). Concentrations were then adjusted to create an equimolar pooled library of 5 nM. The pooled library underwent an additional step of short polymerase chain reaction (PCR). This was conducted through the use of 5µl of the pooled library, and adaptors primers (P5 and P7) with the following conditions: Initialisation cycle at 98 ºC for 30 seconds, followed by 2 cycles of denaturing at 98 ºC for 10 seconds, followed by annealing at 60 ºC for 30 seconds, then elongation at 72 ºC for 60 seconds, ending with a termination cycle at 72 ºC for 5 minutes. Purification of the library was then carried out using Agencourt AMPure XP (Beckman Coulter, Indianapolis, IN, USA).

The library was then quantified using Qubit™ dsDNA High Sensitivity Assay kit (ThermoFisher Scientific, Waltham, MA, USA) and sent to the UCL Genomics Centre. At the UCL Genomics Centre, the library was assessed for both quality and quantity using TapeStation (Agilent Technologies Inc., Santa Clara, CA, USA) utilizing HS D1000 ScreenTape and Qubit respectively. The 4 pM library spiked with 10% 12 pM PhiX was loaded on the MiSeq platform v2 kit (Illumina Inc., San Diego, CA, USA), and processed using 2 x 250 bp paired-end sequencing runs per manufacturer's instructions.

Bioinformatics Processing

De-multiplexed reads were received from the UCL Genomics Centre on the Illumina BaseSpace sequence Hub Account. Data were then imported locally using BaseSpace Downloader in Fastq format. A manifest file was then created using text format then imported to QIIME2 (versions qiime2-2019.4 and qiime2-2022.2) (Bolyen et al., 2019). Reads were then visualised and reviewed in the Interactive Quality Plot tab in qiime2 viewer (https://view.qiime2.org/) to assess the quality of the reads. Then DADA2 plugin (q2-dada2) was used for both merging and denoising the paired-end reads after adjusting the parameter *--p-trunc-len*, to trim out nucleotides with poor quality scores (Callahan et al., 2017). Forward reads were truncated at 250 nucleotides, while reverse reads were truncated at 225 nucleotides. This was done to filter out nucleotides with low quality score (QS) to improve the paired-end merging of the reads. The cutpoint for a low quality score was an average QS of 25 (Estaki et al., 2020). Before processing and merging, the original demultiplexed sequence count was n = 9,519,032. The number of paired-end reads after filtering was n = 7,155,813. Therefore, in processing, 2,363,219 input sequences were discarded after trimming and assembly. Taxonomy analysis was done by first importing both the sequence reads and taxonomy reference files of 16s rRNA gene amplicons from the extended Human Oral Microbiome Database to QIIME2 (HOMD_16S_rRNA_RefSeq_V15.1.fasta.txt and HOMD_16S_rRNA_RefSeq_V15.1.qiime.taxonomy.txt respectively). The Human Oral Microbiome Database was chosen over other popular

alternatives that included more than oral human microbiome data - such as GreenGenes and Silva. The reason why it was selected was that training the classifier using environment-specific data is suggested to improve the accuracy of taxonomy assignment (Estaki et al., 2020). Once taxonomy analysis was complete, qiime feature-classifier using the following parameters *--p-f-primer CCTACGGGNGGCWGCA*G and *--p-r-primer GACTACHVGGGTATCTAATCC* was used to extract the hypervariable regions (V3-V4) of interest from the reference reads to enhance the performance of the classifier. Then classifier was trained using qiime feature-*classifier fit-classifier-naive-bayes*. Taxonomy was then assigned using qiime feature-*classifier classify-sklearn*. Phylogenetic tree reconstruction was done using fragment-insertion plugin (q2-fragment-insertion) (Janssen et al., 2018) against pre-built sepp tree using Silva database (sepp-refs-silva-128.qza). The phylogenetic tree was then visualised using empress plugin (q2-empress). Normal samples were then filtered out of the feature table generated by DADA2, and this was ensured through a visualization. Next, the sequences of normal samples were filtered out based on the filtered feature table. This produced a *.qza file which was utilized in statistical testing.

Statistical Analysis

To compare α diversity between OSCC and OED lesions, Shannon's Index was used, and to compare β diversity, weighted UniFrac distances were used (Kers & Saccenti, 2022). To compare differential abundance, ALDEx2 was used (Gloor, 2023; Gloor et al.,

2017, 2023). For all statistical testing, α was set at 0.05 (Kers & Saccenti, 2022).

To perform the tests, first, the *.qza file was further processed in qiime2 to develop core phylogenic metrics using a sampling depth of 28,000. One drawback of the MiSeq platform is that it creates an uneven sampling depth for each sample. Because of this, a sampling depth of 28,000 was chosen for qiime2 processing, since it is the depth at which the likelihood of identifying taxa increased without losing samples in my data. Next, to conduct the α diversity test, the *alpha-group-significance* command in QIIME2 was run, which conducts Kruskal-Wallace analysis testing on the Shannon Index. For the β diversity test, a PERMANOVA was conducted on the weighted UniFrac distance. For differential abundance, number of significant taxa based upon Wilcoxon tests with Benjamini-Hochberg-corrected p-values from ALDEx2 were considered, as well as those with effect confidence intervals that did not cross zero (Gloor et al., 2023).

Ethics

The Integrated Research Application System (IRAS) has a standing ethical committee approval (IRAS project ID 96630) which covers this study. To ensure complete reporting of this observational study, a STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) checklist was completed for this analysis (von Elm et al., 2007) (see Appendix E).

3.3. Results

A total of 46 participants enrolled in and completed the study, including 21 with OSCC (46%), and 25 with OED (54%). Table 3.1 presents a descriptive statistical summary.

Table 3.1. Sample summary

Category	Level	All n, %	OSCC n, %	OED n, %
All	All	46, 100%	21, 46%	25, 54%
Gender	Male	25, 54%	17, 81%	8, 32%
	Female	21, 46%	8, 38%	13, 52%
Age Group	30-39	4, 9%	3, 14%	1, 4%
	40-49	3, 7%	0, 0%	3, 12%
	50-59	8, 17%	3, 14%	5, 20%
	60-69	13, 28%	11, 52%	2, 8%
	70-79	15, 33%	7, 33%	8, 32%
	80-89	3, 7%	1, 5%	2, 8%
Ethnicity	White	30, 65%	15, 71%	15, 60%
	Asian Indian	6, 13%	5, 24%	1, 4%
	Other Asian	7, 15%	3, 14%	4, 16%
	Black	1, 2%	1, 5%	0, 0%
	Not Reported	2, 4%	1, 5%	1, 4%
Use Alcohol	Yes	25, 54%	13, 62%	12, 48%
Use Tobacco	Current or Former	33, 72%	21, 100%	12, 48%
Use Betel Nut	Current or Former	5, 11%	3, 14%	2, 8%

As shown in Table 3.1, although about half the sample was male, men were overrepresented in the OSCC group (n = 17, 46%). As was described in the background, this is likely due to men having a higher prevalence of OSCC risk factors. In terms of age, one third of the sample (n = 15) were aged 70 to 79, with another 28% (n = 13) in the next lower age group, 60 to 69, making this a largely older sample. Almost two thirds of the sample were of White ethnicity (n = 30, 65%). While only about half (n = 25, 54%) of the sample reported any use of alcohol, almost

three fourths (33, 72%) reported tobacco use, and this was 100% (n = 21) of the OSCC group.

Table 3.2 presents a clinical summary of the sample.

Table 3.2. Clinical summary.

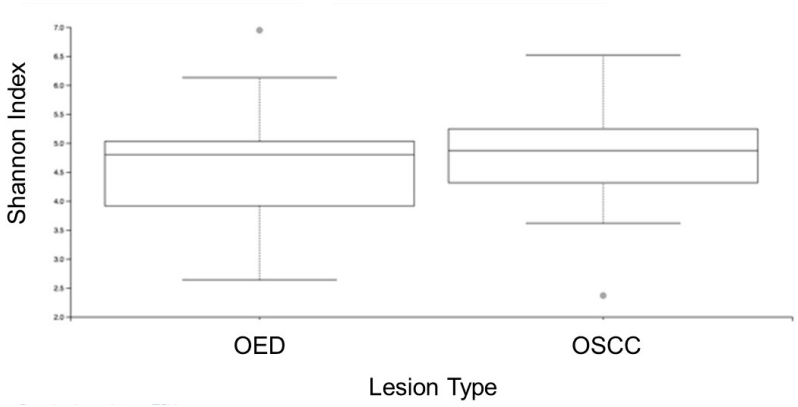
Category	Level	All n, %	OSCC n, %	OED n, %
All	All	46, 100%	21, 46%	25, 54%
Number of Teeth	None	5, 11%	4, 19%	1, 4%
	11 to 17	5, 11%	1, 5%	4, 16%
	20 to 24	9, 20%	6, 29%	3, 12%
	25 to 29	20, 43%	12, 57%	8, 32%
	30 to 32	7, 15%	2, 10%	5, 20%

As shown in Table 3.2, 58% of the sample had at least 25 teeth, which was 67% in the OSCC group (n = 14), and 52% in the OED group (n = 13).

Diversity Results

The α diversity analysis showed that there was not a statistically significant difference in α diversity between OSCC and OED lesions (p = 0.4337). Figure 3.1 compares box plots of Shannon Index between OSCC and OED.

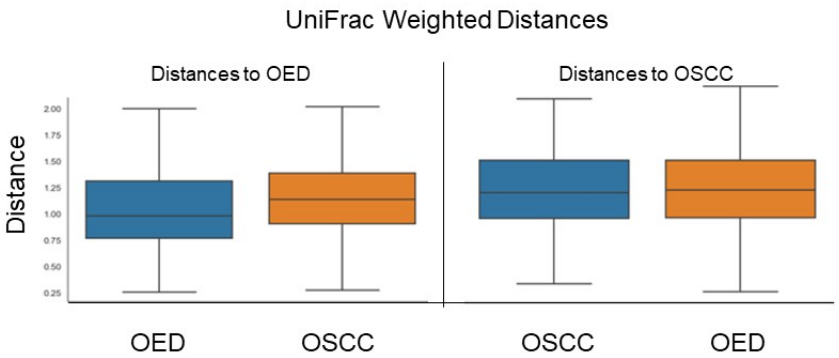
Figure 3.1. Box plots comparing Shannon index: OED vs. OSCC



As can be seen in Figure 3.1, the interquartile range (IQR) of the Shannon Index in both groups overlapped, and the medians were very similar, which is consistent with non-significant statistical test results.

In terms of β diversity, results showed statistically significant β diversity between OED and OSCC lesions (p = 0.013, see Figure 3.2).

Figure 3.2. Box plots comparing β diversity: OED vs. OSCC



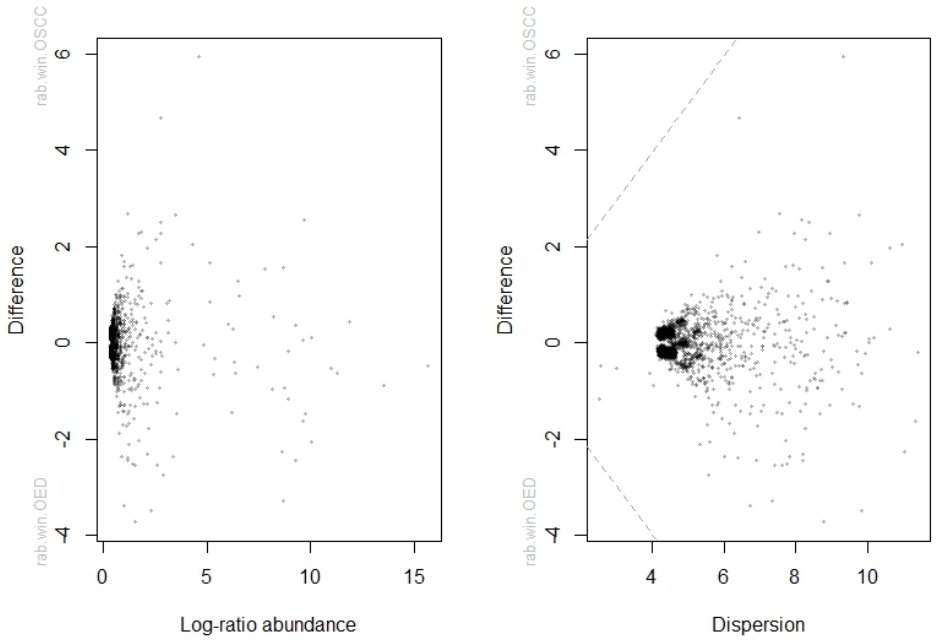
As shown in Figure 3.2, in the PERMANOVA approach, multiple permutations of the data are developed to estimate distances in from the base sample (e.g., OED) to itself as well as other samples (e.g., OSCC). As there were only two samples included, the box plot on the left in

Figure 3.2 visualizes the PERMANOVA results when OED is selected as the base sample, and the one on the right visualizes results when OSCC is selected as the base sample (Kers & Saccenti, 2022). It is not evident from the visualizations why the results were statistically significant, because in both sets of box plots, the IQRs overlap and the medians are visually close.

Differential Abundance Results

The ALDEx2 approach identified 4,201 taxa to compare between OSCC and OED lesions. Bland-Altman and dispersion plots are included in Figure 3.3 to facilitate interpretation of results.

Figure 3.3. Bland-Altman and dispersion plots: OED vs. OSCC



The Bland-Altman plot in Figure 3.3 (left) shows the association between the relative abundance (log-ratio abundance on the x-axis) and the magnitude of difference per sample (on the y-axis). The plot on the right is identical to the one on the left, but it replaces log-ratio abundance

on the x-axis with dispersion. In both plots, black points represent rare taxa, and grey ones represent abundant taxa (Gloor et al., 2016). As can be seen in both plots, many rare taxa coalesce on the left side of the x-axis, while an equally large proportion of more abundant taxa appear dispersed across the x-axis.

Any genera that were differentially abundant between the two groups ($q \leq 0.1$) would be represented by a red dot on the plots in Figure 3.3 (Gloor et al., 2016). As there was a lack of red dots, further analysis was conducted. As part of ALDEx2, for each of the 4,201 taxa identified, a Wilcoxon test was performed, and an expected Benjamini-Hochberg-corrected p-value was generated (Gloor, 2023). Also, for each of the taxa, an effect size with confidence bounds was identified (Gloor, 2023). A descriptive analysis was conducted to determine if any of the Wilcoxon corrected p-values were statistically significant at $p < 0.05$, and to see if there were any taxa with effect sizes with confidence intervals that did not include zero. This analysis found that there were no taxa with Wilcoxon corrected p-values < 0.05, and there were no taxa with effect sizes with confidence intervals that did not include zero. On the basis of Figure 3.3 and this descriptive analysis, the current research fail to reject the null, and interpret the results to say there is not statistically significant differential abundance in the oral bacteriome between OED and OSCC lesions in this study.

3.4. Discussion

In this study of UK cancer centre patients with either OED or OSCC, results showed that in terms of the oral bacteriome, OSCC

lesions did not have statistically significantly more α diversity, but did have statistically significantly more β diversity than OED lesions. Also, there was not a statistically significant difference in the differential relative abundance of taxa between OSCC and OED lesions. It is notable that Peter and colleagues (2022) identified *Fusobacterium* when conducting their meta-analysis of pooled data of OSCC, while in the current study, *Fusobacterium* (nor any other member of the oral bacteriome) was not identified as being associated with OSCC. It is possible that the lack of finding in the current study for *Fusobacterium* is due to low sample, and that a larger study would find a relationship (as was done by the authors who conducted the pooled analysis) (Peter et al., 2022). It is also possible that the reason why the current study did not identify differences in *Fusobacterium* between OSCC and OED because the comparison was with OED controls, not healthy controls.

As described in Chapter 2, many studies in this domain have study design flaws, and this study was designed to overcome those that could be addressed. First, clear research aims were defined before the data were analysed, and an analysis plan was designed *a priori*. Next, efforts were made to clearly characterize the sample descriptively, which is often missing in published articles. Next, the informatics processing approach selected was supported by the literature (Gloor et al., 2023). Finally, the statistical presentation in the results was designed to clearly relate to the research aims stated earlier, and statistical approach defined in the methods.

For reasons cited by other author groups, it is difficult to synthesize findings from this study with the background literature due to lack of consistency and rigor in the background literature (Shen et al., 2023). As an obvious example, the lack of scientific consistency and rigor in studies of the OED oral bacteriome preclude the ability to compare this study's findings to those (Shen et al., 2023). Shen and colleagues (2023) noted in their review of ten studies on OED that there was substantial heterogeneity in the populations and sizes of samples, so there was a consistent issue of comparability between groups. Although many studies of the OSCC bacteriome exist in the literature, due to epidemiologic, operational, and statistical issues, it is not possible to scientifically compare their results (Bugshan & Farooq, 2020; Doddawad et al., 2022; Gloor et al., 2017; Goossens et al., 2015; Kers & Saccenti, 2022; Ou et al., 2021; Peter et al., 2022; Zaura et al., 2021; Zheng, 2018) to the results of the current study.

Admittedly, the current study suffers from some of the same limitations as previous studies, including heterogeneity of sample with respect to OSCC risk factors, and potential lack of comparability between OED and OSCC sample. In other words, those in the OSCC sample may have a different distribution of underlying risk factors for OSCC than the OED sample. In the current study, it would have been ideal to reduce the heterogeneity of the sample to, for example, only a high risk group, such as men aged 60 and over who are current smokers diagnosed with either OSCC or OED (see Table 3.1). Statistically, in a homogeneous high-risk sample, it is more likely that oral bacteriome

profile that is unique to OSCC vs. OED could be identified if it existed. If there were certain members of the oral bacteriome indicated in the development of OSCC in male smokers in the UK, it is likely that using an optimal sampling strategy in this study would have allowed their identification. As described before, an optimal sampling strategy would only filter in a high-risk group so as to reduce the statistical noise. By including such a small, heterogenous sample, if there are different oral bacteriome members involved in different disease processes (e.g., conversion of OED to OSCC outside the presence of tobacco), it would not be possible to identify this. All these characteristics represent bias in the current study.

However, placing high restrictions on qualifications for the study sample to filter in only a high-risk population would have reduced the feasibility of conducting the study at all. Had those restrictions been applied to the current study, only 17 OSCC patients would have qualified (see Table 3.1). Further, it would have been difficult to identify matched OED patients in the same age bracket, as OED is a risk factor for OSCC, and likely there would be fewer patients identified at the OED stage, making the sample unbalanced. Trade-offs in study design were considered acceptable to facilitate research that could be done within the timeframe of a doctoral project. Ideally, in the future, a multi-centre study could be designed with such restrictions eligibility qualifications that proceeds along a longer timeline. That would enable the accumulation of a large enough, homogenous sample to promote the ability to make and test statistical inferences.

Nevertheless, this study makes an improvement upon previous studies on this topic which have been criticized for not using rigorous study design and biostatistical methods. As is recommended, *a priori* statistical testing was conducted, measurement approaches were transparent, and the results were interpreted according to current conventions (Kers & Saccenti, 2022; Zaura et al., 2021). By being transparent about measurement approaches, specimen processing, bioinformatics pipelines, and statistical strategies in their communication, scientific authors can facilitate a standardization of the scientific literature around the study of the oral microbiome (Doddawad et al., 2022; Gloor et al., 2017; Goossens et al., 2015; Ou et al., 2021; Peter et al., 2022).

In conclusion, while there was not a statistically significant difference in α diversity, there was statistically significant differences in β diversity between OSCC and OED lesions in this study of patients in the UK. No differentially relatively abundant taxa were identified that would differentiate OSCC from OED lesions, but the sample was small and heterogenous. Further, this study only investigated differences between OSCC and OED lesions. A study comparing OSCC or OED specimens to healthy tissue in the same individual may provide another opportunity for detecting a signal in the oral bacteriome that could be a marker for disease status or progression. Future studies should also seek to identify target and companion biomarker profiles that can be utilized clinically in the identification and management of OSCC.

Chapter 4: Substudy 2: Study to Characterize How the Oral Microbiome Differs in OSCC and OED from Healthy Tissue in the Same Individual

4.1. Background

This second of three substudies continued to investigate the oral microbiome for biomarkers diagnostic of oral squamous cell carcinoma (OSCC) or oral epithelial dysplasia (OED). This was done by comparing the α diversity, β diversity, and differential abundance between the oral microbiome in lesion tissue and healthy tissue in the same individuals. This section will briefly summarize the epidemiology, risk factors, and treatment approaches for OSCC and OED, and will describe what is known about identifying diagnostic biomarkers for OSCC and OED in the oral bacteriome. Finally, this section will reflect on the findings from Substudy 1 in Chapter 3, and propose a different research approach to investigate differences in the oral bacteriome between diseased and healthy tissues within the same individual.

OSCC and OED: Epidemiology, Risk Factors, and Treatment Approaches

As described in Chapters 2 and 3, OSSC is prevalent globally, posing a worldwide public health problem (Gormley et al., 2022; Salehiniya & Raei, 2020). OSCC incidence rates are higher in geographies where risk factors for OSSC are concentrated (Gormley et al., 2022; Salehiniya & Raei, 2020). The strongest risk factors for OSCC are using tobacco or betel nut, and since these behaviours are more likely to take place among men, men generally have higher rates of

OSCC in all geographies (Gormley et al., 2022; Salehiniya & Raei, 2020). Alcohol use is also a risk factor, and is thought to interact with tobacco use to greatly elevate risk (Gormley et al., 2022; Mello et al., 2019; Salehiniya & Raei, 2020).

Epithelial dysplasia is an anomalous growth of the epithelium that leads to a lesion that exhibits disturbed differentiation and maturation (Tilakaratne et al., 2019). Oral epithelial dysplasia (OED) is considered a pre-malignant condition at high risk for progressing into OSCC, although many cases of OED do not progress to OSCC (Tilakaratne et al., 2019). OED has the same risk factors as OSCC, and once identified, a clinical decision must be made between managing the OED lesion conservatively, or engaging in active treatment, which includes both surgical and non-surgical approaches (Tilakaratne et al., 2019).

While those diagnosed with OED have a pre-malignant condition, most patients diagnosed with head and neck cancer (HNCs) like OSCC are identified in later stages, limiting both treatment options and patient survival (Hashim et al., 2019; Marur & Forastiere, 2016). For the 40% of head and neck squamous cell carcinoma (HNSCC) patients who are diagnosed at stage I or II, the typical treatment is with surgery or radiation alone (Marur & Forastiere, 2016). Chemoradiation and chemotherapy are available for advanced cases, but due to substantial toxicity in some of the regimens, patient quality-of-life and survival needs to be weighed when selecting these treatments (Marur & Forastiere, 2016). Screening for OSCC with the most sensitive and modern technology is invasive, so it would be ideal to identify reliable and valid diagnostic biomarkers that

could be identified through less invasive methods (e.g., saliva) (Hashim et al., 2019).

Identifying Diagnostic Biomarkers of OSCC or OED in the Oral Bacteriome

As described in Chapters 2 and 3, studies of the oral bacteriome seeking to identify diagnostic or prognostic biomarkers must follow rigorous scientific study designs and statistical approaches to be able to detect biomarkers reliably and validly enough to be used in clinical decision-making (Hu & Dignam, 2019; Zaura et al., 2021). More recent studies have used 16s rRNA gene NGS as an approach to identifying biomarkers in the oral bacteriome (Lee et al., 2017; Mukherjee et al., 2017; Pushalkar et al., 2011; Sarkar et al., 2021; Su et al., 2021; Zhang et al., 2020; Zhao et al., 2017). However, methodologists have criticized the rigor of the study design approaches in many of these studies, as well as their lack of scientific statistical hypothesis testing (Goossens et al., 2015; Hu & Dignam, 2019; Ou et al., 2021; Zaura et al., 2021; Zheng, 2018).

Focusing on epidemiologic considerations, studies of the natural history of the oral microbiome are observational, not interventional (Zaura et al., 2021). That is because in order to study the natural history of a condition, it is necessary to observe it, and not apply an intervention that could change the course of the disease (Zaura et al., 2021). Choice of observational study design in oral microbiome studies is important because study designs have different strengths and limitations (Hu & Dignam, 2019; Zaura et al., 2021). Cross-sectional designs, such as the

one used in substudy 1, have a high risk of bias, and this is argued to be even higher in oral microbiome studies due to the complex and dynamic nature of host-microbiome interactions (Zaura et al., 2021). Some of this bias comes from the heterogeneity of the samples typically included in oral microbiome studies, such as including participants with different risk factor statuses with respect to tobacco and betel nut use in the same, small sample study (Zaura et al., 2021). Other primary sources of bias in oral microbiome studies is introduced by using matching approaches, which often inadvertently increase bias rather than reducing it (Zaura et al., 2021).

In substudy 1, lesional swab tissues from OSCC were compared with OED in an effort to evaluate differences in biomarker profiles between the oral microbiome of these tissues. Of particular importance would be a biomarker that could suggest progression from OED to OSCC. Although there was a statistically significant difference in β diversity between the OSCC lesions and OED lesions, this was the only significant finding from substudy 1. Likely due to the heterogeneity and small size of the sample, no statistically significant different abundant members of the oral microbiome were seen between OSCC and OED lesions.

While it was not possible with such a small, heterogenous sample to identify a significantly different oral microbiome biomarker profile between OSCC and OED lesions in substudy 1, it may be possible to identify biomarkers that indicate a difference between diseased and healthy tissues in the same individual. As described in the methods of

substudy 1, for each participant, not only the lesion area was sampled, but another sample was also taken from an anatomically-matched non-lesional area within the same participant. Comparing lesional to non-lesional tissues within the same individual in both OSCC and OED provides an opportunity to identify potential biomarkers in the oral bacteriome that are associated with diseased as compared to healthy tissue in the same person.

OSSC vs. OED in a UK Sample

As described in Chapters 2 and 3, Northern Europe, and specifically the UK, has a distinctive epidemiologic profile of HNC (Bosetti et al., 2020; Salehiniya & Raei, 2020). Regional HNC and OSCC rates in the UK fluctuate based on increases and decreases in the presence of risk factors in the population (Bosetti et al., 2020; Gormley et al., 2022). Regional differences in rates of low socio-economic status (SES), which was discussed as another risk factor for HNC and OSCC in Chapter 2, also contribute to differing rates across the various regions of the UK (Bosetti et al., 2020; Gormley et al., 2022; Salehiniya & Raei, 2020). Across all regions in the UK, incidence rates correlate with mortality rates, as 58.8% of HNCs diagnosed in the UK have already advanced to stage III or IV (Gormley et al., 2022).

Substudy Design

As described in Chapter 3, substudy 1 included a small, heterogeneous sample of OSCC and OED patients from the UK. Although the comparison of the oral microbiome from the surfaces of the lesion did not show significant differences leading to the identification of

target biomarkers, in substudy 1, a comparison with healthy tissue was not conducted. This was purposeful, as it was felt that it would be too overwhelming to present a comparison with healthy tissue in addition to the other results.

Therefore, a second substudy was done. In an effort to identify biomarkers indicating OSCC or OED in the oral bacteriome using 16s rRNA gene sequencing, the aim of this second substudy was to test whether, among patients in the UK with either OSCC or OED, 1) lesional tissues had more α diversity than healthy tissues from the same individual, 2) lesional tissues had more β diversity than healthy tissues in the same individual, and 3) lesional tissues had significantly different relative abundant taxa when compared to healthy tissues in the same individual.

4.2. Methods

As described in substudy 1 (Chapter 3), this is a cross-sectional study of patients from two different UK medical centres diagnosed with either an OED or OSCC lesion. Eligible patients from these centres were consented and enrolled in the study. Patients provided questionnaire information, and data from their medical records was included in the study. Swab samples from both lesional tissue and healthy tissue for all participants were obtained and underwent processing, and statistical tests were employed to answer the research aims. These steps are detailed below. To ensure complete reporting of this observational study, a STROBE checklist was completed (von Elm et al., 2007) (see Appendix F).

Participants and Setting

Please refer to Substudy 1 as the information is identical.

Enrolment and Consent, and Data Collection

Please refer to Substudy 1 as the information is identical.

Specimen Sampling and Analysis, and Bioinformatics Processing

Sample collection, transport and storage, and processing were the same as in substudy 1 (described fully in Chapter 3), so this section will briefly summarize these steps. For sample collection, on all participants, both a lesion area and an anatomically matched non-lesional area were swabbed. The sample from the swab was stored, then transferred to the laboratory.

At the laboratory, the samples underwent DNA extraction and quantification, amplification, and purification, as described in detail in Chapter 3. Additionally, library preparation followed the description in Chapter 3. Bioinformatics processing also took place as described in Chapter 3, only this time, *.qza files were prepared so that pairwise comparisons could be done between lesional and healthy tissue for each participant separately for each group (OSCC vs. OED). The *.qza files were prepared iteratively. Each time a *.qza files was processed, *.qzv files which are visualizations were reviewed to determine the next processing step. Once the final *.qza file was processed, it was output and used for analysis.

Statistical Analysis

To compare α diversity between lesional and healthy tissue in both OSCC and OED, Shannon's Index was used, and paired tests were

performed separately for the OSCC and OED groups (Kers & Saccenti, 2022). To compare β diversity between lesional and healthy tissue, OSCC and OED groups were tested separately using weighted UniFrac distances entered into a PERMANOVA (Kers & Saccenti, 2022). To compare differential abundance taxa between lesional and healthy tissues in each group, ALDEx2 was used (Gloor, 2023; Gloor et al., 2017, 2023), with α set at 0.05 (Kers & Saccenti, 2022).

To develop a data file to support statistical testing, first, the *.qza files were prepared using the same method as described in substudy 1 (Chapter 3) using qiime2. For the α diversity test, the *alpha-group-significance* command in qiime2 was run on each tissue sample group (OSCC lesion vs. healthy, OED lesion vs. OED healthy). This option in qiime2 conducts Kruskal-Wallace analysis testing on the Shannon Index. To evaluate β diversity, the *beta-group-significance* command was run on separate datasets, one with OSCC, and one with OED, which tests using a PERMANOVA. To evaluate differential abundant taxa between healthy and lesional tissues, number of significant taxa based upon pairwise Wilcoxon tests with Benjamini-Hochberg-corrected p-values from ALDEx2 were quantified, as well as the number of taxa with effect confidence intervals that did not cross zero (Gloor et al., 2023).

4.3. Results

As described in substudy 1 (Chapter 3), a total of 46 participants enrolled in and completed the study, including 21 with OSCC (46%), and 25 with OED (54%) (refer to Table 3.1 for a descriptive summary of the sample).

α Diversity Results

Results for the α diversity analysis are visualized in Figure 4.1.

Figure 4.1. Comparison of α diversity results between lesion and non-lesion tissues in OED and OSCC samples

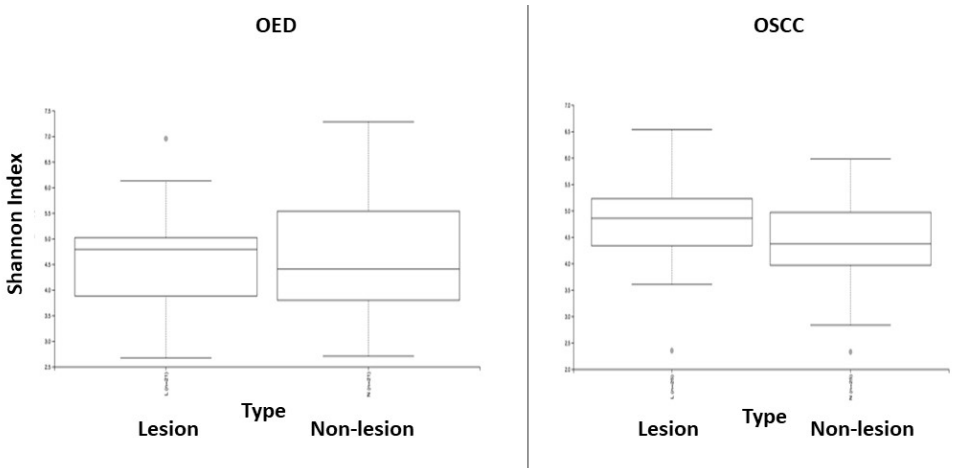
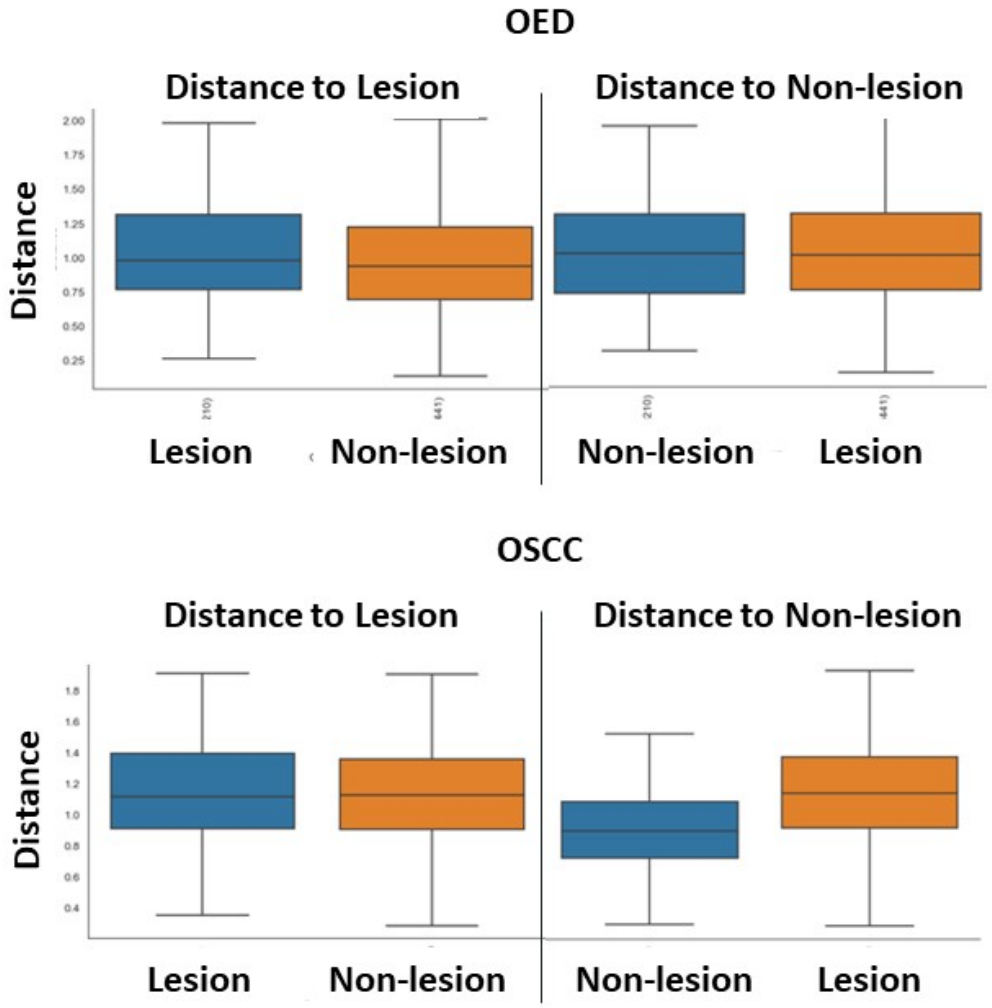


Figure 4.1 shows the box-and-whisker plots for the Shannon Index for OED lesion vs. non-lesion tissue ($p = 0.9298$) on the left side of the figure, and for OSCC lesion vs. non-lesion tissue ($p = 0.1138$) on the right side of the figure. Although there are visual differences in the distributions as depicted in the box plots, these differences did not rise to the level of statistical significance.

β Diversity Results

Results for the β diversity analysis are visualized in Figure 4.2.

Figure 4.2. Comparison of β diversity results between OSCC and OED lesion and non-lesion tissue



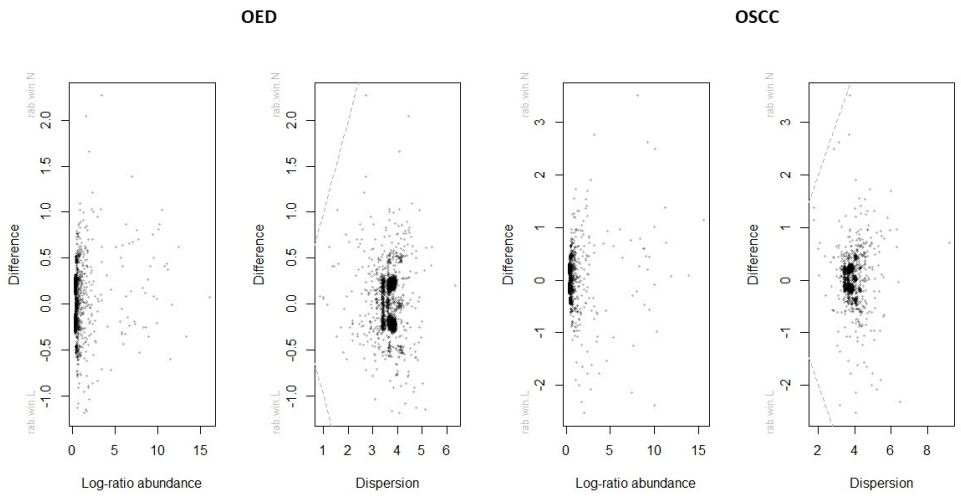
As a reminder, in testing for β diversity, weighted UniFrac distances between the samples that take into account the phylogenetic tree and thus phylogenetic distances between community members was tested, and these are the distances that were examined (Kers & Saccenti, 2022). As can be seen in Figure 4.2, for the OED results at the top, the box plots appear very similar, and the β diversity is not statistically significantly different between healthy and diseased tissue in OED ($p = 0.921$). The results are different for OSCC, shown at the bottom of Figure 4.2, where β diversity was found to be statistically

significantly different between lesion and non-lesion in OSCC (p = 0.001). It is visually apparent in Figure 4.2 that there is a trend toward greater β diversity in the OSCC lesion when viewing the distance to non-lesion (right side of figure), but even though those differences are statistically significant, they are of a small magnitude.

Differential Abundance Results

For the OED analysis, 3,709 taxa were identified, and for the OSCC analysis, 3,516 taxa were identified. Figure 4.3 visualizes the results of the differential relative abundance analysis in Bland-Altman and dispersion plots.

Figure 4.3. Relative differential abundance in OED and OSCC vs. healthy tissue



In Figure 4.3, comparison between relative differential abundance in OED tissue is shown on the left, and the results for OSCC are shown on the right. Both plots are similar in that there are no differentially relatively abundant taxa displayed, because these would produce red dots (Gloor et al., 2016). As black points represent rare taxa, and grey ones represent abundant taxa, both comparisons are similar in that they

plot many rare as well as abundant taxa, and therefore it is difficult to argue that they are different (Gloor et al., 2016). For both the OED and OSCC comparisons, none of the Wilcoxon corrected p-values were statistically significant at $p < 0.05$, and all of the effect sizes included 0. For these reasons, it appears that in this analysis, there are not significantly differentially relatively abundant taxa in OED tissue vs. healthy from the same individual, or in OSCC tissue vs. healthy from the same individual.

4.4. Discussion

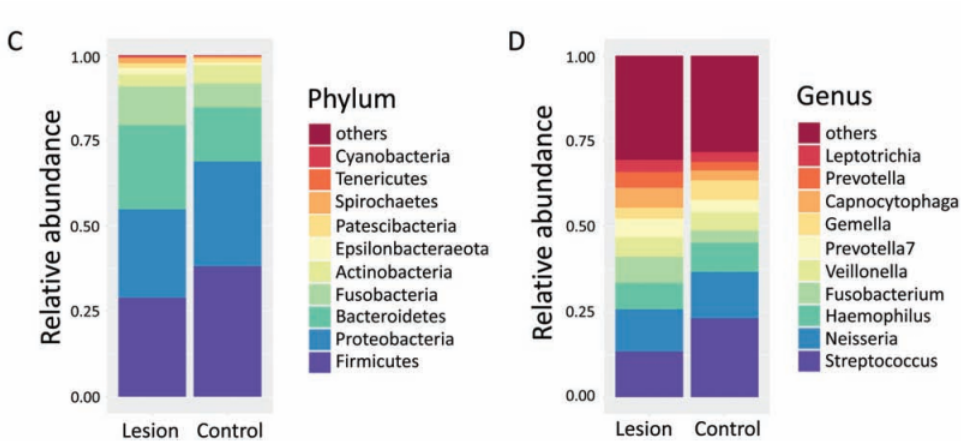
In this UK study comparing both OED and OSCC tissue to healthy tissue in the same individual, in terms of the oral microbiome as measured using 16s rRNA gene sequencing, it was found that neither type of lesional tissues had more α diversity than healthy tissues. However, while it was found that in OED, there was not a significant difference in β diversity between lesional and non-lesional tissues, in OSCC, there was a statistically significant difference in β diversity between lesional and non-lesional tissues. Nevertheless, no taxa were found to be differentially relatively abundant in either OED or OSCC compared to healthy tissues.

Comparing these results to the scientific literature is challenging for the reasons stated in Chapter 2, in that both statistical methods as well as processing configuration pipelines are not comparable (Goossens et al., 2015; Ou et al., 2021). However, it is fair to compare the results to Substudy 1, where only lesional tissues between OED and OSCC were compared. Substudy 1 also found significant differences in

β diversity in OSCC compared to OED, with OSCC appearing more β diverse. In the current study, OSCC tissues were found to be significantly more β diverse than healthy tissues. This seems to be a robust finding, in that it suggests that OSCC tissue that originated as OED tissue may became more β diverse as it evolved. This is purely speculation based on the findings arising from these two studies. If this speculation were to be the case, the challenge is to identify the new taxa that are being added as the OSCC lesion develops.

In both substudy 1 and the current analysis, studies of differential relative abundance was done on taxa numbering in the thousands, and no differentially relatively abundant taxa were detected. This may be an artifact of classification, in that in compositional data like oral microbiome data, how members of the microbiome are classified can greatly impact the analytic approach and the potential interpretation of the results (Gloor et al., 2017). Su and colleagues (2021) provide a descriptive figure from their study of the oral microbiome of Taiwanese oral cancer patients that illustrates this issue, which is reprinted as Figure 4.4.

Figure 4.4. Challenges with oral microbiome classification approaches to compositional data. Reprinted from Su et al. (2021)



As can be seen in Figure 4.4, depending upon whether the members of the oral microbiome were classified to the phylum level or the genus level changed not only the number of levels of the classification, but the patterns of relative abundance identified. It is possible that by changing the level of classification of the current data used in this study and repeating the analysis, significant results might be obtained. Although higher level of classification might not provide sufficient information about specific member/members of the OSCC microbiome, robust statistical findings can provide a better guidance on the development of diagnostic microbiome biomarkers.

The results of this study are consistent with the background scientific literature which has not identified specific biomarkers that signal the development of OED or OSCC, or that signal the progression from OED to OSCC (Doddawad et al., 2022). However, this analysis identified that there appears to be significantly increased β diversity in OSCC swabbed tissues when compared to healthy tissues in the same individual (current study results), and when compared to OED tissues

from other individuals (results of substudy 1). It would be interesting to observe the results from other studies of OSCC, OED and heathy tissue samples from different populations that replicate the same approaches used in substudies 1 and 2 to see if the same patterns appear in different samples. Perhaps as the OSCC lesion forms, the microbiome becomes significantly more β diverse, and if a biomarker signal from the oral microbiome indicating healthy or OED tissue was achieving greater β diversity compared to biomarkers arising from the surrounding tissue could be identified, it could be seen as an early indication of OSCC formation.

While this study has strengths in terms of its statistical and compositional approach to analysing the oral microbiome, it also has significant limitations. As noted in the discussion to substudy 1, the underlying sample is diverse in terms of its OSCC risk factors, so it was unlikely that any particular member of the oral microbiome could be identified as responsible for lesion formation given the likely diverse underlying aetiologies of the pathology seen. Nevertheless, this study used similar methods as existing studies in the literature (Sarkar et al., 2021; Zhang et al., 2020). The current study and similar studies had small, heterogenous samples, likely due to time limitations with respect to positive lesion sample accumulation (Sarkar et al., 2021; Zhang et al., 2020). Due to lack of standardization in the background literature, it is difficult to compare the results of this study to other, similarly-designed studies (Sarkar et al., 2021; Zhang et al., 2020). As highlighted by the inclusion of Figure 4.4, in this study, the level of classification of the taxa

may have impacted the findings relating to relative differential abundance. This study is limited to a UK population with a particular OSCC risk factor profile; different results would like be seen in populations with different risk factor profiles (Gormley et al., 2022; Salehiniya & Raei, 2020). There was enough sample in the current study to allow for the identification of some statistically significant associations. However, obtaining a larger sample would provide greater insight into whether there are indeed members of the oral microbiome that serve as biomarkers that signal the conversion of OED tissue or healthy tissue to OSCC.

In conclusion, although this study found that there was significant differences in β diversity between OSCC lesions and healthy tissues in the same individual, it was not apparent what member of the oral microbiome were responsible for this finding. Taken together, the findings from this substudy and substudy 1 suggest that in the population studied, as OSCC lesions form, their associated oral microbiome becomes more diverse. This appears to not be true of OED lesions, so an increase in diversity beyond a certain threshold could signal progression in OSCC. Future studies should seek to replicate this finding in other OED and OSCC patient populations, as well as to identify which members of the oral microbiome are responsible for the increasing diversity as the OSCC lesion forms.

Chapter 5: Substudy 3: A Reanalysis of OSCC Microbiome Repository Data Shows Consistent Results with Current Study Data

5.1. Background

In substudy 1 presented in Chapter 3, the oral bacteriome of lesional tissues from swab samples of oral squamous cell carcinoma (OSCC) and oral epithelial dysplasia (OED) patients in the United Kingdom (UK) were analysed using 16s rRNA gene sequencing and compared, and the only significant finding was that OSCC lesions had significantly more β diversity than OED lesions. Next, in substudy 2 presented in Chapter 4, lesional and healthy tissue swabs were compared within the same individuals - using the same sample used in substudy 1. Again, the only significant finding was greater β diversity between OSCC lesions and healthy tissues within the same individuals; OED tissues did not show this pattern. These findings suggest that as tissue evolves from healthy or OED to OSCC, it becomes more diverse in terms of taxa. If the additional taxa being added could be identified, they could serve as clinical biomarkers for disease progression.

An important limitation of substudy 2 was the small number of heterogenous samples coming from the UK, an area of the world with a unique risk factor pattern for OSCC (Bosetti et al., 2020). It would be advantageous to analyse OSCC data from a similar study design taking place in a different part of the world, such as China or India, where the risk factor patterns are different, using the same analytic methods (Gormley et al., 2022; Salehiniya & Raei, 2020; Sarkar et al., 2021;

Zhang et al., 2020). That way, the differences in both statistical approach as well as data processing approaches can be analysed for possible effects on study outcomes, and results can be directly compared to Substudy 2 (Gloor et al., 2017; Goossens et al., 2015; Kers & Saccenti, 2022; Ou et al., 2021).

Replication of Substudy 2 Approach

The Sequence Read Archive (SRA) is a public service sponsored by the United States (US) National Institutes of Health (NIH) (National Institutes of Health, n.d.). The SRA is a public data repository that accepts data from sequencing projects that involve human participants or their metagenomes (National Institutes of Health, n.d.). The SRA sequencing data were available for several of the studies that were the subject of the comprehensive review presented in Chapter 2, including a case-series of OSCC patients in China that used a similar design as substudy 2 (Zhang et al., 2020). This study included 50 OSCC patients who had a tumour site and an opposite healthy site swabbed, and the oral microbiome from these tissues were analysed using 16s rRNA gene sequencing (Zhang et al., 2020). The raw reads were placed in the SRA repository under PRJNA533177 (National Institutes of Health, n.d.). These data are available for download and re-analysis by other researchers (National Institutes of Health, n.d.).

Similar to substudy 2, the participants studied by Zhang et al. (2020) were heterogenous in terms of risk factors. This means the risk factors for OSCC were not distributed evenly across the sample, just has was the case in substudy 2. For example, the sample included

individuals with different tobacco use, alcohol use, and betel nut use profiles, and since these are risk factors for OSCC, this sample was considered heterogenous (Zhang et al., 2020). In addition, the sample studied by Zhang et al. (2020) had tumours at all four clinical stages, making it even more heterogeneous.

The availability of the samples studied by Zhang et al. (2020) in the SRA provided the opportunity for the raw sequencing microbiome data to be pre-processed and reanalysed in the same way as the data in substudy 2. This facilitated a direct comparison of the results to that of substudy 2. In other words, the raw reads data from the SRA generated by Zhang et al. (2020) could be processed and analysed to compare the findings directly to the findings of substudy 2, so this became the aim. Specifically, to test whether 1) OSCC lesional tissue swabs had more α diversity than healthy tissue swabs from the same individual, 2) OSCC lesional tissue swabs had more β diversity than healthy tissue swabs in the same individual, and 3) OSCC lesional tissue swabs had significantly different relative abundance of certain taxa when compared to healthy tissue swabs in the same individual.

5.2. Methods

This was a reanalysis of data placed in the SRA that was generated by a previous study (National Institutes of Health, n.d.; Zhang et al., 2020). First, raw sequences from the data underwent a data reformatting and cleaning process to make them comparable to data used in substudy 2. Next, the data underwent further bioinformatic analysis similar to that conducted previously, and then statistical analysis

took place (see chapter 3 and 4). These processes are described in this section. To ensure complete reporting of this observational study, a STROBE checklist was completed (von Elm et al., 2007) (see Appendix G).

Data Reformatting and Cleaning

First, data for project PRJNA533177 were downloaded from the SRA (National Institutes of Health, n.d.; Zhang et al., 2020). For this project, the 100 runs described in the article were identified (Zhang et al., 2020). These were paired-end samples, in that each run had two reads per spot. The library layout was 2 × 300 bp paired-end Illumina MiSeq (Zhang et al., 2020). The primers used to amplify the V3-V4 hypervariable regions in this project were 338F (5′-CCTACGGGNGGCWGCAG-3′) and 806R (5′-GACTACHVGGGTATCTAATCC-3′).

Once the format of the library was determined, data were imported locally in SRA format using the SRA toolkit (Estaki et al., 2020; National Institutes of Health, n.d.). SRA runs were converted to fastq format using *fastq dump* in SRAtoolkit, and split into forward and reverse reads for each sample. As the data were demultiplexed paired-end sequences, a manifest file was created locally using text format accordingly (Estaki et al., 2020).

Next the QIIME 2 environment was activated (qiime2-2023.5), and the manifest file was imported into QIIME 2 (Estaki et al., 2020). The data were visualized using the *demux summarize* command to allow for data cleaning (Estaki et al., 2020). Data were trimmed using the *cutadapt*

plugin to remove primers and adaptors and to filter out short reads (Estaki et al., 2020). As much as possible, the processing followed substudy 2's processing. After this, denoising and merging of the reads was performed using the DADA2 QIIME 2 plug-in. Phylogenetic reconstruction took place using the *sepp-refs-silva-128.qza* file as reference as was used previously. Fragment-insertion tree was created and diversity metrics were generated (details in Chapter 3). The sample size was limited to the sample provided from the original analysis, so no sample size calculation was completed.

Statistical Analysis

As with substudy 2, α diversity between lesional and healthy tissue swabs was compared using Shannon's Index, and β diversity was compared between lesional and healthy tissue swabs using weighted UniFrac distances via a PERMANOVA run in QIIME 2 (Kers & Saccenti, 2022). UniFrac and weighted Unifrac distances between two samples are derived from taking into account the phylogenetic tree (Kers & Saccenti, 2022). In doing so, they are essentially measuring the phylogenetic distances between community members (Kers & Saccenti, 2022). The UniFrac distance is calculated as a fraction of the branch length (Kers & Saccenti, 2022). To create a weighted UniFrac, the branch lengths are weighted by the relative abundance of sequences, thus giving them more influence (Kers & Saccenti, 2022). To compare differential abundance taxa between lesional and healthy tissue swabs as was done in Substudy 2, ALDEx2 was used (Gloor, 2023; Gloor et

al., 2017, 2023). For all statistical tests, α was set at 0.05 (Kers & Saccenti, 2022).

5.3. Results

As described earlier, data analysed were from a study of 50 Chinese individuals diagnosed with OSCC (Zhang et al., 2020). Table 5.1 is adapted from the original article, and provides summary statistics about the sample in the study (Zhang et al., 2020).

Table 5.1 Sample summary.

Category	Level	N	%
All	All	50	100%
Clinical Stage	I	23	46%
	II	16	32%
	III	8	16%
	IV	3	6%
Alcohol Use	Previous	20	40%
	Current	17	34%
	Non-drinker	13	26%
Tobacco Smoking	Previous	17	34%
	Current	9	18%
	Non-smoker	24	48%
Betel Nut Chewing	Previous	4	8%
	Current	2	4%
	Non-chewer	44	88%

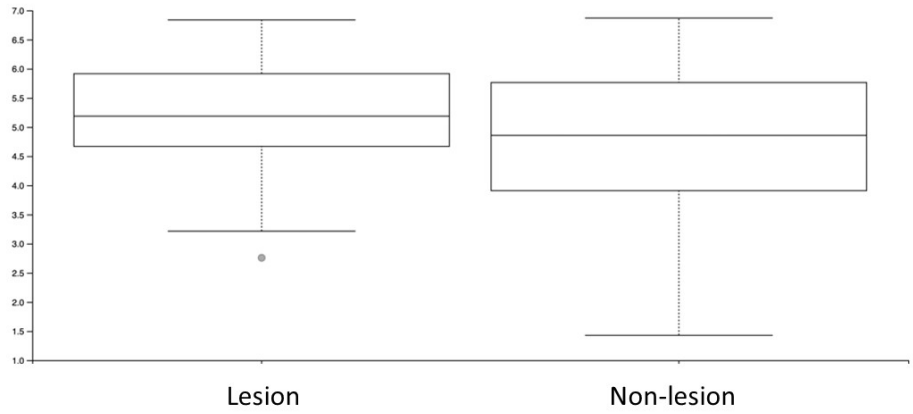
As shown in Table 5.1, the first row shows all of the members of the sample, with n = 50. Almost half the sample (46%) had lesions classified as Stage I. In Table 5.1, 74% of the sample were either previous or current alcohol users, while in substudy 2, 54% were current alcohol users, suggesting that the samples are not comparable in terms of alcohol use -as one is more than 20% greater than the other. In Table 5.1, 52% are either previous or current tobacco smokers, while in

substudy 2, 72% were either current or former tobacco users, suggesting that results from the UK sample may be more dominated by the oral microbiome of smokers compared to the Chinese sample. In Table 5.1, 12% of participants were current or previous betel nut chewers, and this is relatively similar to substudy 2, where 11% were current or former betel nut chewers.

Diversity Results

Figure 5.1 visualizes a comparison of α diversity between lesion and non-lesion tissue using Shannon's Index, which was found not to be statistically significant (p = 0.1496).

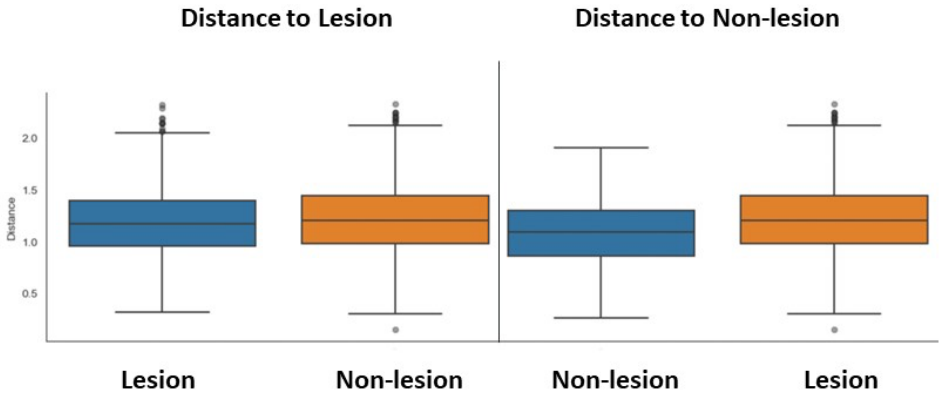
Figure 5.1 Comparison of α diversity: lesion vs. non-lesion



As seen in Figure 5.1, although the box plot shape shows there is more diversity in the non-lesional samples, the medians are close, and the test results show the α diversity between the lesional and non-lesional tissues are not statistically significantly different. These results mirror the results of substudy 2.

Figure 5.2 visualizes the comparison of β diversity between lesional and non-lesional tissues, which like with substudy 2, resulted in a statistically significant PERMANOVA test (p = 0.001).

Figure 5.2 Comparison of β diversity: lesion vs. non-lesion

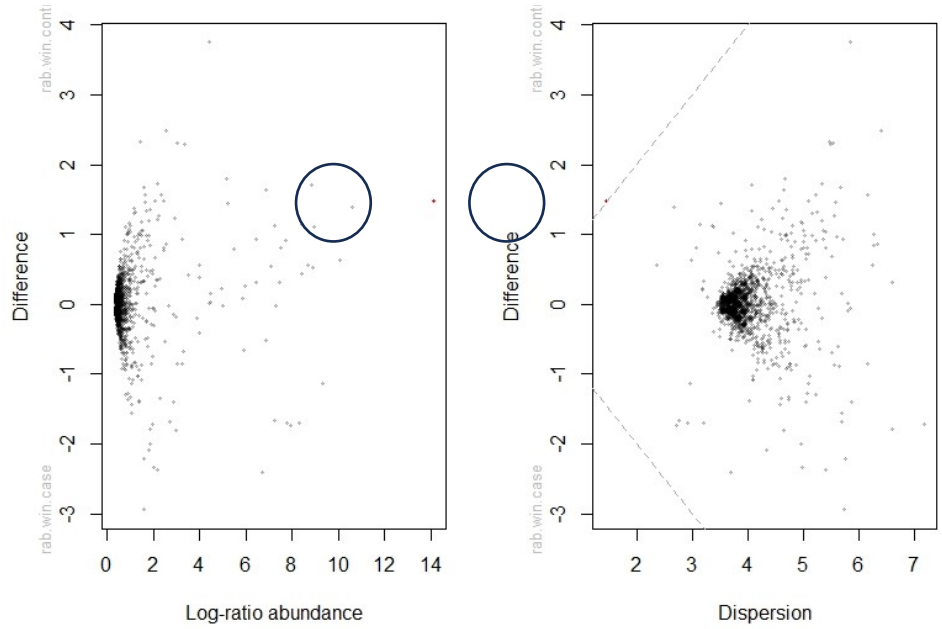


As shown in Figure 5.2, while the box plots look very similar on the left side of the figure (visualizing distance to lesion), on the right side of the figure, it appears that there is a visual difference in box plot with the distance to lesion. In other words, the distribution of distances within the lesion vs. the non-lesion were visually very similar when considering the UniFrac distances to the lesion, but were not as similar when considering the equation with a different reference in terms of distance to non-lesion. As a reminder, the distance being calculated is the weighted UniFrac distance between different microbiome community members. Although this visual difference is subtle, like with the β diversity results from substudy 2, the analysis of β diversity results in a statistically significant PERMANOVA test ($p = 0.001$).

Differential Abundance Results

The differential relative abundance analysis identified 3,692 taxa for comparison. As with substudy 2, Bland-Altman and dispersion plots were created to interpret the differential abundance results (see Figure 5.3).

Figure 5.3. Bland-Altman and dispersion plots



Although the results in Figure 5.3 look similar to the differential relative abundance results shown in substudy 2, this analysis identified one member of the microbiome that was statistically significantly differentially abundant in comparing lesional to non-lesional tissues ($q \leq 0.1$), as shown by the single red dot (circled for clarification) on Figure 5.3. Admittedly, it is difficult to identify the red dot, and the plot itself is challenging to interpret because it is mostly in black and grey. This has been acknowledged by those who use the Aldex2 approach; in cases where the distances are greater between members, the results appear much clearer in this plot (Gloor et al., 2016, 2017). With only one dot as shown here, the result is considered an artifact (Gloor et al., 2016, 2017). Recalling that black points represent rare taxa, and grey ones represent abundant taxa (Gloor et al., 2016), these results in Figure 5.3 are similar to the results of substudy 2.

As with substudy 2, for each of the 3,692 taxa identified, further analysis found that there were no taxa with Wilcoxon corrected p-values < 0.05, and there were no taxa with effect sizes with confidence intervals not including zero. For these reasons, this research fails to reject the null, and concludes that there was not a statistically significant difference in the relative differential abundance between lesional and non-lesional tissues.

5.4. Discussion

Although it would be typical to begin this discussion by presenting the results of the current reanalysis study, instead, this discussion will start by presenting the results from the study from which the reanalysed data in the current study was sourced, which was written by Zhang et al. (2020). In their study, which was set in China, they enrolled 50 participants with OSCC, and compared microbiota compositions between tumour sites and normal tissues opposite to those sites in the buccal mucosa using 16S rRNA sequencing (Zhang et al., 2020). Although inferential statistics were not applied in this study, the authors reported that they found richness and diversity of microbiota to be higher in tumour sites when compared to control tissues (Zhang et al., 2020). They have also reported some taxa to be higher in differential abundance in tumour sites as compared to healthy sites. These taxa were assigned to species level and included *Fusobacterium nucleatum*, *Prevotella intermedia*, *Aggregatibacter segnis* and other seven species.

The results of the current study, which was a reanalysis of existing data from a study by Zhang et al. (2020) designed similarly to substudy

2, are consistent with the results from substudy 2. There was not a statistically significant difference between lesional and non-lesional tissue swabs in terms of α diversity, but when comparing β diversity between lesional and non-lesional tissue swabs, there was a statistically significant difference. However, relative differential abundance analysis was unable to identify the additional taxa present in the lesion that were not present in healthy tissues and were responsible for the statistically significant increase in β diversity in the lesion tissue compared to healthy tissues.

The advantage of this comparison between the data obtained by Zhang and colleagues (2020) and the data used in substudies 1 and 2 is that they have features that make them directly comparable. First, they all used similar study designs, in terms of sampling participants, gathering measurements, and processing specimen. These features of the original studies make them comparable. Next, through the research design, there was a consistency of method imposed by pre-processing the data. This feature ensured that data from substudy 2 and 3 not only used consistent processing methods, but also used the same statistical approach. Because of these consistencies, the results of all three studies can be directly compared.

This approach provides a way to reanalyse data to make the results comparable to results from datasets from other sources. With respect to the original study by Zhang and colleagues (2020) in their scientific paper, the authors reported pre-processing the data differently than was done in the current studies presented in Chapters 3 and 4, and

as reported earlier, did not use formal hypothesis testing. The ability to reprocess the raw data from the SRA allows the researcher to impose a different processing protocol than was reported in the original scientific paper. This act will cause the newly processed data to be able to be analysed in a similar way to other data from other datasets that are processed in the same way, making them directly comparable. Likewise, imposition of a formal statistical testing approach used here allowed the results to be directly compared to Substudies 2. This led to the ability to consider if findings were consistent across studies in the current analysis, because now they could be directly compared. In the current reanalysis study, while α diversity did not differ significantly between lesional and healthy tissue samples in the same individual, β diversity differed significantly in both analyses. Yet, relative differential abundance analysis could not identify the additional taxa responsible for this finding of greater β diversity.

One of the main challenges seen in the reanalysis in the present chapter, and in the analysis done for Substudy 2 presented in Chapter 4, is that the process results in thousands of taxa being compared (as opposed to hundreds) (Zhang et al., 2020). The large number of taxa groupings becomes statistically challenging because of these large numbers. It is computationally difficult to identify differentially abundant taxa when the number of taxa groupings are in the thousands. As a thought experiment, consider calculated weighted UniFrac distances among thousands compared to hundreds of community members. The smaller the community, the easier it would be to identify clusters, as

these members would clearly separate away from the other members. But with thousands of members, the clustering would have to be extremely strong to be identified statistically.

If there are indeed differentially-abundant taxa contributing to the elevated β diversity seen in tumour tissue, then these will need to be grouped in such a way as to be identified in the analysis. The grouping approach now produces thousands of groups; without an improvement in this area, it will be difficult to find clusters statistically (Gloor et al., 2017; Kers & Saccenti, 2022). Although some studies have attempted to solve this problem by reducing the classifications to the genus or phylum level, these approaches have not been successful (Goossens et al., 2015; Su et al., 2021). These approaches have indeed reduced the classifications, but have not aided in the identification of clusters (Goossens et al., 2015; Su et al., 2021). Also, many argue this approach does not seem logical (Goossens et al., 2015; Su et al., 2021). As such classifications to higher levels, for example to phylum or even genus, would provide broader categories of taxa which might include both commensal and oncogenic taxa under the same classification.

Doddawad and colleagues (2022) summarized different approaches to classifying biomarkers in oral cancer as a way to facilitate improved prognostication and survival predication in patients. They first reviewed the biomarker classification system developed by Speight and Morgan in 1993 (Daniel & Lalitha, 2016; Doddawad et al., 2022; Speight & Morgan, 1993). In this system, biomarkers were classified into the following groups: proliferative markers, genetic markers, oncogenes,

tumour suppressor markers, cytokines, blood group antigens, and integrins extracellular matrix (ECM) ligands (Daniel & Lalitha, 2016; Doddawad et al., 2022). While this classification system may appear to be rather random, it was designed to accommodate the identified biomarkers at the time, such as histones as proliferative markers, ploidy as a genetic marker, and P53 mutations as tumour suppressor markers (Daniel & Lalitha, 2016; Doddawad et al., 2022). This classification system was improved upon by Schliephake (2003), who reclassified the biomarkers into several categories: tumour growth markers, markers of tumour invasion and metastatic potential, intracellular markers, angiogenesis markers, cells surface makers, arachidonic acid products, enzymes, markers of tumour suppression and antitumor response, and makers of anomalous keratinization (Daniel & Lalitha, 2016; Doddawad et al., 2022).

Both of these systems attempted to use current knowledge on biomarker identification and function to develop classifications (Daniel & Lalitha, 2016; Doddawad et al., 2022; Schliephake, 2003). However, classifications can be based on site, separating epithelial markers from mesenchymal markers; biomarkers can also be classified as prognostic or biochemical (Doddawad et al., 2022). Based on current knowledge Doddawad et al. (2022) propose a higher-level classification into four categories: prognostic biomarkers, biochemical markers, hormone receptors, and proliferation biomarkers. The reason why they chose four categories is they were hoping to empirically establish a more scientific way of thinking about forming hypotheses regarding biomarker detection

(Doddawad et al., 2022). However, using just four categories would not be workable for OSCC, as there are many hypothesized members of the oral microbiome that deserve investigation.

Reclassifying members of the microbiome into four categories as recommended by Doddawad et al. (2022) will not solve the challenge of comparing differential relative abundance of taxa between samples. Nevertheless, their exercise provides insight into what a more useful classification system would look like. Clearly, it would have to be based on some empirical knowledge not taken into account in the current approaches. As evidenced by papers that recommend new classification systems, scientific discussions in the published peer-reviewed literature are evolving towards developing functional classifications (Doddawad et al., 2022; Ou et al., 2021). In a functional classification system, the biomarker's classification relates to its function in the course of the disease (Doddawad et al., 2022; Ou et al., 2021).

Ou and colleagues (2021) are proponents of a functional classification system. These authors recommend identifying which biomarkers are present in relation to the course of the disease, comparing those seen in asymptomatic disease to those seen during symptomatic disease, diagnosis and treatment, and surveillance monitoring after treatment (Ou et al., 2021). Had the current study been able to leverage such classification systems, it would have been possible to develop hypothesis-driven analyses. For example, if there were proposed functional classifications of microbiome biomarkers involved in

the development of OED and OSCC, Substudy 1 could have been developed specifically to test hypotheses based on these classifications.

Unlike many of the studies reviewed in Chapter 2, the current study proposed several hypothesis-driven analyses, and the results were difficult to interpret. Although it is possible to compare α and β diversity specifically, the true aim of the line of research is to identify specific biomarkers associated with disease states, this the current study could not accomplish. Hence, the limitations of the results of the current study highlight the need to find evidence-based biomarker classification systems that can be imposed on oral microbiome data to aid in hypothesis-driven analysis. (Doddawad et al., 2022; Ou et al., 2021). As evidenced in this current study and the literature, current classification approaches do not generate clusters of taxa that relate to clinical expression of disease (Doddawad et al., 2022; Ou et al., 2021).

The current reanalysis study presented in this chapter has both strengths and limitations. The main strength of this reanalysis is that it provides a case study in how to reanalyse data from a public repository to provide a comparison to an analysis of a dataset generated from primary data collection. While this has been done by others, this adds another example. The term case study refers to a particular approach taken in a particular case. It is acknowledged that many researchers conducted case studies on the data in repositories. What is special about this case study is that it demonstrates how to reanalyse data from a repository in such a way as to make it directly comparable to data that was prospectively collected and analysed. Again, while this is novel it

has been done by others, this project provides yet another example that can be used as a guideline.

By selecting data from a study with a similar study design to Substudy 2, pre-processing the data similarly as was done in substudies 1 and 2, and conducting the statistical analysis in a similar way as was done in substudies 1 and 2, the research produced results that were directly comparable to the results presented in substudies 1 and 2. As a reminder, the reanalysed data were from a study set in China, where researchers enrolled 50 participants with OSCC (Zhang et al., 2020). In this study by Zhang et al. (2020), researchers compared microbiota compositions between tumour sites and normal tissues sampled from opposite those sites in buccal mucosal using 16S rRNA sequencing. They analysed the data using 16S rRNA sequencing and reported their results in their original paper (i.e., Zhang et al., 2020). In their original study, Zhang et al. (2020) did not use inferential studies, but the authors reported that they found the richness and diversity of microbiota were higher in tumour sites when they were compared to control tissues.

However, while the results of the reanalysed data are directly comparable to substudies 1 and 2, the limitations inherent in the original study designs are also relevant here. The sample from the UK used in substudy 1 and 2 (presented in chapter 3 and 4), and the data used from the repository for substudy 3 from China (i.e., Zhang et al., 2020) were all based on a small sample of participants with heterogenous levels of OSCC risk factors. As described in substudy 1, improvements need to be made to the study design and statistical approaches in researching

the oral microbiome if reliable biomarkers for OSCC are to be identified (Goossens et al., 2015; Hamada et al., 2019; Hu & Dignam, 2019; Ou et al., 2021; Peter et al., 2022).

In conclusion, the reanalysis of data obtained from a repository from a similarly-designed study (i.e., Zhang et al., 2020) as substudy 2 produced similar findings as substudy 2. While α diversity was found to be similar between OSCC lesion and non-lesion tissue in the same individual, there was significantly more β diversity in lesion tissue compared to non-lesion tissue. Nevertheless, specific differentially relatively abundant taxa responsible for the increased lesion β diversity were not identified, likely due to the large number of taxa being compared. In other words, no taxa were identified statistically that were differentially relatively abundant in the microbiome of OSCC lesion tissue compared to healthy tissue. Future studies should seek to use analytic approaches consistent with existing literature so results can be directly comparable. Additionally, more useful biomarker classifications based on empirical classifications should be developed in OSCC. If this can be done effectively, it can facilitate the identification of differentially abundant taxa associated with OSCC disease progression.

Chapter 6: Discussion and Conclusion

6.1. Discussion

This dissertation presents three analyses aimed at increasing knowledge about the human oral microbiome in oral squamous cell carcinoma (OSCC). In the first analysis, in substudy 1, the oral microbiome from tissue samples from a series of patients in the United Kingdom (UK) with either OSCC or oral epithelial dysplasia (OED) were analysed using 16s rRNA gene Next Generation Sequencing (NGS). Diversity and differential relative abundance were compared between lesional tissues, it was found that there was statistically significant β diversity between the OSCC and OED microbiomes. However, no differentially relatively abundant taxa were identified as being responsible for the increased diversity in the OSCC lesion.

In the second analysis in substudy 2, the same approach was used to compare the oral microbiome in OSCC and OED lesional tissues to healthy tissues in the same patients. While there was not a statistically significant difference in β diversity between OED and healthy tissues in the same individual, there was a statistically significant difference in OSCC compared to healthy tissues in the same individual. However, as with substudy 1, no differentially relatively abundant taxa were identified as being responsible for this increased diversity in the OSCC tissues.

In the third analysis in substudy 3, existing data from a Chinese study with a similar design as substudy 2 were obtained from an online repository, and reprocessed the same way data were processed in substudy 2, to make the results from this re-analysis comparable (Zhang

et al., 2020). The findings in substudy 3 were similar to substudy 2, where statistically significant β diversity was found when comparing OSCC lesional tissues to healthy tissues, but no relatively abundant taxa were identified as being present in the OSCC tissues and not present in the healthy tissues.

6.2. Consistency with Prior Studies

The three substudies in this dissertation have findings that fit together in a logical pattern. First, β diversity of the microbiome was increased in the OSCC lesions as compared to OED lesions from other participants in substudy 1. Next, β diversity of the microbiome in OSCC was increased compared to healthy tissues from the same individual in substudies 2 and 3, but differential relative abundance analysis was not able to detect the taxa responsible for this increased diversity. These findings contrast with those from the two studies from the comprehensive review in Chapter 2 whose results were interpretable statistically (Lee et al., 2017; Su et al., 2021). In one study, five genera had statistically significant differential relative abundance between the epithelial precursor lesion group and the cancer group: *Bacillus, Enterococcus, Parvimonas, Peptostreptococcus,* and *Slackia* (Lee et al., 2017). In the other study, it was found that that the genus *Fusobacterium* was relatively more abundant and genus *Streptococcus* was relatively less abundant in OSCC tumour tissue compared to control tissue (Su et al., 2021). Both of these studies were of Taiwanese cohorts, and had study design differences from the current substudies presented in Chapters 3, 4, and 5, such as their qualification criteria, type and size of sample, and

bioinformatics and statistical methods (Lee et al., 2017; Su et al., 2021). These and other differences between studies may have been responsible for these different results.

Substudy 3 represents a best practices approach when trying to compare findings between studies with similar design, because in this approach, the same methods were used for preparing and analysing the data. The methods developed in substudy 2 were applied to an existing, well-characterized dataset in substudy 3, and this facilitates an accurate comparison of findings between substudy 2 and substudy 3 (Zhang et al., 2020). Efforts like these seek to standardize bioinformatics processing and other research methods between studies to make their results directly comparable (Goossens et al., 2015; Ou et al., 2021; Zheng, 2018). Potentially, there could emerge scientific standards that would then guide microbiome biomarker studies in general that would apply to the study of OSCC (Hamada et al., 2019; Peter et al., 2022).

Nevertheless, as described at the end of Chapter 5, it still will be necessary to reduce the number of taxa included in relative differential abundance analysis in order to identify clusters, but this cannot come at the cost of misclassification (Daniel & Lalitha, 2016; Doddawad et al., 2022). An immediate improvement that could be implemented in oral biomarker studies is promoting the use of more rigorous study designs and statistical approaches as recommended through published guidance (Gloor et al., 2017; Goossens et al., 2015; Hu & Dignam, 2019; Kers & Saccenti, 2022; Ou et al., 2021). Because mastering all these knowledge domains is challenging, a new integrative transdisciplinary

field of molecular pathological epidemiology (MPE), which prepares researchers to address heterogeneous effects of different exposures on disease outcomes, should be expanded (Hamada et al., 2019).

6.3. Strengths and Limitations of this Research

Although this research did not identify a particular member of the oral microbiome associated with OSCC, negative results are not necessarily a limitation. Compared to similar studies in the scientific literature, the substudies in this dissertation have several important strengths. First, all three adhere to rigorous study design principles recommended for biomarker studies, and second, all three used formal statistical testing (Gloor et al., 2017; Goossens et al., 2015; Hu & Dignam, 2019; Kers & Saccenti, 2022; Ou et al., 2021). As part of this, the most up-to-date and evidence-based approaches were used, including QIIME 2 and ALDEx2 (Estaki et al., 2020; Gloor et al., 2023; Nearing et al., 2022). Authors of a study that compared methods for assessing differential abundance in the microbiome across 38 datasets concluded that the tools that produced the most consistent results were ALDEx2 and ANCOM-II (Nearing et al., 2022). These authors highlighted the lack of specific standards, and recommended that either ALDEx2 or ANCOM-II be a main tool for measuring differential abundance compared to the others reviewed (Nearing et al., 2022). They discussed whether or not multiple differential abundance analyses should be conducted in the same study using multiple tools, and while they suggested it may improve consensus, they underscored the findings from their analysis, which is the results are inconsistent so they do not

facilitate a consensus (Nearing et al., 2022). Further, there are issues with *post hoc* analysis; only one tool should be used to answer hypotheses developed *a priori* (Gloor et al., 2017; Kers & Saccenti, 2022). Next, the results of the three substudies were compatible, even when including an external dataset as was done in substudy 3 (Zhang et al., 2020).

In addition to strengths, there are also many limitations to this research that could not be overcome. Many of these limitations were identified throughout the literature in Chapter 2. First, no significantly differentially abundant taxa or clusters of taxa were found in any of these studies. However, in consideration of the increased β diversity differences seen with statistical testing, one could speculate that this null finding might be due to study design features (such as small sample size, or high levels of heterogeneity in the sample). If there had been more sample, this may not have been the result, but more sample was not available. Next, as mentioned in the previous paragraph, due to timeline considerations, the sample analysed in substudies 1 and 2 was not ideal. The samples were heterogeneous with respect to risk factors for OSCC, which means that they had different distributions of risk factors for OSCC. Ideally, all of the samples would have included enough individuals who were positive for strong risk factors so that a subgroup analysis could take place. Alternatively, a stronger study design feature would have included small samples that focused on specific populations with known risk factors (e.g., studies restricted to tobacco users only). Because the samples included were both small and heterogeneous, it

decreased the power to be able to detect unique microbiome profiles. This also decreases the validity of the results. Finally, it is acknowledged that using different bioinformatics and statistical approaches may produce different results, as has been demonstrated in simulations (Gloor et al., 2016; Nearing et al., 2022). Therefore, the results of the substudies presented in this project should be understood to be tentative, and would need to be replicated with a similar study design before being considered robust findings. This is because scientifically, it is considered necessary to validate findings across rigorous studies before results can be accepted as scientifically factual (Kers & Saccenti, 2022).

6.4. Future Research Directions

The three substudies described in this thesis did not identify any differentially relatively abundant taxa present in OSCC tissues compared to other tissues in this sample of UK patients. The β diversity results from the analyses suggest that it may be possible to identify genera associated specifically with OSCC. Because the sample was small and heterogeneous with respect to OSCC risk factors, the power to detect differences was low. If a hypothesis-driven study was developed focusing on candidate members of the microbiome, it would have more statistical power to detect differences. Also, studies with more homogenous samples and larger samples would have more statistical power to detect differences. As described in Chapter 2, Lee and colleagues (2017) identified *Bacillus, Enterococcus, Parvimonas, Peptostreptococcus,* and *Slackia* as potential biomarker targets in

OSCC, and Su and colleagues (2021) found that *Fusobacterium* was relatively more abundant and *Streptococcus* was relatively less abundant in OSCC tissue compared to control, identifying two additional biomarker targets. Although the current studies presented in this thesis did not identify any differentially abundant taxa, as a next step, future studies could seek to validate the results from these other studies (Lee et al., 2017; Su et al., 2021). This might be by quantifying the presence of these identified biomarkers, for example *Fusobacterium* in OSCC tissues of different cohort. Researchers conducting exploratory studies like the ones in the three substudies in this dissertation should also consider ways to improve the classification of taxa into functional groups to reduce the number of taxa analysed, thus producing results that are less noisy and easier to interpret (Doddawad et al., 2022).

When reviewing the datasets in the SRA, it was found that many could not be used because the documentation about the data and informatics processing was not sufficient. Some datasets could not be used because the variables in the actual datasets shared were not clear. Even the cleanest, most appropriate and well-documented dataset that was selected for reanalysis presented many logistical data-cleaning challenges (Zhang et al., 2020). If microbiome repositories professionalize their dataset presentation through including the necessary documentation and curation, this would greatly facilitate future research in this area. Such an upgrade would allow authors to easily re-analyse different external datasets using a standardised protocol with appropriate bioinformatics and statistical processing. Such

action would greatly promote comparison to current studies, as was done with substudies 2 and 3 in this project. This both conserves research resources and expand the scientific evidence base.

Once OSCC biomarker research advances to the point that reliable biomarkers are identified that are specific to OSCC, then efforts at developing standards will need to take place, as recommended in the article where authors compared differential abundance measures across 38 datasets (Hu & Dignam, 2019; Nearing et al., 2022; Zaura et al., 2021; Zheng, 2018). Consensus across the scientific domain will need to be agreed-upon as to how specific biomarkers are measured, what they mean clinically, and how to make treatment decisions based on their values (Hu & Dignam, 2019; Zaura et al., 2021; Zheng, 2018). Until that point, research methods into OSCC biomarkers will need to be transparent and replicable so as achieve a level of standardization that would promote arriving at such a consensus.

Overall, future research direction could be summarised into four overarching key points. First, there is a need for the development of OSSC microbiome study guidelines for researchers reporting their results. These could be developed by a consensus team of multi-disciplinary researchers including epidemiologists, statisticians, clinicians, oral microbiome microbiologists, and bioinformaticians. Second, multi-centre OSSC microbiome studies conducted by multi-disciplinary research teams are needed. These studies would include large sample sizes and homogenous groups of participants, and would follow a standardised research protocol. These features would make it

more possible to identify differences in the oral microbiome between subgroups. Third, researchers using hypothesis-driven inquiries should focus on identifying candidate members of the oral microbiome that have oncogenic potentials (e.g, *Fusobacterium nucleatum*). Identifying and publicizing these candidates will help guide researchers as to priority biomarkers to assess. Finally, *in silico* and laboratory studies mapping genetic and epigenetic changes in OSCC should seek to characterise the potential pathogenicity of members of the oral microbiome.

6.5. Conclusion

In conclusion, this dissertation presented three substudies seeking to identify members of the oral microbiome, using 16s rRNA gene NGS, that play an important role in the development of OSCC. The results of the thee substudies suggest that the diversity of the bacteriome of OSSC lesion was increased compared to OED and normal tissues. The exact differences in the members of the bacteriome between OSCC as compared to that of OED and normal controls should be identified. However, at the present time, there are many logistical barriers to moving forward with such research, including a lack of transdisciplinary knowledge of research and statistical methods required to impart rigor to such studies (Gloor et al., 2017; Hamada et al., 2019; Ou et al., 2021; Zheng, 2018). Improvements in the quality of microbiome research repositories could greatly facilitate the progress of research in the area of the OSCC microbiome because it would promote the standardisation of methodology from study to study (T. Chen et al., 2010; *HOMD ::*

Human Oral Microbiome Database, n.d.; Human Microbiome Project, n.d.).

Ultimately, the aim of OSCC microbiome research is to improve the diagnostic and prognostic capabilities and therapeutic opportunities afforded OSCC patient population (Hu & Dignam, 2019). Research that is ultimately successful at identifying consistent, evidence-based OSCC biomarkers in the oral microbiome could result in earlier detection, better prognosis, and superior treatment of OSCC patients worldwide (Gormley et al., 2022; Hu & Dignam, 2019; Salehiniya & Raei, 2020; Zaura et al., 2021).

References

Al-Hebshi, N. N., Nasher, A. T., Idris, A. M., & Chen, T. (2015). Robust species taxonomy assignment algorithm for 16S rRNA NGS reads: Application to oral carcinoma samples. *Journal of Oral Microbiology*, *7*, 10.3402/jom.v7.28934. https://doi.org/10.3402/jom.v7.28934

Al-Hebshi, N. N., Nasher, A. T., Maryoud, M. Y., Homeida, H. E., Chen, T., Idris, A. M., & Johnson, N. W. (2017). Inflammatory bacteriome featuring Fusobacterium nucleatum and Pseudomonas aeruginosa identified in association with oral squamous cell carcinoma. *Scientific Reports*, *7*(1), 1834. https://doi.org/10.1038/s41598-017-02079-3

Ambatipudi, S., Cuenin, C., Hernandez-Vargas, H., Ghantous, A., Le Calvez-Kelm, F., Kaaks, R., Barrdahl, M., Boeing, H., Aleksandrova, K., Trichopoulou, A., Lagiou, P., Naska, A., Palli, D., Krogh, V., Polidoro, S., Tumino, R., Panico, S., Bueno-de-Mesquita, B., Peeters, P. H., … Herceg, Z. (2016). Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. *Epigenomics*, *8*(5), 599–618. https://doi.org/10.2217/epi-2016-0001

American Statistical Association. (2022). *Ethical Guidelines for Statistical Practice*. Default. https://www.amstat.org/your-career/ethical-guidelines-for-statistical-practice

Batta, N., & Pandey, M. (2019). Mutational spectrum of tobacco associated oral squamous carcinoma and its therapeutic

significance. *World Journal of Surgical Oncology*, *17*, 198. https://doi.org/10.1186/s12957-019-1741-2

Bezerra, C. T., Grande, A. J., Galvão, V. K., dos Santos, D. H. M., Atallah, Á. N., & Silva, V. (2022). Assessment of the strength of recommendation and quality of evidence: GRADE checklist. A descriptive study. *São Paulo Medical Journal*, *140*(6), 829–836. https://doi.org/10.1590/1516-3180.2022.0043.R1.07042022

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bosetti, C., Carioli, G., Santucci, C., Bertuccio, P., Gallus, S., Garavello, W., Negri, E., & La Vecchia, C. (2020). Global trends in oral and pharyngeal cancer incidence and mortality. *International Journal of Cancer*, *147*(4), 1040–1049. https://doi.org/10.1002/ijc.32871

Brooks, J. P. (2016). Challenges for case-control studies with microbiome data. *Annals of Epidemiology*, *26*(5), 336-341.e1. https://doi.org/10.1016/j.annepidem.2016.03.009

Bugshan, A., & Farooq, I. (2020). Oral squamous cell carcinoma: Metastasis, potentially associated malignant disorders, etiology

and recent advancements in diagnosis. *F1000Research*, *9*, 229. https://doi.org/10.12688/f1000research.22941.1

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, *11*(12), 2639–2643. https://doi.org/10.1038/ismej.2017.119

Chang, C., Geng, F., Shi, X., Li, Y., Zhang, X., Zhao, X., & Pan, Y. (2019). The prevalence rate of periodontal pathogens and its association with oral squamous cell carcinoma. *Applied Microbiology and Biotechnology*, *103*(3), 1393–1404. https://doi.org/10.1007/s00253-018-9475-6

Chen, J.-W., Wu, J.-H., Chiang, W.-F., Chen, Y.-L., Wu, W.-S., & Wu, L.-W. (2021). Taxonomic and functional dysregulation in salivary microbiomes during oral carcinogenesis. *Frontiers in Cellular and Infection Microbiology*, *11*, 663068. https://doi.org/10.3389/fcimb.2021.663068

Chen, T., Yu, W.-H., Izard, J., Baranova, O. V., Lakshmanan, A., & Dewhirst, F. E. (2010). The Human Oral Microbiome Database: A web accessible resource for investigating oral microbe taxonomic and genomic information. *Database: The Journal of Biological Databases and Curation*, *2010*, baq013. https://doi.org/10.1093/database/baq013

Cristaldi, M., Mauceri, R., Di Fede, O., Giuliana, G., Campisi, G., & Panzarella, V. (2019). Salivary biomarkers for oral squamous cell carcinoma diagnosis and follow-up: Current status and

perspectives. *Frontiers in Physiology*, *10*, 1476. https://doi.org/10.3389/fphys.2019.01476

Daniel, D., & Lalitha, R. M. (2016). Tumor markers – A bird's eye view. *Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology*, *28*(6), 475–480. https://doi.org/10.1016/j.ajoms.2016.07.006

de la Iglesia, J. V., Slebos, R. J. C., Martin-Gomez, L., Wang, X., Teer, J. K., Tan, A. C., Gerke, T. A., Aden-Buie, G., Van Veen, T., Masannat, J., Chaudhary, R., Song, F., Fournier, M., Siegel, E. M., Schabath, M. B., Wadsworth, J. T., Caudell, J., Harrison, L., Wenig, B. M., … Chung, C. H. (2020). Effects of tobacco smoking on the tumor immune microenvironment in head and neck squamous cell carcinoma. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *26*(6), 1474–1485. https://doi.org/10.1158/1078-0432.CCR-19-1769

Doddawad, V. G., Bannimath, G., Shivakumar, S., & Bannimath, N. (2022). Biomarkers of oral cancer: A current views and directions. *Biomedical and Biotechnology Research Journal (BBRJ)*, *6*(1), 33. https://doi.org/10.4103/bbrj.bbrj_204_21

Ernani, V., & Saba, N. F. (2015). Oral cavity cancer: Risk factors, pathology, and management. *Oncology*, *89*(4), 187–195. https://doi.org/10.1159/000398801

Escapa, I. F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F. E., & Lemon, K. P. (2018). New insights into human nostril microbiome from the Expanded Human Oral Microbiome Database (eHOMD): A

resource for the microbiome of the human aerodigestive tract. *mSystems*, *3*(6), e00187-18. https://doi.org/10.1128/mSystems.00187-18

Estaki, M., Jiang, L., Bokulich, N. A., McDonald, D., González, A., Kosciolek, T., Martino, C., Zhu, Q., Birmingham, A., Vázquez-Baeza, Y., Dillon, M. R., Bolyen, E., Caporaso, J. G., & Knight, R. (2020). QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. *Current Protocols in Bioinformatics*, *70*(1). https://doi.org/10.1002/cpbi.100

Falagas, M., Pitsouni, E., Malietzis, G., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *FASEB Journal*, *22*(2), 338–342. https://doi.org/10.1096/fj.07-9492lsf

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer*, *149*(4), 778–789. https://doi.org/10.1002/ijc.33588

Gaziano, L., Giambartolomei, C., Pereira, A. C., Gaulton, A., Posner, D. C., Swanson, S. A., Ho, Y.-L., Iyengar, S. K., Kosik, N. M., Vujkovic, M., Gagnon, D. R., Bento, A. P., Barrio-Hernandez, I., Rönnblom, L., Hagberg, N., Lundtoft, C., Langenberg, C., Pietzner, M., Valentine, D., … VA Million Veteran Program COVID-19 Science Initiative. (2021). Actionable druggable genome-wide Mendelian randomization identifies repurposing

opportunities for COVID-19. *Nature Medicine*, *27*(4), 668–676. https://doi.org/10.1038/s41591-021-01310-z

Gehanno, J.-F., Rollin, L., & Darmoni, S. (2013). Is the coverage of Google Scholar enough to be used alone for systematic reviews. *BMC Medical Informatics and Decision Making*, *13*, 7. https://doi.org/10.1186/1472-6947-13-7

Ghantous, Y., Schussel, J. L., & Brait, M. (2018). Tobacco and alcohol induced epigenetic changes in oral carcinoma. *Current Opinion in Oncology*, *30*(3), 152–158. https://doi.org/10.1097/CCO.0000000000000444

Gibbons, S. M., Duvallet, C., & Alm, E. J. (2018). Correcting for batch effects in case-control microbiome studies. *PLoS Computational Biology*, *14*(4), e1006102. https://doi.org/10.1371/journal.pcbi.1006102

Gloor, G. B. (2023). *ANOVA-Like differential expression tool for high throughput sequencing data*. Bioconductor. https://www.bioconductor.org/packages/devel/bioc/vignettes/ALDEx2/inst/doc/ALDEx2_vignette.html

Gloor, G. B., Fern, A., es, Macklaim, J., Albert, A., Links, M., Quinn, T., Wu, J. R., Wong, R. G., Br, & Lieng, on. (2023). *ALDEx2: Analysis of differential abundance taking sample variation into account* (Version 1.32.0) [Computer software]. Bioconductor version: Release (3.17). https://doi.org/10.18129/B9.bioc.ALDEx2

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, *8*, 2224. https://doi.org/10.3389/fmicb.2017.02224

Gloor, G. B., Macklaim, J. M., Vu, M., & Fernandes, A. D. (2016). Compositional uncertainty should not be ignored in high-throughput sequencing data analysis. *Austrian Journal of Statistics*, *45*(4), 73–87. https://doi.org/10.17713/ajs.v45i4.122

Goodrich, J. K., Di Rienzi, S. C., Poole, A. C., Koren, O., Walters, W. A., Caporaso, J. G., Knight, R., & Ley, R. E. (2014). Conducting a microbiome study. *Cell*, *158*(2), 250–262. https://doi.org/10.1016/j.cell.2014.06.037

Goossens, N., Nakagawa, S., Sun, X., & Hoshida, Y. (2015). Cancer biomarker discovery and validation. *Translational Cancer Research*, *4*(3), 256–269. https://doi.org/10.3978/j.issn.2218-676X.2015.06.04

Gopinath, D., Kunnath Menon, R., Chun Wie, C., Banerjee, M., Panda, S., Mandal, D., Behera, P. K., Roychoudhury, S., Kheur, S., George Botelho, M., & Johnson, N. W. (2021). Salivary bacterial shifts in oral leukoplakia resemble the dysbiotic oral cancer bacteriome. *Journal of Oral Microbiology*, *13*(1), 1857998. https://doi.org/10.1080/20002297.2020.1857998

Gopinath, D., Menon, R. K., Wie, C. C., Banerjee, M., Panda, S., Mandal, D., Behera, P. K., Roychoudhury, S., Kheur, S., Botelho, M. G., & Johnson, N. W. (2021). Differences in the bacteriome of swab,

saliva, and tissue biopsies in oral cancer. *Scientific Reports*, *11*(1), 1181. https://doi.org/10.1038/s41598-020-80859-0

Gormley, M., Creaney, G., Schache, A., Ingarfield, K., & Conway, D. I. (2022). Reviewing the epidemiology of head and neck cancer: Definitions, trends and risk factors. *British Dental Journal*, *233*(9), 780–786. https://doi.org/10.1038/s41415-022-5166-x

Granato, D. C., Neves, L. X., Trino, L. D., Carnielli, C. M., Lopes, A. F. B., Yokoo, S., Pauletti, B. A., Domingues, R. R., Sá, J. O., Persinoti, G., Paixão, D. A. A., Rivera, C., De Sá Patroni, F. M., Tommazetto, G., Santos-Silva, A. R., Lopes, M. A., De Castro, G., Brandão, T. B., Prado-Ribeiro, A. C., … Paes Leme, A. F. (2021). Meta-omics analysis indicates the saliva microbiome and its proteins associated with the prognosis of oral cancer patients. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1869*(8), 140659. https://doi.org/10.1016/j.bbapap.2021.140659

Grantham, N. S., Guan, Y., Reich, B. J., Borer, E. T., & Gross, K. (2020). MIMIX: A Bayesian mixed-effects model for microbiome data From designed experiments. *Journal of the American Statistical Association*, *115*(530), 599–609. https://doi.org/10.1080/01621459.2019.1626242

Hamada, T., Nowak, J. A., Milner, D. A., Song, M., & Ogino, S. (2019). Integration of microbiology, molecular pathology, and epidemiology: A new paradigm to explore the pathogenesis of microbiome-driven neoplasms. *The Journal of Pathology*, *247*(5), 615–628. https://doi.org/10.1002/path.5236

Harrison, R., Jones, B., Gardner, P., & Lawton, R. (2021). Quality assessment with diverse studies (QuADS): An appraisal tool for methodological and reporting quality in systematic reviews of mixed- or multi-method studies. *BMC Health Services Research*, *21*, 144. https://doi.org/10.1186/s12913-021-06122-y

Hashim, D., Genden, E., Posner, M., Hashibe, M., & Boffetta, P. (2019). Head and neck cancer prevention: From primary prevention to impact of clinicians on reducing burden. *Annals of Oncology*, *30*(5), 744–756. https://doi.org/10.1093/annonc/mdz084

Hashimoto, K., Shimizu, D., Hirabayashi, S., Ueda, S., Miyabe, S., Oh-iwa, I., Nagao, T., Shimozato, K., & Nomoto, S. (2019). Changes in oral microbial profiles associated with oral squamous cell carcinoma vs leukoplakia. *Journal of Investigative and Clinical Dentistry*, *10*(4). https://doi.org/10.1111/jicd.12445

Hemminki, O., dos Santos, J. M., & Hemminki, A. (2020). Oncolytic viruses for cancer immunotherapy. *Journal of Hematology & Oncology*, *13*, 84. https://doi.org/10.1186/s13045-020-00922-1

Hernandez, B. Y., Zhu, X., Goodman, M. T., Gatewood, R., Mendiola, P., Quinata, K., & Paulino, Y. C. (2017). Betel nut chewing, oral premalignant lesions, and the oral microbiome. *PLoS ONE*, *12*(2), e0172196. https://doi.org/10.1371/journal.pone.0172196

HOMD :: Human Oral Microbiome Database. (n.d.). Retrieved August 9, 2023, from https://www.homd.org/

Hooper, S. J., Crean, S. J., Lewis, M. A. O., Spratt, D. A., Wade, W. G., & Wilson, M. J. (2006). Viable bacteria present within oral

squamous cell carcinoma tissue. *Journal of Clinical Microbiology*, *44*(5), 1719–1725. https://doi.org/10.1128/JCM.44.5.1719-1725.2006

Hooper, S. J., Crean, S.-J., Fardy, M. J., Lewis, M. A. O., Spratt, D. A., Wade, W. G., & Wilson, M. J. (2007). A molecular analysis of the bacteria present within oral squamous cell carcinoma. *Journal of Medical Microbiology*, *56*(Pt 12), 1651–1659. https://doi.org/10.1099/jmm.0.46918-0

Hu, C., & Dignam, J. J. (2019). Biomarker-driven oncology clinical trials: Key design elements, types, features, and practical considerations. *JCO Precision Oncology*, *3*, PO.19.00086. https://doi.org/10.1200/PO.19.00086

Huber, M. A., & Tantiwongkosi, B. (2014). Oral and oropharyngeal cancer. *The Medical Clinics of North America*, *98*(6), 1299–1321. https://doi.org/10.1016/j.mcna.2014.08.005

Human Microbiome Project. (n.d.). *iHMP Data Portal*. Retrieved August 13, 2023, from https://portal.hmpdacc.org/

Janssen, S., McDonald, D., Gonzalez, A., Navas-Molina, J. A., Jiang, L., Xu, Z. Z., Winker, K., Kado, D. M., Orwoll, E., Manary, M., Mirarab, S., & Knight, R. (2018). Phylogenetic placement of exact amplicon sequences improves associations with clinical information. *mSystems*, *3*(3), e00021-18. https://doi.org/10.1128/mSystems.00021-18

Jethwa, A. R., & Khariwala, S. S. (2017). Tobacco-related carcinogenesis in head and neck cancer. *Cancer Metastasis*

Reviews, *36*(3), 411–423. https://doi.org/10.1007/s10555-017-9689-6

Jiang, X., Wu, J., Wang, J., & Huang, R. (2019). Tobacco and oral squamous cell carcinoma: A review of carcinogenic pathways. *Tobacco Induced Diseases*, *17*, 29. https://doi.org/10.18332/tid/105844

Kerr, S. (2016, November 23). Honeymoon is over for new Saudi leader as reform pain kicks in. *Financial Times*. https://www.ft.com/content/6526fbfe-b090-11e6-a37c-f4a01f1b0fa1

Kers, J. G., & Saccenti, E. (2022). The power of microbiome studies: Some considerations on which alpha and beta metrics to use and how to report results. *Frontiers in Microbiology*, *12*. https://www.frontiersin.org/articles/10.3389/fmicb.2021.796025

Lee, W.-H., Chen, H.-M., Yang, S.-F., Liang, C., Peng, C.-Y., Lin, F.-M., Tsai, L.-L., Wu, B.-C., Hsin, C.-H., Chuang, C.-Y., Yang, T., Yang, T.-L., Ho, S.-Y., Chen, W.-L., Ueng, K.-C., Huang, H.-D., Huang, C.-N., & Jong, Y.-J. (2017). Bacterial alterations in salivary microbiota and their association in oral cancer. *Scientific Reports*, *7*(1), 16540. https://doi.org/10.1038/s41598-017-16418-x

Leung, K. (2011). Presenting post hoc hypotheses as a priori: Ethical and theoretical issues. *Management and Organization Review*, *7*(3), 471–479.

Li, Y., Tan, X., Zhao, X., Xu, Z., Dai, W., Duan, W., Huang, S., Zhang, E., Liu, J., Zhang, S., Yin, R., Shi, X., Lu, Z., & Pan, Y. (2020).

Composition and function of oral microbiota between gingival squamous cell carcinoma and periodontitis. *Oral Oncology*, *107*, 104710. https://doi.org/10.1016/j.oraloncology.2020.104710

Liu, B., Shen, M., Xiong, J., Yuan, Y., Wu, X., Gao, X., Xu, J., Guo, F., & Jian, X. (2015). Synergistic effects of betel quid chewing, tobacco use (in the form of cigarette smoking), and alcohol consumption on the risk of malignant transformation of oral submucous fibrosis (OSF): A case-control study in Hunan Province, China. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, *120*(3), 337–345. https://doi.org/10.1016/j.oooo.2015.04.013

Macey, R., Walsh, T., Brocklehurst, P., Kerr, A. R., Liu, J. L. Y., Lingen, M. W., Ogden, G. R., Warnakulasuriya, S., & Scully, C. (2015). Diagnostic tests for oral cancer and potentially malignant disorders in patients presenting with clinically evident lesions. *The Cochrane Database of Systematic Reviews*, *2015*(5), CD010276. https://doi.org/10.1002/14651858.CD010276.pub2

Mager, D. L., Haffajee, A. D., Devlin, P. M., Norris, C. M., Posner, M. R., & Goodson, J. M. (2005). The salivary microbiota as a diagnostic indicator of oral cancer: A descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. *Journal of Translational Medicine*, *3*, 27. https://doi.org/10.1186/1479-5876-3-27

Makin, T. R., & Orban de Xivry, J.-J. (2019). Ten common statistical mistakes to watch out for when writing or reviewing a manuscript. *eLife*, *8*, e48175. https://doi.org/10.7554/eLife.48175

Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., & Peddada, S. D. (2015). Analysis of composition of microbiomes: A novel method for studying microbial composition. *Microbial Ecology in Health and Disease*, *26*, 10.3402/mehd.v26.27663. https://doi.org/10.3402/mehd.v26.27663

Markowitz, V. M., Chen, I.-M. A., Chu, K., Szeto, E., Palaniappan, K., Jacob, B., Ratner, A., Liolios, K., Pagani, I., Huntemann, M., Mavromatis, K., Ivanova, N. N., & Kyrpides, N. C. (2012). IMG/M-HMP: A metagenome comparative analysis system for the Human Microbiome Project. *PLoS ONE*, *7*(7), e40151. https://doi.org/10.1371/journal.pone.0040151

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160–1177. https://doi.org/10.1016/j.joi.2018.09.002

Marur, S., & Forastiere, A. A. (2016). Head and neck squamous cell carcinoma: Update on epidemiology, diagnosis, and treatment. *Mayo Clinic Proceedings*, *91*(3), 386–396. https://doi.org/10.1016/j.mayocp.2015.12.017

Mello, F. W., Melo, G., Pasetto, J. J., Silva, C. A. B., Warnakulasuriya, S., & Rivero, E. R. C. (2019). The synergistic effect of tobacco

and alcohol consumption on oral squamous cell carcinoma: A systematic review and meta-analysis. *Clinical Oral Investigations*, *23*(7), 2849–2859. https://doi.org/10.1007/s00784-019-02958-1

Migliavaca, C. B., Stein, C., Colpani, V., Munn, Z., Falavigna, M., & Prevalence Estimates Reviews – Systematic Review Methodology Group (PERSyst). (2020). Quality assessment of prevalence studies: A systematic review. *Journal of Clinical Epidemiology*, *127*, 59–68. https://doi.org/10.1016/j.jclinepi.2020.06.039

Miranda-Filho, A., & Bray, F. (2020). Global patterns and trends in cancers of the lip, tongue and mouth. *Oral Oncology*, *102*, 104551. https://doi.org/10.1016/j.oraloncology.2019.104551

Mok, S. F., Karuthan, C., Cheah, Y. K., Ngeow, W. C., Rosnah, Z., Yap, S. F., & Ong, H. K. A. (2017). The oral microbiome community variations associated with normal, potentially malignant disorders and malignant lesions of the oral cavity. *The Malaysian Journal of Pathology*, *39*(1), 1–15.

Moskalewicz, A., & Oremus, M. (2020). No clear choice between Newcastle-Ottawa Scale and Appraisal Tool for Cross-Sectional Studies to assess methodological quality in cross-sectional studies of health-related quality of life and breast cancer. *Journal of Clinical Epidemiology*, *120*, 94–103. https://doi.org/10.1016/j.jclinepi.2019.12.013

Mukherjee, P. K., Wang, H., Retuerto, M., Zhang, H., Burkey, B., Ghannoum, M. A., & Eng, C. (2017). Bacteriome and mycobiome

associations in oral tongue cancer. *Oncotarget*, *8*(57), 97273–97289. https://doi.org/10.18632/oncotarget.21921

Nagy, K., Sonkodi, I., Kovács, A., Szöke, I., Mari, A., & Nagy, E. (1998). [Microbial study of the surface of malignant tumors of the oral cavity. *Fogorvosi szemle*, *91*(8–9), 281–284.

Namburi, P. R., Mahalakshmi, K., Sankari, S. L., & Kumar, V. N. (2020). Low abundance of capnophiles in the saliva of oral squamous cell carcinoma patients: A metagenomic analysis. *European Journal of Molecular and Clinical Medicine*, *7*(5), 1375–1379.

National Institutes of Health. (n.d.). *The Sequence Read Archive (SRA)*. Retrieved September 18, 2023, from https://www.ncbi.nlm.nih.gov/sra/docs/

Nearing, J. T., Douglas, G. M., Hayes, M. G., MacDonald, J., Desai, D. K., Allward, N., Jones, C. M. A., Wright, R. J., Dhanani, A. S., Comeau, A. M., & Langille, M. G. I. (2022). Microbiome differential abundance methods produce different results across 38 datasets. *Nature Communications*, *13*(1), 342. https://doi.org/10.1038/s41467-022-28034-z

Nguyen, N.-P., Warnow, T., Pop, M., & White, B. (2016). A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *Npj Biofilms and Microbiomes*, *2*(1), Article 1. https://doi.org/10.1038/npjbiofilms.2016.4

NIH Human Microbiome Portfolio Analysis Team. (2019). A review of 10 years of human microbiome research activities at the US

National Institutes of Health, Fiscal Years 2007-2016. *Microbiome*, *7*, 31. https://doi.org/10.1186/s40168-019-0620-y

Odell, E., Kujan, O., Warnakulasuriya, S., & Sloan, P. (2021). Oral epithelial dysplasia: Recognition, grading and clinical significance. *Oral Diseases*, *27*(8), 1947–1976. https://doi.org/10.1111/odi.13993

Ogoke, U. P. (2023). Determination of the receiver operating characteristics (ROC) curve of the logistic regression model accuracy using some breast measurements in the presence of multicollinearity. *IPS Journal of Public Health*, *3*(1), Article 1. https://doi.org/10.54117/ijph.v3i1.11

Ou, F.-S., Michiels, S., Shyr, Y., Adjei, A. A., & Oberg, A. L. (2021). Biomarker discovery and validation: Statistical considerations. *Journal of Thoracic Oncology*, *16*(4), 537–545. https://doi.org/10.1016/j.jtho.2021.01.1616

Perera, M., Al-hebshi, N. N., Perera, I., Ipe, D., Ulett, G. C., Speicher, D. J., Chen, T., & Johnson, N. W. (2017). A dysbiotic mycobiome dominated by Candida albicans is identified within oral squamous-cell carcinomas. *Journal of Oral Microbiology*, *9*(1), 1385369. https://doi.org/10.1080/20002297.2017.1385369

Perera, M., Al-Hebshi, N. N., Perera, I., Ipe, D., Ulett, G. C., Speicher, D. J., Chen, T., & Johnson, N. W. (2018). Inflammatory bacteriome and oral squamous cell carcinoma. *Journal of Dental Research*, *97*(6), 725–732. https://doi.org/10.1177/0022034518767118

Peter, T. K., Withanage, M. H. H., Comnick, C. L., Pendleton, C., Dabdoub, S., Ganesan, S., Drake, D., Banas, J., Xie, X. J., & Zeng, E. (2022). Systematic review and meta-analysis of oral squamous cell carcinoma associated oral microbiome. *Frontiers in Microbiology*, *13*, 968304. https://doi.org/10.3389/fmicb.2022.968304

Pushalkar, S., Ji, X., Li, Y., Estilo, C., Yegnanarayana, R., Singh, B., Li, X., & Saxena, D. (2012). Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. *BMC Microbiology*, *12*, 144. https://doi.org/10.1186/1471-2180-12-144

Pushalkar, S., Mane, S. P., Ji, X., Li, Y., Evans, C., Crasta, O. R., Morse, D., Meagher, R., Singh, A., & Saxena, D. (2011). Microbial diversity in saliva of oral squamous cell carcinoma. *FEMS Immunology & Medical Microbiology*, *61*(3), 269–277. https://doi.org/10.1111/j.1574-695X.2010.00773.x

Radoï, L., & Luce, D. (2013). A review of risk factors for oral cavity cancer: The importance of a standardized case definition. *Community Dentistry and Oral Epidemiology*, *41*(2), 97–109. https://doi.org/10.1111/j.1600-0528.2012.00710.x

Rai, A. K., Panda, M., Das, A. K., Rahman, T., Das, R., Das, K., Sarma, A., Kataki, A. Ch., & Chattopadhyay, I. (2021). Dysbiosis of salivary microbiome and cytokines influence oral squamous cell carcinoma through inflammation. *Archives of Microbiology*, *203*(1), 137–152. https://doi.org/10.1007/s00203-020-02011-w

Salehiniya, H., & Raei, M. (2020). Oral cavity and lip cancer in the world: An epidemiological review. *Biomedical Research and Therapy*, *7*(8), Article 8. https://doi.org/10.15419/bmrat.v7i8.619

Sarkar, P., Malik, S., Laha, S., Das, S., Bunk, S., Ray, J. G., Chatterjee, R., & Saha, A. (2021). Dysbiosis of oral microbiota during oral squamous cell carcinoma development. *Frontiers in Oncology*, *11*, 614448. https://doi.org/10.3389/fonc.2021.614448

Sawant, S., Dugad, J., Parikh, D., Srinivasan, S., & Singh, H. (2021). Identification and correlation of bacterial diversity in oral cancer and long-term tobacco chewers- A case-control pilot study. *Journal of Medical Microbiology*, *70*(9). https://doi.org/10.1099/jmm.0.001417

Schifano, E. D. (2019). A review of analysis methods for secondary outcomes in case-control studies. *Communications for Statistical Applications and Methods*, *26*(2), 103–129. https://doi.org/10.29220/CSAM.2019.26.2.103

Schliephake, H. (2003). Prognostic relevance of molecular markers of oral cancer—A review. *International Journal of Oral and Maxillofacial Surgery*, *32*(3), 233–245. https://doi.org/10.1054/ijom.2002.0383

Schmidt, B. L., Kuczynski, J., Bhattacharya, A., Huey, B., Corby, P. M., Queiroz, E. L. S., Nightingale, K., Kerr, A. R., DeLacure, M. D., Veeramachaneni, R., Olshen, A. B., & Albertson, D. G. (2014). Changes in abundance of oral microbiota associated with oral

cancer. *PLoS ONE*, *9*(6), e98741. https://doi.org/10.1371/journal.pone.0098741

Semrau, F., Aidelsburger, P., & Israel, C. W. (2023). Common misunderstandings of evidence-based medicine. *Herzschrittmachertherapie & Elektrophysiologie*. https://doi.org/10.1007/s00399-023-00957-0

Sepich-Poore, G. D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J. A., & Knight, R. (2021). The microbiome and human cancer. *Science (New York, N.Y.)*, *371*(6536), eabc4552. https://doi.org/10.1126/science.abc4552

Serdar, C. C., Cihan, M., Yücel, D., & Serdar, M. A. (2021). Sample size, power and effect size revisited: Simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia Medica*, *31*(1), 010502. https://doi.org/10.11613/BM.2021.010502

Shen, X., Zhang, Y.-L., Zhu, J.-F., & Xu, B.-H. (2023). Oral dysbiosis in the onset and carcinogenesis of oral epithelial dysplasia: A systematic review. *Archives of Oral Biology*, *147*, 105630. https://doi.org/10.1016/j.archoralbio.2023.105630

Speight, P. M., & Morgan, P. R. (1993). The natural history and pathology of oral cancer and precancer. *Community Dental Health*, *10 Suppl 1*, 31–41.

Sperandio, M., Warnakulasuriya, S., Soares, A. B., Passador-Santos, F., Mariano, F. V., Lima, C. S. P., Scarini, J. F., Dominguete, M. H. L., De Camargo Moraes, P., Montalli, V. A. M., Hellmeister, L., &

De Araújo, V. C. (2023). Oral epithelial dysplasia grading: Comparing the binary system to the traditional 3-tier system, an actuarial study with malignant transformation as outcome. *Journal of Oral Pathology & Medicine*, *52*(5), 418–425. https://doi.org/10.1111/jop.13365

Su, S.-C., Chang, L.-C., Huang, H.-D., Peng, C.-Y., Chuang, C.-Y., Chen, Y.-T., Lu, M.-Y., Chiu, Y.-W., Chen, P.-Y., & Yang, S.-F. (2021). Oral microbial dysbiosis and its performance in predicting oral cancer. *Carcinogenesis*, *42*(1), 127–135. https://doi.org/10.1093/carcin/bgaa062

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Sutton, A., Clowes, M., Preston, L., & Booth, A. (2019). Meeting the review family: Exploring review types and associated information retrieval requirements. *Health Information & Libraries Journal*, *36*(3), 202–222. https://doi.org/10.1111/hir.12276

Takahashi, Y., Park, J., Hosomi, K., Yamada, T., Kobayashi, A., Yamaguchi, Y., Iketani, S., Kunisawa, J., Mizuguchi, K., Maeda, N., & Ohshima, T. (2019). Analysis of oral microbiota in Japanese oral cancer patients using 16S rRNA sequencing. *Journal of Oral Biosciences*, *61*(2), 120–128. https://doi.org/10.1016/j.job.2019.03.003

Tilakaratne, W. M., Jayasooriya, P. R., Jayasuriya, N. S., & De Silva, R. K. (2019). Oral epithelial dysplasia: Causes, quantification, prognosis, and management challenges. *Periodontology 2000*, *80*(1), 126–147. https://doi.org/10.1111/prd.12259

Toporcov, T. N., Znaor, A., Zhang, Z.-F., Yu, G.-P., Winn, D. M., Wei, Q., Vilensky, M., Vaughan, T., Thomson, P., Talamini, R., Szeszenia-Dabrowska, N., Sturgis, E. M., Smith, E., Shangina, O., Schwartz, S. M., Schantz, S., Rudnai, P., Richiardi, L., Ramroth, H., … Filho, V. W. (2015). Risk factors for head and neck cancer in young adults: A pooled analysis in the INHANCE consortium. *International Journal of Epidemiology*, *44*(1), 169–185. https://doi.org/10.1093/ije/dyu255

UCL. (2018, April 19). *Eastman Dental Hospital*. Eastman Dental Institute. https://www.ucl.ac.uk/eastman/about-us/eastman-dental-hospital

UCLH. (n.d.). *UCLH Head and Neck Centre*. University College London Hospitals NHS Foundation Trust. Retrieved August 21, 2023, from https://www.uclh.nhs.uk/our-services/find-service/head-and-neck-centre

Vogtmann, E., Chaturvedi, A. K., Blaser, M. J., Bokulich, N. A., Caporaso, J. G., Gillison, M. L., Hua, X., Hullings, A. G., Knight, R., Purandare, V., Shi, J., Wan, Y., Freedman, N. D., & Abnet, C. C. (2023). Representative oral microbiome data for the US population: The National Health and Nutrition Examination

Survey. *The Lancet. Microbe*, *4*(2), e60–e61. https://doi.org/10.1016/S2666-5247(22)00333-0

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., & STROBE Initiative. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Bulletin of the World Health Organization*, *85*(11), 867–872. https://doi.org/10.2471/blt.07.045120

Warnakulasuriya, S., Kujan, O., Aguirre-Urizar, J. M., Bagan, J. V., González-Moles, M. Á., Kerr, A. R., Lodi, G., Mello, F. W., Monteiro, L., Ogden, G. R., Sloan, P., & Johnson, N. W. (2021). Oral potentially malignant disorders: A consensus report from an international seminar on nomenclature and classification, convened by the WHO Collaborating Centre for Oral Cancer. *Oral Diseases*, *27*(8), 1862–1880. https://doi.org/10.1111/odi.13704

Wolf, A., Moissl-Eichinger, C., Perras, A., Koskinen, K., Tomazic, P. V., & Thurnher, D. (2017). The salivary microbiome as an indicator of carcinogenesis in patients with oropharyngeal squamous cell carcinoma: A pilot study. *Scientific Reports*, *7*, 5867. https://doi.org/10.1038/s41598-017-06361-2

Yang, C.-Y., Yeh, Y.-M., Yu, H.-Y., Chin, C.-Y., Hsu, C.-W., Liu, H., Huang, P.-J., Hu, S.-N., Liao, C.-T., Chang, K.-P., & Chang, Y.-L. (2018). Oral microbiota community dynamics associated with oral squamous cell carcinoma staging. *Frontiers in Microbiology*, *9*, 862. https://doi.org/10.3389/fmicb.2018.00862

Yang, J., Wang, Z.-Y., Huang, L., Yu, T.-L., Wan, S.-Q., Song, J., Zhang, B.-L., & Hu, M. (2021). Do betel quid and areca nut chewing deteriorate prognosis of oral cancer? A systematic review, meta-analysis, and research agenda. *Oral Diseases*, *27*(6), 1366–1375. https://doi.org/10.1111/odi.13456

Zaura, E., Pappalardo, V. Y., Buijs, M. J., Volgenant, C. M. C., & Brandt, B. W. (2021). Optimizing the quality of clinical studies on oral microbiome: A practical guide for planning, performing, and reporting. *Periodontology 2000*, *85*(1), 210–236. https://doi.org/10.1111/prd.12359

Zhang, L., Liu, Y., Zheng, H. J., & Zhang, C. P. (2020). The oral microbiota may have influence on oral cancer. *Frontiers in Cellular and Infection Microbiology*, *9*, 476. https://doi.org/10.3389/fcimb.2019.00476

Zhao, H., Chu, M., Huang, Z., Yang, X., Ran, S., Hu, B., Zhang, C., & Liang, J. (2017). Variations in oral microbiota associated with oral cancer. *Scientific Reports*, *7*(1), 11773. https://doi.org/10.1038/s41598-017-11779-9

Zheng, Y. (2018). Study design considerations for cancer biomarker discoveries. *The Journal of Applied Laboratory Medicine*, *3*(2), 282–289. https://doi.org/10.1373/jalm.2017.025809

Zhou, X., Hao, Y., Peng, X., Li, B., Han, Q., Ren, B., Li, M., Li, L., Li, Y., Cheng, G., Li, J., Ma, Y., Zhou, X., & Cheng, L. (2021). The clinical potential of oral microbiota as a screening tool for oral squamous cell carcinomas. *Frontiers in Cellular and Infection*

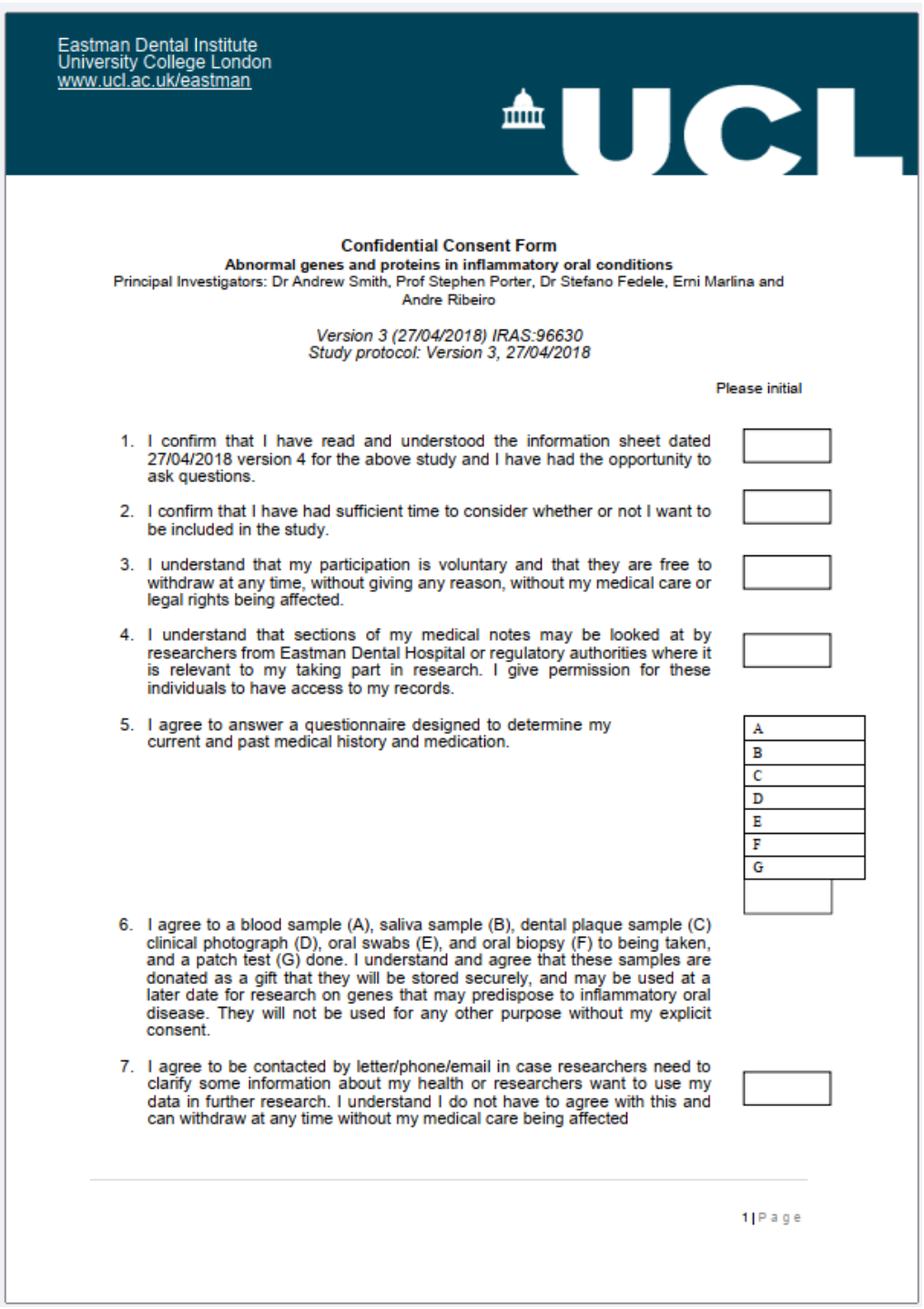
Microbiology, *11*, 728933. https://doi.org/10.3389/fcimb.2021.728933

Appendices

Appendix	Title
A	Consent form and participant information sheet
B	Blank questionnaire
C	DNeasy PowerSoil Kit modified instructions
D	NGS specific primers with overhang adaptors for metabarcoding
E	STROBE Checklist for Substudy 1
F	STROBE Checklist for Substudy 2
G	STROBE Checklist for Substudy 3

Appendix A: Consent Form and Participant Information Sheet

Consent Form



Eastman Dental Institute
University College London
www.ucl.ac.uk/eastman

Confidential Consent Form

Abnormal genes and proteins in inflammatory oral conditions

Principal Investigators: Dr Andrew Smith, Prof Stephen Porter, Dr Stefano Fedele, Erni Marlina and Andre Ribeiro

Version 3 (27/04/2018) IRAS:96630
Study protocol: Version 3, 27/04/2018

Please initial

1. I confirm that I have read and understood the information sheet dated 27/04/2018 version 4 for the above study and I have had the opportunity to ask questions. ☐
2. I confirm that I have had sufficient time to consider whether or not I want to be included in the study. ☐
3. I understand that my participation is voluntary and that they are free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. ☐
4. I understand that sections of my medical notes may be looked at by researchers from Eastman Dental Hospital or regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to my records. ☐
5. I agree to answer a questionnaire designed to determine my current and past medical history and medication. ☐
6. I agree to a blood sample (A), saliva sample (B), dental plaque sample (C) clinical photograph (D), oral swabs (E), and oral biopsy (F) to being taken, and a patch test (G) done. I understand and agree that these samples are donated as a gift that they will be stored securely, and may be used at a later date for research on genes that may predispose to inflammatory oral disease. They will not be used for any other purpose without my explicit consent.

   A ☐
   B ☐
   C ☐
   D ☐
   E ☐
   F ☐
   G ☐
7. I agree to be contacted by letter/phone/email in case researchers need to clarify some information about my health or researchers want to use my data in further research. I understand I do not have to agree with this and can withdraw at any time without my medical care being affected ☐

1 | Page

8. I agree to my GP being informed of my involvement in this study. ☐

9. I agree to take part in the above study. ☐

10. I would like to receive information relating to the results of the study (Please provide contact details) ☐

Name of patient	Date	Signature (optional)
Name of person taking consent	Date	Signature

Eastman Dental Institute
University College London
www.ucl.ac.uk/eastman

Confidential Information Sheet
Abnormal genes and proteins in inflammatory oral conditions
Principal Investigators: Dr Andrew Smith, Prof Stephen Porter, Dr Stefano Fedele, Erni Marlina and Andre Ribeiro

Version 4 (27/04/2018) IRAS: 96630
Study protocol: Version 3, 27/04/2018

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of this study?
We are investigating the causes of inflammatory oral conditions including orofacial granulomatosis, Lichen planus, Behçet's disease, vesiculobullous disorders, gum disease, precancerous/cancerous conditions, recurrent ulcers, foreign body reaction and Sjögrens disease. There is limited research into the causes of these conditions, but the immune system is thought to play a role. Other studies at the University College London have shown that there are differences in the way particular white blood cells (macrophages and neutrophils) work in patients with other inflammatory conditions such as Crohn's disease and ulcerative colitis. In patients with Crohn's disease the body has a weak response to bacteria, whereas patients with ulcerative colitis the body has a stronger response that lasts longer. As the tissue inside the mouth is similar to tissue of the gut, we hope to investigate to see whether these same differences in the way white blood cells work are true for oral inflammatory conditions.
This study will include analysis of genes, proteins, lipids and other molecules in white blood cells and various tissue samples. We will also investigate the microbes in the mouth and contrast this with the individual diseases.
If we can understand more about the cause of these conditions, we may be able to improve patient treatment. Results of this study may lead to further research studies.

This study received favorable opinion from the East of England - Cambridgeshire and Hertfordshire Research Ethics Committee (16/EE/0328) and Health Research Authority approval.

Why have I been chosen?
You have been asked to participate in this study because you have an oral inflammatory condition. We are investigating this in order to better understand the underlying cause.

What will I have to do to take part?
We are running several studies to investigate these molecules in patients with inflammatory oral conditions. We may ask you for one or more of the following:

(1) Blood samples (10-100ml. 10ml is less than 1 tablespoon. 100ml is less than 7 tablespoons). This will be used to analyse genes, proteins, lipids and other molecules in white blood cells and/or serum.

(2) Saliva samples (2-5ml; ~1 teaspoon). This will be used to analyse genes from the cells in the saliva as well as microflora.

(3) **Dental plaque samples.** The dental plaque will be wiped off the surfaces of the teeth using a small dental instrument from just below the gum line. This will be used to screen what bacteria are currently in your mouth.

(4) **Oral swab.** The oral swab will be taken from the inside of the cheek. This will be used to analyse the microflora.

(5) **Clinical photograph.** A clinical photograph of the oral cavity and lips might be taken at the time of sample collection. It will not be possible to identify the face from the photograph. This will be used to document the oral health and identify the sites where the samples were taken.

(6) **Oral tissue sample.** We may ask whether we can have a small portion of tissue from inside your mouth. This will only be requested in cases where you need a biopsy for diagnostic/medical reasons. In such circumstances we may ask your permission to cut off a small piece (approximately 2x2mm) from the sample your doctor has taken so we can analyse different proteins, lips and genes within the tissue.

(7) **Patch test.** This will only be performed on patients who have been diagnosed with orofacial granulomatosis and will be done to test for hypersensitivity to known allergens associated with oral inflammation. The test will be arranged separately and will be conducted at the Eastman Dental Institute and involve 3 clinical visits over 5 day period. Initial appointment will involve application of up to 20 patches with different common food and dental allergens. These would be removed 48 hours later at the second appointment. The final reading will be done at the third appointment - 96 hours after the initial appointment. We will determine the strength of the reactions through three different ways (visual score, photographs and using a Laser Doppler Imager, which records changes in blood flow. There is no ionising radiation involved). With your consent the results from this test will be added to your clinical notes and sent to your GP.

(8) **Questionnaires.** We will use two forms to collect details needed for this study. The first form will comprise of the "Records Sheet" which will be completed using your medical records and contain disease type, current and past medication and GP's details. The second form "Confidential clinical data collection record", which is a questionnaire, will be collected on the day of the appointment by the clinical investigator. The questions will concentrate on your medical/oral history and determine any family history of related diseases. The questionnaire should take no longer than 10 mins to complete.

We will not ask you to undergo any medical procedures beyond those required for clinical reasons. We will always ask you again for consent before using any additional samples.

Do I have to take part?
No, taking part is voluntary. If you do not want to take part, you do not have to tell us why. You can change your mind and withdraw from the study at any stage. Agreeing to contribute to the study will not affect the care or treatment you receive. If you do not wish to take part or

decide to withdraw, your treatment and relationship with your doctor will not be affected in any way.

What patient information will be collected?
We shall record your date of birth, gender and clinical data as recorded in your hospital notes. These will be used to run our sample analysis on oral inflammation. These data will be anonymized using study IDs.

Confidentiality – who will have access to the samples and findings?
The samples will be given a code and stored securely within the Eastman Dental Institute, University College London. Participant information will be stored securely on a password-protected computer within the Department accessible only by researchers of this study. All patient information will be treated in the strictest confidence, in accordance with the Data Protection Act 1998. The samples and participant information may be used for future ethically approved studies, but only to investigate genes that may cause oral inflammatory conditions.

What are the possible risks of taking part?
There are no risks or side effects. For example, 100ml of blood is about 20% of that taken at the time of blood donation or 7 tablespoons. We will not perform any medical procedures that you do not require for clinical reasons.

What are the implications of participating in a genetic study?
We shall only investigate genes that are believed to cause inflammation. We anticipate no risks from investigating your genes. Due to the nature of the tests we will use in this study, there is a chance that we may identify clinically significant genes, for example BRCA1 and BRAC2 which have been linked to causing breast and ovarian cancer. Our study is not designed to be diagnostic and we are unable to test for clinically significant genes. If we inadvertently discover that you may be a carrier of a clinically significant gene then we would write a letter to your GP in order to provide you with the relevant NHS guidelines on the best course of action. Some of the clinically relevant genes are hereditary and direct family members can also be affected. Your GP may advise testing your relatives if the procedure is not available through the NHS we will endeavor to perform the relevant test and pass the results onto the GP. If you are not happy with this course of action please let us know.

What if new information becomes available?
Sometimes during the course of a research project, new information becomes available. If this happens, we shall tell you about it and discuss whether you want to continue in the study. If you decide to continue you will be asked to sign an updated consent form.

Will my GP be informed?
With your consent we would like to inform your GP of your participation in this study by sending a letter.

What happens if I decide to withdraw from the study?
This will not affect your medical care in any way. If you decide to withdraw, all patient information and specimens will be destroyed.

What will happen to the samples I give?
All samples will be stored in a pseudoanonymous form and in accordance with the Human Tissue Act. Initially we will use the samples to study the genes, proteins, lipids and other molecules in white blood cells and various tissues. During the study period all samples will be stored in a secure location within UCL. Once the study is completed any remaining tissue will be transferred to the Eastman Dental Institutes tissue bank. Remaining DNA, RNA and proteins will be stored by the research team and unwanted material disposed of in

accordance with the Human Tissue Act. Long-term storage of the samples will allow us to conduct future studies if new relevant information becomes available.

What will happen to the study results?
The samples you kindly provide will be used to perform medical research for publication in medical/scientific journals. We hope that this will help in the treatment of oral inflammatory disorders. No details that specifically identify you will be included. We can provide you with details of any publication, at your request. If you are interested in obtaining the results from our study then we will be able to send them to you either via your GP or directly. In order for us to communicate the results to you we will need to have either a postal address or e-mail address.
Data generated by the study will be stored for up to 20 years on university and NHS secure computers and used for presentations in scientific and clinical conferences and meetings, publications in scientific and medical literature, patient information leaflets if appropriate. Only completely anonymised data will be used for these purposes.
Patients' health records (including a copy of consent form for research study) will be stored by the university and hospital for at least 10 years. All activities are conducted in accordance with the Data Protection Act and UCL Data Protection Policy.

Will this study benefit me?
No, this study will not directly benefit you at the moment, but we hope that it will benefit patients with oral disease in the future. However, in case you participate in patch testing and are found to be allergic to something, you will know which allergens to avoid.

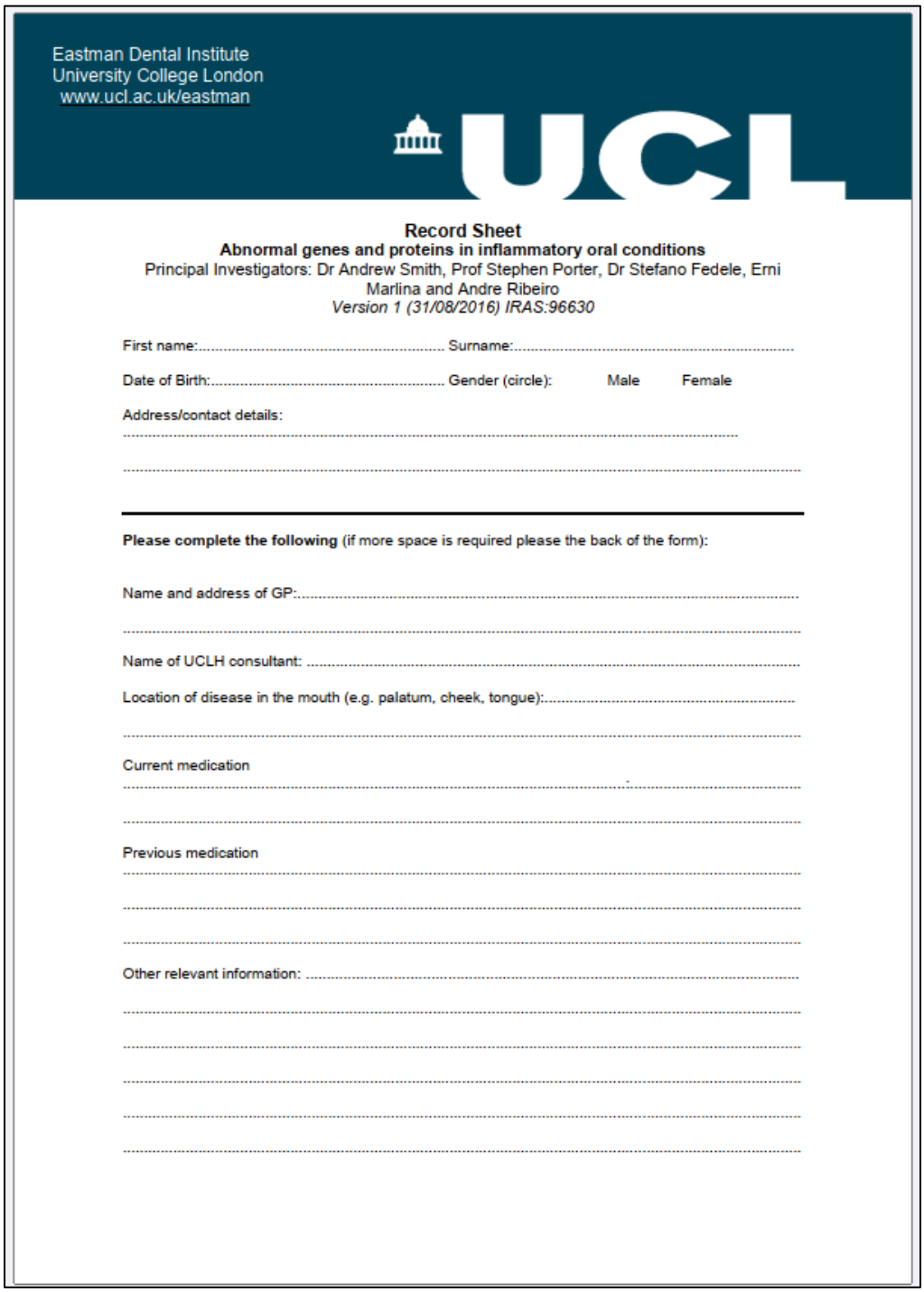
What do I do now?
If you are happy to take part in this study, you will be asked to sign a consent form to show that you have understood what the study entails.

What if something goes wrong?
If you are concerned about any aspect of this study, please speak to the researchers who will do their best to answer your questions. Please contact Dr Andrew Smith (andrew.m.smith@ucl.ac.uk). If you remain unhappy, you can make a formal complaint through the National Health Service (NHS) complaints procedure. Details can be obtained through the University College London Hospitals (UCLH) Patient Advice and Liaison Service (PALS) on 0207 3447 3041, email: PALS@uclh.nhs.uk, address: PALS, Ground Floor Atrium, University College Hospital, 235 Euston Road, London, NW1 2BU.

University College London (UCL) holds insurance against claims from participants for harm caused by their participation in this clinical study. Participants may be able to claim compensation if they can prove that UCL has been negligent. However, if this clinical study is being carried out in a hospital, the hospital continues to have a duty of care to the participant of the clinical study. University College London does not accept liability for any breach in the hospital's duty of care, or any negligence on the part of hospital employees. This applies whether the hospital is an NHS Trust or otherwise.

For further information, please contact Dr Andrew Smith, email address:
andrew.m.smith@ucl.ac.uk.

Appendix B: Blank Questionnaire



Eastman Dental Institute
University College London
www.ucl.ac.uk/eastman

Record Sheet
Abnormal genes and proteins in inflammatory oral conditions
Principal Investigators: Dr Andrew Smith, Prof Stephen Porter, Dr Stefano Fedele, Erni Marlina and Andre Ribeiro
Version 1 (31/08/2016) IRAS:96630

First name:..... Surname:.....

Date of Birth:..... Gender (circle): Male Female

Address/contact details:
.....

.....

Please complete the following (if more space is required please the back of the form):

Name and address of GP:.....

.....

Name of UCLH consultant: .....

Location of disease in the mouth (e.g. palatum, cheek, tongue):.....

.....

Current medication
.....

.....

Previous medication
.....

.....

.....

Other relevant information: .....

.....

.....

.....

.....

.....

ORION samples laboratory processing 1 - Extraction of Metagenomics DNA using DNeasy PowerSoil Kit (QIAGEN):

Materials & equipment:

1. Gloves.
2. Mask.
3. DNeasy PowerSoil Kit (QIAGEN):
    - Small (2 ml) screw-top (external thread) MoBio collection tubes containing 750 µl of specimen collection fluid (MoBio buffer) (i.e, PowerBead tubes).
    - PowerSoil® 2 ml Collection Tubes
    - PowerSoil® Spin Filters (units in 2 ml tubes)
    - PowerSoil® Solution C1
    - PowerSoil® Solution C2
    - PowerSoil® Solution C3
    - PowerSoil® Solution C4
    - PowerSoil® Solution C5
    - PowerSoil® Solution C6
4. Vortexer.
5. MO BIO Vortex Adapter tube.
6. Microcentrifuge.
7. 4°C fridge.
8. -20°C freezer.
9. 5 ml sterile tubes.
10. Filter tips (50 ul – 1000 ul).
11. Pipettes.

DNA isolation:

Wear gloves at all times.

- Solutions preparations:
  1. Check solution C1. If solution C1 is precipitated, heat solution to 60°C until dissolved before use.
     - Prepare one tube of C1 for each participant to accommodate 60 µl for each sample.
  2. Solutions C2, C3, C4 (shake to mix C4 before use) and C5
     For each sample prepare 1 tube of solutions C2, C3, C4 and C5 from the kit.
     - Tube C2= 250 µl in 2 ml collection tube
     - Tube C3= 200 µl in 2 ml collection tube
     - Tube C4= 1200 µl in 5 ml sterile collection tube
     - Tube C5= 500 µl in 2 ml collection tube
  3. Solution C6

- Prepare one tube of C6 for each participant to accommodate 100 µl for each sample.

- Procedure:
  1. Gently vortex the PowerBead tubes to mix.
  2. Add 60 µl of solution C1 and vortex briefly (10 seconds).
  3. Secure PowerBead Tubes horizontally using the MO BIO Vortex adapter tube holder for the vortex. Vortex at maximum speed for 20 min.
  4. Make sure the PowerBead Tubes rotate freely in your centrifuge without rubbing, Centrifuge tubes at 10,000 x g for 30 sec at room temperature.

     CAUTION: be careful not to exceed 10,000 x g or tubes may break.

  5. Transfer the supernatant (expect 400-500 µl don't take any more than this) to tube C2 and vortex for 5 sec.
  6. Incubate at 4°C for 5 min (supernatant may still contain some sample particles).
  7. Centrifuge tubes at 10,000 x g for 1 min at room temperature.
  8. Avoiding the pellet, transfer up to, but no more than, 600 µl of supernatant to tube C3 and vortex briefly.
  9. Incubate at 4°C for 5 min.
  10. Centrifuge tubes at 10,000 x g for 2 min at room temperature.
  11. Avoiding the pellet, transfer up to, but no more than, 750 µl of supernatant to tube C4 and vortex for 5 sec.

12. Load approximately 675 µl onto Spin filter column and centrifuge at 10,000 x g for 1 minute at room temperature.
13. Discard the flow-through and add an additional 675 µl of supernatant to the Spin Filter and centrifuge at 10,000 x g for 1 minute at room temperature.
14. Discard the flow-through and add the remaining 675 µl of supernatant to the Spin Filter and centrifuge at 10,000 x g for 1 minute at room temperature.

    NOTE: a total of three loads for each sample processed are required.

15. Add 500 µl of Solution C5 and centrifuge at room temperature for 30 sec at 10,000 x g.
16. Discard the flow-through.
17. Centrifuge again at room temperature for 1 min at 10,000 x g.
18. Carefully place Spin filter column into a clean 2 ml Collection Tube. Avoid splashing any solution C5 onto the column.
19. Add 70 µl of 60C pre-heated molecular ddH2O (sterile DNA-free PCR-grade water (cat. no. 17000-10) to the centre of the white filter membrane.
20. Centrifuge at room temperature for 30 sec at 10,000 x g.
21. Discard the spin column. The DNA in the tube should be stored at -20 °C or lower. If these DNA samples will bel labelled according to the study ID ex. If the subject ID was ORION EDH-0001 for a normal swab, then the DNA will bel labelled ORION EDH-0001 DNA normal swab.

Appendix D: NGS specific primers with overhang adaptors for metabarcoding

NGS (MiSeq) primers for V3-V4 regions

Foward Primers (V1)

PrimerID	P5(29bp)	Index(8bp)	Tm Padding (11bp)	Forward Template - 341F (17bp)	Complete Sequence	Length (bp)
P5_V3_FWD01	AATGATACGGCGACCACCGAGATCTACAC	TAGATCGC	AGTACGTACGT	CCTACGGGNGGCWGCAG	AATGATACGGCGACCACCGAGATCTACACTAGATCGCAGTACGTACGTCCTACGGGNGGCWGCAG	65
P5_V3_FWD02	AATGATACGGCGACCACCGAGATCTACAC	CTCTCTAT	AGTACGTACGT	CCTACGGGNGGCWGCAG	AATGATACGGCGACCACCGAGATCTACACCTCTCTATAGTACGTACGTCCTACGGGNGGCWGCAG	65
P5_V3_FWD03	AATGATACGGCGACCACCGAGATCTACAC	TATCCTCT	AGTACGTACGT	CCTACGGGNGGCWGCAG	AATGATACGGCGACCACCGAGATCTACACTATCCTCTAGTACGTACGTCCTACGGGNGGCWGCAG	65
P5_V3_FWD04	AATGATACGGCGACCACCGAGATCTACAC	AGAGTAGA	AGTACGTACGT	CCTACGGGNGGCWGCAG	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAAGTACGTACGTCCTACGGGNGGCWGCAG	65
P5_V3_FWD05	AATGATACGGCGACCACCGAGATCTACAC	GTAAGGAG	AGTACGTACGT	CCTACGGGNGGCWGCAG	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGAGTACGTACGTCCTACGGGNGGCWGCAG	65
P5_V3_FWD06	AATGATACGGCGACCACCGAGATCTACAC	ACTGCATA	AGTACGTACGT	CCTACGGGNGGCWGCAG	AATGATACGGCGACCACCGAGATCTACACACTGCATAAGTACGTACGTCCTACGGGNGGCWGCAG	65
P5_V3_FWD07	AATGATACGGCGACCACCGAGATCTACAC	AAGGAGTA	AGTACGTACGT	CCTACGGGNGGCWGCAG	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAAGTACGTACGTCCTACGGGNGGCWGCAG	65
P5_V3_FWD08	AATGATACGGCGACCACCGAGATCTACAC	CTAAGCCT	AGTACGTACGT	CCTACGGGNGGCWGCAG	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTAGTACGTACGTCCTACGGGNGGCWGCAG	65

Reverse Primers (V3)

PrimerID	P7(24bp)	Index(8bp)	Tm Padding(15bp)	Reverse Template - 805R (21bp)	Complete Sequence	Length (bp)
P7_V4_REV01	CAAGCAGAAGACGGCATACGAGAT	TCGCCTTA	GCGAGTCAGTCAGCG	GACTACHVGGGTATCTAATCC	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	68
P7_V4_REV02	CAAGCAGAAGACGGCATACGAGAT	CTAGTACG	GCGAGTCAGTCAGCG	GACTACHVGGGTATCTAATCC	CAAGCAGAAGACGGCATACGAGATCTAGTACGGCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	68
P7_V4_REV03	CAAGCAGAAGACGGCATACGAGAT	TTCTGCCT	GCGAGTCAGTCAGCG	GACTACHVGGGTATCTAATCC	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	68
P7_V4_REV04	CAAGCAGAAGACGGCATACGAGAT	GCTCAGGA	GCGAGTCAGTCAGCG	GACTACHVGGGTATCTAATCC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	68
P7_V4_REV05	CAAGCAGAAGACGGCATACGAGAT	AGGAGTCC	GCGAGTCAGTCAGCG	GACTACHVGGGTATCTAATCC	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	68
P7_V4_REV06	CAAGCAGAAGACGGCATACGAGAT	CATGCCTA	GCGAGTCAGTCAGCG	GACTACHVGGGTATCTAATCC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	68
P7_V4_REV07	CAAGCAGAAGACGGCATACGAGAT	GTAGAGAG	GCGAGTCAGTCAGCG	GACTACHVGGGTATCTAATCC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	68
P7_V4_REV08	CAAGCAGAAGACGGCATACGAGAT	CCTCTCTG	GCGAGTCAGTCAGCG	GACTACHVGGGTATCTAATCC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	68
P7_V4_REV09	CAAGCAGAAGACGGCATACGAGAT	AGCGTAGC	GCGAGTCAGTCAGCG	GACTACHVGGGTATCTAATCC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	68
P7_V4_REV10	CAAGCAGAAGACGGCATACGAGAT	CAGCCTCG	GCGAGTCAGTCAGCG	GACTACHVGGGTATCTAATCC	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	68
P7_V4_REV11	CAAGCAGAAGACGGCATACGAGAT	TGCCTCTT	GCGAGTCAGTCAGCG	GACTACHVGGGTATCTAATCC	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	68
P7_V4_REV12	CAAGCAGAAGACGGCATACGAGAT	TCCTCTAC	GCGAGTCAGTCAGCG	GACTACHVGGGTATCTAATCC	CAAGCAGAAGACGGCATACGAGATTCCTCTACGCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	68

Read Primers (V3-V5)

PrimerID	Complete Sequence	Tm	Length (bp)
Read 1 Primer_V3	AGTACGTACGTCCTACGGGNGGCWGCAG	66.9	28
Read 2 Primer_V4	GCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC	66.2	36
Index read Primer_V4	GGATTAGATACCCBDGTAGTCCGCTGACTGACTCGC	66.2	36
	(Reverse complement of Read 2 Primer_V4)		

Appendix E: STROBE Checklist for Substudy 1

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	*(a)* Indicate the study's design with a commonly used term in the title or the abstract	3
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	131-140
Objectives	3	State specific objectives, including any prespecified hypotheses	141
Methods			
Study design	4	Present key elements of study design early in the paper	141
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	141-143
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	142, 143
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	142
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	142, 143
Bias	9	Describe any efforts to address potential sources of bias	142, 143
Study size	10	Explain how the study size was arrived at	142
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	144-148
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	148, 149
		(*b*) Describe any methods used to examine subgroups and interactions	N/A
		(*c*) Explain how missing data were addressed	N/A
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	144
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	150-152
		(b) Indicate number of participants with missing data for each variable of interest	N/A
Outcome data	15*	Report numbers of outcome events or summary measures	152-155

1

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	N/A
		(*b*) Report category boundaries when continuous variables were categorized	N/A
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
Discussion			
Key results	18	Summarise key results with reference to study objectives	155-156
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	157, 158
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	155-157
Generalisability	21	Discuss the generalisability (external validity) of the study results	155-158
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	N/A

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Appendix F: STROBE Checklist for Substudy 2

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	3
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	160-165
Objectives	3	State specific objectives, including any prespecified hypotheses	165
Methods			
Study design	4	Present key elements of study design early in the paper	164, 165
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	166
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	166
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	166
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	166
Bias	9	Describe any efforts to address potential sources of bias	166
Study size	10	Explain how the study size was arrived at	166
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	166
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	166, 167
		(*b*) Describe any methods used to examine subgroups and interactions	N/A
		(*c*) Explain how missing data were addressed	N/A
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	167
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	167
		(b) Indicate number of participants with missing data for each variable of interest	167
Outcome data	15*	Report numbers of outcome events or summary measures	167-172

1

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	N/A
		(*b*) Report category boundaries when continuous variables were categorized	N/A
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
Discussion			
Key results	18	Summarise key results with reference to study objectives	172
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	175, 176
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	172-175
Generalisability	21	Discuss the generalisability (external validity) of the study results	176
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	N/A

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	*(a)* Indicate the study's design with a commonly used term in the title or the abstract	3
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	177-179
Objectives	3	State specific objectives, including any prespecified hypotheses	179
Methods			
Study design	4	Present key elements of study design early in the paper	178-179
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	178-179
Participants	6	*(a)* Give the eligibility criteria, and the sources and methods of selection of participants	178-179
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	178-179
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	178
Bias	9	Describe any efforts to address potential sources of bias	N/A
Study size	10	Explain how the study size was arrived at	181
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	179-181
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	181-182
		(*b*) Describe any methods used to examine subgroups and interactions	N/A
		(*c*) Explain how missing data were addressed	N/A
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	178
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	182, 179
		(b) Indicate number of participants with missing data for each variable of interest	N/A
Outcome data	15*	Report numbers of outcome events or summary measures	183-188

1

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	N/A
		(*b*) Report category boundaries when continuous variables were categorized	N/A
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
Discussion			
Key results	18	Summarise key results with reference to study objectives	187, 188
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	194, 195
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	187-195
Generalisability	21	Discuss the generalisability (external validity) of the study results	194-196
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	N/A

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.