

From Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

AN EPIDEMIOLOGICAL STUDY ON NASOPHARYNGEAL CARCINOMA

Yufeng Chen

陈裕锋



**Karolinska
Institutet**

Stockholm 2021

All previously published papers were reproduced with permission from the publishers.

Published by Karolinska Institutet.

Printed by Universitetsservice US-AB, 2021

© Yufeng Chen, 2021

ISBN 978-91-8016-374-3

An Epidemiological Study on Nasopharyngeal Carcinoma

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Yufeng Chen

Time: Tuesday 16 November 2021, 13:00 pm.

Location: Atrium lecture hall, Nobels väg 12B, Karolinska Institutet, Solna, Stockholm

Principal Supervisor:

Professor Weimin Ye
Karolinska Institutet
Department of Medical Epidemiology and
Biostatistics

Co-supervisor(s):

Professor Rob Knight
University of California San Diego
Department of Pediatrics

Dr. Zhiwei Liu
National Cancer Institute (USA)
Division of Cancer Epidemiology and Genetics

Opponent:

Dr. Gary Clifford
International Agency for Research on Cancer
Infections and Cancer Epidemiology (ICE) group

Examination Board:

Associate Professor Linda Lindström
Karolinska Institutet
Department of Oncology and Pathology

Professor Keith Humphreys
Karolinska Institutet
Department of Medical Epidemiology and
Biostatistics

Associate Professor Johan Dicksved
Swedish University of Agricultural Sciences
Department of Animal Nutrition and Management

To my beloved family

至我亲爱的家人

ABSTRACT

Nasopharyngeal carcinoma (NPC) is a malignancy known for its unique geographic and ethnic distribution characterized by particularly high incidence in southern China, Southeast Asia, and North and East Africa, but being rare in the rest of the world. Its etiology has remained enigmatic and the accumulated evidence suggests that genetic predisposition, environmental factors, and infection with Epstein-Barr virus (EBV) are involved. Nevertheless, the association between environmental exposures and NPC carcinogenesis remains largely elusive to date, and how environmental risk factors interact with EBV in the development of NPC has been rarely studied. In addition, oral microbiome is emerging as a vital factor contributing to carcinogenic processes, but its role in NPC is largely unknown.

The overall aim of this thesis is to provide more solid evidence and precise insights into the etiology of NPC, with focus on the associations of NPC with environmental, viral, and other microbial factors. To address these knowledge gaps, we carried out a large-scale population-based case-control study entitled NPC Genes, Environment, and EBV (NPCGEE) in southern China from 2010 to 2014. In short, 2,554 histopathologically confirmed, incident NPC cases, and 2,648 controls frequency matched to cases on age, sex and geographic area from general population were recruited.

In Study I, we investigated the magnitude and pattern of associations between NPC and various residential exposures. Based on the NPCGEE study questionnaire data of a lifelong residential history, we found poor residential conditions including living in inferior housing types, using less clean fuels for cooking, house with poor ventilation, using untreated water sources, exposure to smoke when cooking, burning incense, and residential proximity to a factory area, were associated with a higher risk of NPC, and the associations were notably stronger for exposures at young ages.

In Study II, we examined the relationship of NPC risk with occupational exposures based also on the NPCGEE study. With the analysis of the complete occupational history data, we found significantly elevated risk of NPC associated with exposures to broad categories of occupational pollutants, including dusts, chemical vapors, exhausts/smokes, and acids/alkalis. These associations were primarily explained by 14 subtypes of occupational agents within the above-mentioned broad categories. Moreover, the strengths of the associations were generally stronger with increasing duration of exposure.

In Study III, we assessed whether there is an association between environmental factors and EBV reactivation, a critical step in the NPC carcinogenesis, in the healthy population controls of the NPCGEE study. Overall, we found no associations between EBV reactivation and extensive environmental factors, including alcohol or tea drinking, a history of chronic ear/nose/throat diseases, use of medications or herbs, consumption of salted fish or other

preserved foods, oral hygiene, sibship structure, and various residential and occupational exposures. Cigarette smoking is the only factor associated with EBV reactivation.

In Study IV, we investigated the relationship between oral fungal microbiome and NPC status using Internal Transcribed Spacer (ITS)-2 sequencing in a subset of NPCGEE study samples (538 NPC patients, 537 controls). We found a significantly reduced fungal community richness and diversity in NPC patients compared with those in controls, and the global fungal community compositions significantly differed between cases and controls. Furthermore, a number of differentially abundant oral fungal organisms in NPC cases and healthy controls were identified.

In conclusion, findings from this thesis work support that individuals exposed to both residential and occupational risk factors are at increased risk of NPC. Most environmental factors, except for cigarette smoking, are not likely to induce EBV reactivation; other mechanisms such as host genetic and viral variations in the EBV reactivation deserve to be further studied. Dysbiotic oral mycobiome characterized by reduced community richness and diversity, as well as an increased abundance in pathogenic fungi and a decrease in commensal fungi may contribute to the development of NPC.

LIST OF SCIENTIFIC PAPERS

This thesis is based on the following papers, which will be referred to in the content by their Roman numerals (I-IV). * Contribute equally.

- I. Chen Y, Chang ET, Liu Z, Liu Q, Cai Y, Zhang Z, Chen G, Huang QH, Xie SH, Cao SM, Jia WH, Zheng Y, Li Y, Lin L, Ernberg I, Zhao H, Feng R, Huang G, Zeng Y, Zeng YX, Adami HO, Ye W. Residence characteristics and risk of nasopharyngeal carcinoma in southern China: a population-based case-control study. *Environment International*, 2021, 151:106455.
- II. Chen Y, Chang ET, Liu Q, Cai Y, Zhang Z, Chen G, Huang QH, Xie SH, Cao SM, Jia WH, Zheng Y, Li Y, Lin L, Ernberg I, Wang D, Chen W, Feng R, Huang G, Zeng YX, Adami HO, Ye W. Occupational exposures and risk of nasopharyngeal carcinoma in a high-risk area: a population-based case-control study. *Cancer*, 2021, 127(15):2724-2735.
- III. Chen Y, Chang ET, Liu Q, Cai Y, Zhang Z, Chen G, Huang QH, Xie SH, Cao SM, Jia WH, Zheng Y, Li Y, Lin L, Ernberg I, Huang G, Zeng YX, Adami HO, Ye W. Environmental risk factors for Epstein-Barr virus reactivation in a high-risk area for nasopharyngeal carcinoma: a population-based study. *Manuscript submitted*.
- IV. Chen Y *, Li W *, Chang ET, Debelius JW, Manoharan L, Zhang Z, Zheng Y, Li Y, Huang G, Adami HO, Knight R, Cai Y, Ye W. Oral fungal profiling and risk of nasopharyngeal carcinoma: a population-based case-control study. *Manuscript*.

CONTENTS

1	Introduction	1
2	Background.....	3
2.1	Descriptive epidemiology	3
2.2	Residential exposures and NPC	6
2.3	Occupational exposures and NPC.....	8
2.4	Environmental factors and EBV reactivation	10
2.5	Oral fungal communities.....	11
3	Research aims	15
4	Materials and methods	17
4.1	Data sources and study design	17
4.1.1	Study base.....	17
4.1.2	NPC case recruitment.....	18
4.1.3	Population control recruitment	18
4.2	Data collection and sample collection	18
4.3	Laboratory methods.....	19
4.3.1	EBV serological tests	19
4.3.2	Oral fungal microbiome profiling.....	20
4.4	Statistical analyses.....	20
4.4.1	Residential exposures and NPC.....	20
4.4.2	Occupational exposures and NPC	21
4.4.3	Environmental factors and EBV reactivation.....	21
4.4.4	Oral fungal microbiome and NPC	22
4.5	Ethical considerations.....	22
5	Results.....	25
5.1	Residential exposures and risk of NPC.....	25
5.2	Occupational exposures and risk of NPC	29
5.3	Environmental factors and EBV reactivation	32
5.4	Oral fungal profiling and risk of NPC	35
6	Discussion.....	38
6.1	Findings, interpretations and implications.....	38
6.1.1	Residential exposures and risk of NPC	38
6.1.2	Occupational exposures and risk of NPC	39
6.1.3	Environmental factors and EBV reactivation.....	40
6.1.4	Oral fungal profiling and risk of NPC	40
6.2	Methodological considerations	41
6.2.1	Study design	41
6.2.2	Sample size	42
6.2.3	Selection bias.....	43
6.2.4	Information bias (misclassification)	43
6.2.5	Confounding	44
6.2.6	Generalizability	44

7	Conclusions	46
8	Funding.....	47
9	Acknowledgements	48
10	References	51

LIST OF ABBREVIATIONS

ASV	Amplicon sequence variant
BMI	Body mass index
CI	Confidence interval
DNA	Deoxyribonucleic acid
EA	Early antigen
EBV	Epstein-Barr virus
EBERs	EBV-encoded small RNAs
EBNA1	EBV nuclear antigen 1
ELISA	Enzyme-linked immunosorbent assay
ENT	Ear, nose, and throat
FEP	Fibro-epithelial polyp
FDR	False discovery rate
GWAS	Genome-wide association study
HAP	Household air pollution
HLA	Human leukocyte antigen
HNSCC	Head and neck squamous cell carcinoma
IARC	International Agency for Research on Cancer
ITS	Internal transcribed spacer
LDA	Linear discriminant analysis
LEfSe	Linear discriminant analysis effect size
LMP	Latent membrane protein
NCI	National Cancer Institute
NGS	Next-generation sequencing
NPC	Nasopharyngeal carcinoma
NPCGEE	NPC Genes, Environment, and EBV
OLP	Oral lichen planus
OR	Odds ratio
OSCC	Oral squamous cell carcinoma
PCoA	Principal coordinates analysis
PCR	Polymerase chain reaction

PERMANOVA	Permutational multivariate analysis of variance
RERI	Relative excess risk due to interaction
rRNA	Ribosomal ribonucleic acid
SES	Socioeconomic status
SMR	Standardized mortality ratio
VCA	Viral capsid antigen
WGS	Whole-genome sequencing

1 INTRODUCTION

Nasopharyngeal carcinoma (NPC) is a malignant cancer arising from epithelial lining of the nasopharynx and known for its unique geographic and racial distribution. It is a rare disease in most parts of the world, but it poses heavy disease burden in southern China, Southeast Asia, and North and East Africa. Infection with Epstein-Barr virus (EBV) has been recognized as a necessary cause for NPC development. However, the ubiquitous infection over 95% of the world population suggests EBV itself is not sufficient to cause this malignancy. Further, the unique geographic and ethnic distribution of NPC worldwide implies that other factors including environmental factors and genetic traits also contribute to its development.

Although the role of environmental factors in NPC has been widely studied over decades, the link between environmental factors and the risk of NPC, however, remains largely inconclusive to date. The inconsistent results in previous studies may be due in part to various shortcomings in methodology, including without a strict population-based design, insufficient statistical power, and inadequate control of confounding.

EBV infection is strongly associated with the carcinogenesis of undifferentiated NPC, the predominant histopathological type in endemic regions. However, the exact biological mechanism of how EBV contributes to the carcinogenesis is unclear. Evidence from EBV serologic studies and increasing output from molecular research suggest that EBV reactivation from latent to lytic infection plays a pivotal role in NPC carcinogenesis. Whether environmental risk factors can act as inducers for EBV reactivation and how environmental risk factors interact with EBV in the development of NPC have been rarely investigated.

With the advent of next-generation sequencing technologies, the role of microbiome in human health and disease is being uncovered. The oral microbiome, a vital member of human microbiome, is emerging as an important factor contributable to carcinogenic processes. In some most recent findings, dysbiosis of oral bacterial microbiome has been linked to NPC risk, but the role of oral mycobiome in NPC development is still unknown.

2 BACKGROUND

2.1 DESCRIPTIVE EPIDEMIOLOGY

Nasopharyngeal carcinoma (NPC) is a head-neck malignancy arising from the epithelial lining of the nasopharynx, with over 133,000 new cases and 80,000 deaths worldwide in 2020 [1]. It is a rare disease with less than 1 per 100,000 person-years throughout most parts of the world, whereas, it is particularly prevalent in some regions including southern China, south-eastern Asia, northern and eastern Africa (**Figure 1, Figure 2**). The highest incidence rate of NPC in the world was observed among Cantonese-speaking population in southern China, which can reach as high as 30 per 100,000 person-years [2].

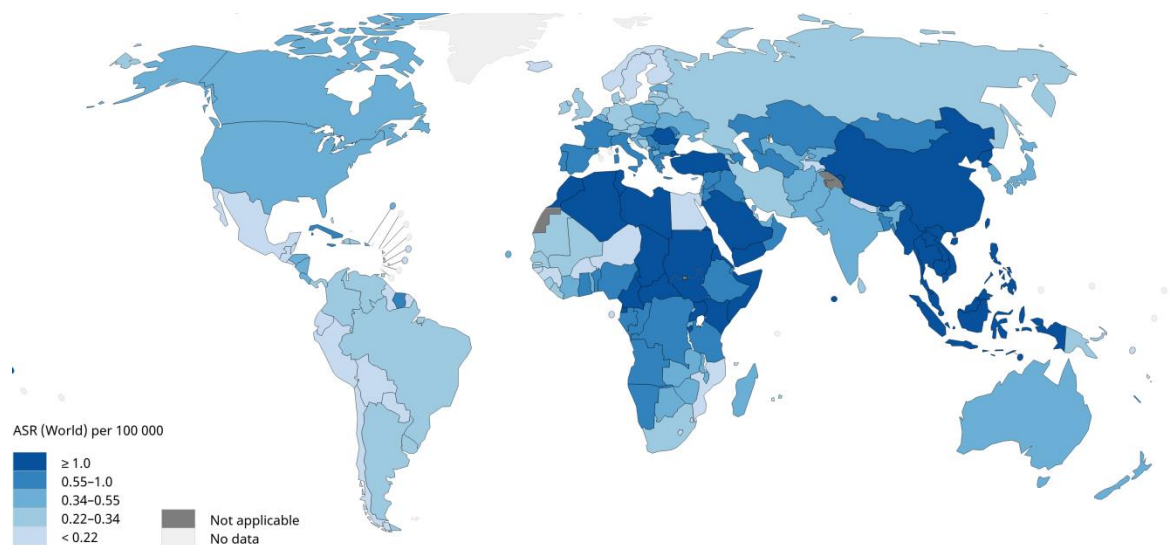


Figure 1. Age-standardized (world) incidence rates of nasopharyngeal carcinoma worldwide, both sexes, ages 0-74. (Source: <https://gco.iarc.fr/today>)

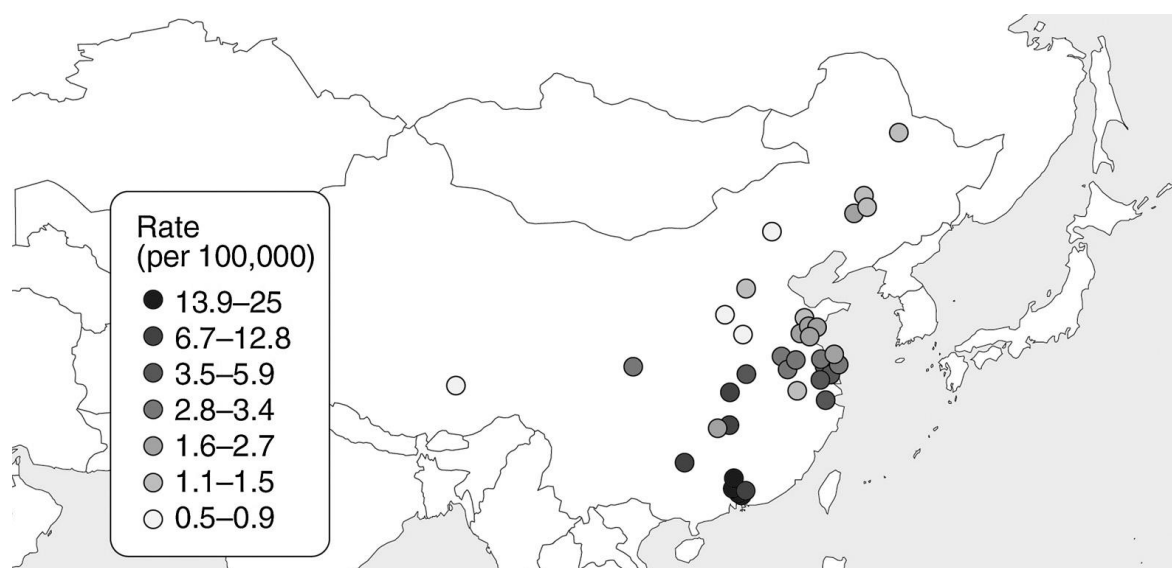


Figure 2. Age-standardized (world) incidence rates of nasopharyngeal carcinoma among men by city or county, China, 2008–2012. (Source: Ref. [3])

Aside from geographical difference, the distribution of NPC varies greatly across ethnic groups even within the same country. Almost half of NPC cases in the world occur in China [2,4], however, even within China, the incidence of this cancer differs significantly, with the incidence in southern China (10-30/100,000 person-years) much higher than that in northern China ($< 2/100,000$ person-years) [5]. Moreover, the incidence in Cantonese ethnicity is much higher than those of other ethnic groups [5,6]. Although the secular trend of incidence of NPC showed a significant decline in some parts of the world, such as Hong Kong, Taiwan, Singapore, and some urban cities in China, the incidence remained stable in the traditional high-incidence areas in southern China over the past decades [5-8].

The pattern of age-specific incidence of NPC is disparate in low- and high-risk areas. The incidence in endemic areas appears to be in a steady increase with age and peaks at age 45 to 60 years, followed by a decrease or plateauing thereafter. In contrast, in non-endemic areas, the incidence of NPC shows a bimodal mode, with a minor peak emerging at late teen and early adulthood (age 15 to 24 years), followed by a second peak later in life (age 65 to 79 years) [9]. NPC occurs more commonly in males than females, with a consistent male to female ratio of 2-3:1 regardless in low- or high-incidence populations [6,9,10].

In accordance to the WHO classification, the histopathology of NPC is categorized into three subtypes, i.e. keratinizing squamous cell carcinoma, nonkeratinizing carcinoma, and basaloid squamous cell carcinoma. Nonkeratinizing carcinoma is subdivided into differentiated and undifferentiated carcinoma [11]. Undifferentiated nonkeratinizing carcinoma comprises the vast majority ($> 95\%$) of the NPC cases in endemic areas such as southern China [12-14]. By contrast, keratinizing squamous cell carcinoma is more common in non-endemic areas [11,15-17]. Moreover, the histological subtype is closely related to geographical distribution, epidemiological risk factors, as well as treatment response and prognosis [11,16].

Radiotherapy is the primary treatment for NPC patients. However, the majority of patients are at an advanced clinical stage (stage III and IV) at the time of diagnosis, owing to the primary anatomic site of the tumor which is located in a silent area with atypical symptoms [14] (**Figure 3**). The prognoses in early stage and advanced stage NPC cases are remarkably different. A 10-year survival rate for patients at stage I could be as high as 98% if treatment is employed promptly and 60% for stage II [18]. By contrast, the 10-year survival rate is 50% or less in patients receiving treatment at advanced stages [19]. Therefore, early diagnosis is critical for improving the outcome of treatment and eventually reducing mortality from NPC.

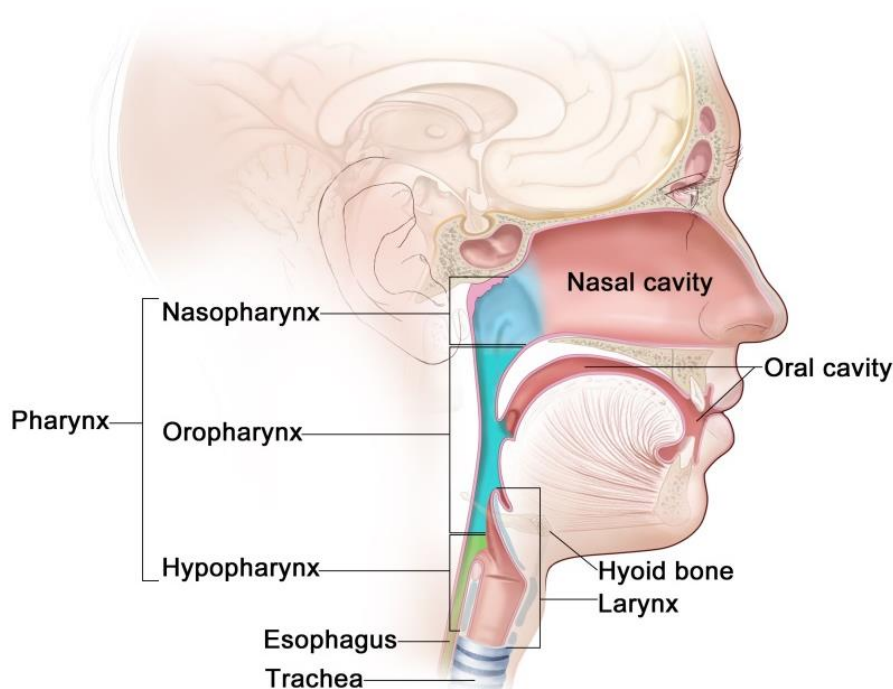


Figure 3. Anatomy of the head and neck (Source: <http://www.cancer.gov/types/head-and-neck>)

Current evidence suggests the oncogenesis of NPC is multifactorial and multistep; genetic factors, infection with EBV, other environmental factors, and their interactions are speculated to be involved in the etiology of this disease [3,20]. The distinct geographic and ethnic distribution of NPC indicates genetic susceptibility might contribute largely to the carcinogenesis of NPC. Over the past two decades, more than 100 association studies, including six genome-wide association studies (GWAS), have been performed to assess the associations of genetic factors with NPC risk. Consistent evidence for the associations were reported for a set of genes, including immune-related human leukocyte antigen (HLA) Class I genes, metabolism genes, DNA repair genes, cell cycle control genes, and cell adhesion/migration genes [3,20,21]. Besides, familial clusters and aggregation of NPC are well documented both in high- and low-incidence regions. Epidemiological studies demonstrate 4- to 10-fold higher risk of NPC in individuals having a first-degree relative with NPC than those without such familial history [13,22-26]. Migrant studies have reported a consistent decline of incidence among Chinese immigrants with longer duration of residency, and a further decrease in incidence among second-generation Chinese migrants compared to first-generation migrants and native Chinese [27,28]. By contrast, the risk of NPC increases for those who migrated from low-incidence areas to high-incidence areas [29]. These clues from migrant studies indicate that except for genetic factor, environmental factors might also contribute significantly to the oncogenesis of NPC.

Infection with EBV has been consistently considered as the most prominent risk factor for the development of NPC, although the exact mechanism through which the virus is involved in the disease process has not been fully clarified. EBV genome can be detected both in epithelial cells with pre-invasive lesions and tumor tissues in the nasopharynx [30,31].

However, the ubiquitous infection with EBV over 95% of the world's population cannot explain the distinct geographical distribution of NPC, and it is still elusive that among the ubiquitously infected people why only a minor fraction of them develop NPC.

Apart from genetic and viral factors, numerous environmental factors have been suggested to be associated with excess risk of NPC in epidemiologic studies, for instance, consumption of salted-fish and other preserved food, cigarette smoking, exposure to occupational risk factors, lower socioeconomic status (SES), and utility of herbal medicines [20,32,33]. Consuming salted fish is the most consistently and well established environmental risk factor for NPC, the relative risk reported in previous studies ranges from 1.4 to 20.2 for intake of salted fish during childhood and from 1.4 to 17.2 for adulthood [33]. However, consumption of salted fish is declining dramatically [34], and in our previous study, Barrett et al. [35] reported a null association between NPC risk and intake of hard Chinese-style salted fish in adulthood, and only a moderately increased risk for consumption during adolescence, suggesting that salted fish may play a smaller role in NPC development than previously reported. In terms of the associations between other environmental factors and NPC, findings from previous studies are often inconsistent. The small sample size, hospital-based study design, targeted populations with discrepant genetic backgrounds, or lack of complete assessment of exposures, and residual confounding might account for the conflicting findings in previous studies.

In addition, the advancement of next-generation sequencing technologies provides novel opportunities to explore the role of microbial factor in NPC development. In some most recent reports, the relationship of oral microbiome with NPC is gradually uncovered and a dysbiosis of oral bacterial microbiome has been linked to NPC [36,37], but the role of oral mycobiome, another core member of oral microbes, in NPC is largely unknown.

In the following content, an overview of the latest progress of epidemiologic studies on NPC will be provided, with emphases on the specific topics related to this thesis work.

2.2 RESIDENTIAL EXPOSURES AND NPC

Residential exposures account for an essential part of the non-viral environmental risk factors of NPC. The associations of NPC with a range of residential exposures have been widely investigated in previous studies, including cooking fuels, type of housing, ventilation condition of home, indoor burning incense, and burning anti-mosquito coil. However, the findings in previous studies are largely inconclusive.

The earliest study that reported the relationship between residential risk factor and NPC dates back as far as 50 years ago [38], in which a significantly increased risk of NPC was found in males who burn incense, suggesting burning incense at home may be a risk factor of NPC. In southern China, a case-control study conducted by Zheng et al. [39] found a 2.3-fold higher risk of NPC for those living in a rural dwelling compared to living in an apartment, but the association was not statistically significant; however, the association with housing type was tested only in childhood, while the association for exposure in adulthood was not examined.

Another study in northeast India, an intermediate-risk region of NPC, reported people living in mud houses had a significant excess risk of NPC than those living in brick/concrete houses (OR=3.46, 95%CI 1.19-10.08) [40]. The association remained significant after adjustment for multiple other environmental factors; but SES was not adjusted, which is a strong confounder for this association. A study carried out in Malaysia Chinese also indicated inferior housing type as an indicator of lower SES was associated with a higher relative risk of NPC [41], whereas the conclusion was derived from univariate analysis without adjusting any potential confounders.

The relationship of NPC with different types of cooking fuels was examined mainly in intermediate- and high-risk areas, and results from previous studies are also not consistent. Three studies conducted by Yu et al. [42-44] in Hong Kong, Guangxi, and Guangzhou in southern China, and one study in north Africa [45] showed no association with using woodfire for cooking; whereas another three studies carried out also in southern China, and some studies in northeast India and Malaysia found a positive association between using wood as a domestic cooking fuel and NPC risk, with odds ratios (ORs) ranging from 2.0 to 5.8 [39,40,46-48].

The results from studies of burning incense have also been largely controversial. Four case-control studies [38,47,49,50] conducted in populations from southern China (three in Hong Kong and one in Guangzhou) found a significantly increased risk of NPC associated with burning incense, and two of them [47,50] even observed a significant exposure-response trend. By contrast, five other studies [41-44,51] (four case-control studies and one register-based cohort study) were not able to replicate those findings.

Most studies examining the association between burning anti-mosquito coil and NPC risk reported a null result [42-44,47,52], with an exception of a study conducted by West et al. in Philippines [53], in which they found after adjustment of potential confounders, individuals who daily burned anti-mosquito coils had a 5.9-fold excess risk of NPC compared with never users, but no excess risk was observed among those with less than daily use. Of the six studies which tested the association of NPC risk with ventilation conditions in house or in kitchen, four have noted a positive association between poor ventilation and NPC risk [52,54-56]. For the other two studies, the one conducted in north Africa [45] found no significant relationship with kitchen ventilation in childhood or adulthood, but an increased risk of NPC was observed among males in childhood in stratified analysis; another one found an interaction between using firewood for cooking and house ventilation, although no overall association was found for house ventilation alone [40].

A handful of residential exposures appear to be associated with NPC risk, e.g., inferior housing type, use of firewood for cooking, and ill-ventilation. Nevertheless, those findings in previous studies are largely inconclusive, and the extent to which residential factors and their interaction with other covariates contribute to the etiology of NPC remains unclear.

2.3 OCCUPATIONAL EXPOSURES AND NPC

Occupational exposure to inhalants and chemicals of various kinds has also been linked to an elevated risk of NPC in epidemiologic studies carried out in both endemic and non-endemic areas [32]. However, the majority of prior evidence comes from studies carried out in low-incidence regions, including United States [57-65], Nordic countries (i.e. Sweden, Denmark, Finland) [15,66-70], United Kingdom [71-73], New Zealand [74], and Northern China (i.e. Shanghai, Tianjin) [75-77]; some evidence is derived from studies conducted in areas with intermediate incidence [53,54,78-86], such as Malaysia, Northeast Thailand, North Africa, Turkey, Philippine, and Taiwan; only a few studies [42,43,46,87,88] from high incidence regions, namely areas of Southern China, for instance, Guangzhou, Guangxi, and Hong Kong, contribute to this research context.

In low-incidence areas, a number of occupational cohorts [64,65,71-73] or national register-based cohort studies [15,68,69] have examined the risk of NPC in association with occupational exposures, such as wood dust and formaldehyde, but results are often null or conflicting. A major concern for these studies is the number of NPC cases is very limited given the rarity of this disease in occupational cohort studies, or the exposure is rare in register-based studies, leading to not sufficient statistical power to detect even moderately excess risks for NPC. For example, a pooled occupational cohort consisting of 28,704 workers in wood related industries presented a significantly elevated risk of NPC (standardized mortality ratio (SMR) = 2.4; 95% confidence interval (CI) = 1.1-4.5) in relation to wood dust exposure; however, the finding is based on only 9 NPC deaths [71]. Two national register-based cohort studies in Denmark [68] and Finland [69] reported no association of NPC risk with exposure to formaldehyde, but there were only 4 and 5 cases in these two cohorts, respectively; whereas a US National Cancer Institute formaldehyde worker cohort study including 10 plants observed a statistically significant increased risk of NPC among formaldehyde exposed workers [89]. However, the positive finding in the NCI cohort was driven heavily by the result in one study plant (6 NPC deaths) with a large SMR (7.34, 95 % CI = 2.69-15.97) while the remaining nine study plants (5 NPC deaths) found no association (SMR = 0.82, 95 % CI = 0.17-2.41) [89]. Therefore, the validity and reliability of evidence from these cohort studies in this research field is still limited. In contrast, evidence from cohort study in high-incidence regions remains lacking.

Compared with cohort study, case-control study is more efficient and frequently employed to identify risk factors for diseases with low incidence. Therefore, the vast majority of currently available evidence on the relationship of occupational exposures with NPC risk is produced from case-control studies. Ideally, case-control studies should include all incident cases with the disease in a defined population during a specified time period; control subjects without the disease should be randomly selected from the source population that generates the cases. However, most previous case-control studies on this topic were implemented with compromise, thereby resulting in, if any, inconsistent findings across studies. For example, a large proportion of previous case-control studies that investigated the link of occupational exposures with NPC are of hospital-based design [53,56,57,79,81,82,88], in which cases were

recruited from hospitals and controls were patients with other diseases from the same hospitals. The problem for a hospital-based design is patient controls may have more comorbidities than controls from the source population, thus selection bias is of great concern. Because the hospital-based controls are not representative of the source population which the cases arise, hence resulting in a distortion of the relative risk, in an unknown direction. It is also problematic that some studies [42,43,54,75,78,87] enrolled neighbors, family members, or friends of the cases as control group. These types of controls may improve the feasibility and participation rate during control recruitment, but they may also introduce an underestimate or a negative bias in the study results. The controls may be too similar to the cases with respect to exposure status, namely, they may not represent the true levels of exposure that exist in the source population due to their similarity to the cases. Using prevalent cases and small sample size could also cause uncertainty to their findings in some previous case-control studies.

Among the specific types of occupational exposures that have been examined for NPC, much attention has been focused on occupational exposure to formaldehyde [90,91] and wood dust or related jobs [92,93]. Although the International Agency for Research on Cancer (IARC) recently drew the conclusion that sufficient evidence from epidemiologic studies supports the carcinogenicity of formaldehyde on NPC [94], one previous meta-analysis and a more recent review show that the relationship between occupational formaldehyde exposure and NPC remains inconclusive, and more additional evidence is warranted; it is also notable that majority of the evidence is derived from low-risk population while evidence from intermediate- or high-risk population is relatively sparse [90,91]. Regarding the association of occupational exposure to wood dust, the findings are quite diverse between studies regardless of the population under study is of low- or high-risk, although recent two meta-analyses reported there was a positive association between occupational exposure to wood dust and NPC [92,93]. Apart from studies that investigated formaldehyde and wood dust, most of previous epidemiological studies, with a few exceptions [77,80,88], examined occupational exposures in a broad or overall fashion.

Almost all previous case-control studies claimed that they have collected a complete occupational history from study subjects. In fact, the lifetime exposure information was usually summarized into cumulative exposure value or roughly categorized into exposed and unexposed in previous studies. The associations with these exposures were usually estimated using general logistic regression, which did not directly account for changes in exposures over lifetime. The time-varying effects of exposures are therefore ignored, which may cause imprecise estimates.

Given the various limitations in previous research, the role of occupational exposures in NPC development remains uncertain, thus evidence in this research field is still very limited, especially in the high-incidence areas.

2.4 ENVIRONMENTAL FACTORS AND EBV REACTIVATION

EBV infects over 95% of the world's population, and it is the first virus found to be associated with a wide range of human cancers including several types of lymphoma and NPC [95,96]. The life cycle of EBV includes latent and lytic phases. After primary infection at early age, EBV is in a form of asymptomatic, life-long latent infection phase in the resting B memory cells in healthy individuals [97]. During EBV latent infection, the virus exists as episome form in B cells, and expresses only a few viral genes. Patterns of latent gene expression differ among EBV-associated malignancies; in the context of NPC, the virus adopts a specific form of latent infection, latency II, and viral genes including EBV-encoded small RNAs (EBERs), nuclear antigen1 (EBNA1), latent membrane proteins (LMPs) are expressed [98]. Furthermore, certain genes have been shown to play a role in viral maintenance (e.g. EBNA1), immune evasion (i.e. by restraining expression of latent genes), and viral carcinogenesis (e.g. LMP1) [98]. In addition, latent EBV infection genes are detected in cancer cells of virtually all cases of NPC from endemic regions, indicating latent EBV infection is a necessary etiologic factor of NPC in endemic areas [31]. In healthy carriers, EBV can be reactivated periodically under certain environmental stress, in which the latently infected cells are triggered into a lytic (i.e. replicative) phase. Upon reactivation to the lytic phase, large numbers of lytic cycle related genes are expressed and viral particle assembly and release increase, which can activate host immune response. Such virus reactivation is reflected by aberrantly elevated antibody levels against multiple EBV antigens, including *BZLF1* transcription activator protein (Zta), viral capsid antigen (VCA), and early antigen (EA) [99,100].

Serological studies have proved that the levels of EBV antibodies among NPC patients are significantly higher than normal population, and individuals with increased EBV antibody levels have a significant excess risk of NPC [101,102]. In addition, several prospective cohort studies have demonstrated that EBV antibodies can elevate several years preceding NPC diagnosis [101-103]. Several serological biomarkers, e.g., VCA/IgA, EA/IgA, and EBNA1/IgA have been exerted to screening and early diagnosis for NPC in endemic areas [104-107]. However, EBV alone is not a sufficient cause of NPC, given virtually all adults worldwide are infected with the virus, while only a fairly small proportion of individuals develop NPC [20]. Moreover, increasing evidence from molecular research shows that the EBV lytic phase also contributes to oncogenesis of NPC, primarily through two pathways: the production of infectious particles to infect more cells; and the regulation of cellular oncogenic pathways by both cell-autonomous and non-cell-autonomous signalling mechanisms [108]. In summary, previous clues indicate that although EBV latent infection is essential in carcinogenesis, EBV lytic infection or reactivation may also contribute substantially to the development of NPC. Although the mechanisms behind EBV reactivation and NPC oncogenesis are not fully understood, identifying environmental factors that can induce EBV reactivation may contribute to the primary prevention for NPC.

Evidence from experimental studies have shown that EBV can be reactivated by a number of chemicals and environmental factors, such as phorbol esters (i.e., TPA), sodium butyrates,

nitrosamines, and extracts from Cantonese-style salted fish, Chinese herbs, or cigarette smoke extracts [100,109,110]. In contrast, epidemiological evidence for the relationship between environmental inducers and EBV reactivation is relatively rare. A study by Xu et al. [100] found that, among the eight life-style factors examined (i.e., family history of NPC, tobacco smoking, and consumption of alcohol, tea, Chinese herbal tea, Cantonese slow-cooked soup, salted fish, and preserved vegetables), only smoking was associated with VCA/IgA seropositivity, and dose-response relationship was observed with various smoking indicators. This observation was established in three independent populations (two populations from a high-risk area with 1,571 male subjects and 1,657 healthy males, respectively; and the other one from a low-risk area with 1,961 healthy males). He et al. [99] examined whether those environmental factors were associated with the levels of other EBV antibodies, including Zta/IgA, EBNA1/IgA, and LMP1/IgA, using the same study populations but fewer subjects (1,498 male subjects) from a high-risk area. Likewise, smoking was the only factor associated with seropositivity for EBNA1/IgA and Zta/IgA, with dose-response effects, but not associated with the levels of LMP1-IgA. However, both studies were conducted among healthy male subjects, and whether the study findings (including the null results) are applicable for female population is unknown. In a screening-based cohort study in southern China, Hu et al. [111] reported that among non-NPC individuals, smoking was associated with higher levels of EBV antibodies of VCA/IgA and EBNA1/IgA tested at baseline (N = 10,110) as well as 3-5 years at follow-up (N = 2,737), while no associations were observed with the other two factors (namely, family history of NPC and salted food consumption). A more recent hospital-based study with 1,026 non-NPC subjects in Hong Kong suggests a possible association between seropositivity of VCA/IgA and sunlight exposure, but no association with vitamin D level, a molecular mediator of sunlight exposure [112]. However, the populations under study might be problematic in these epidemiologic studies, i.e., study subjects were recruited from physical examination programs in the former two studies, voluntary participants were recruited in the screening cohort, and hospital-based control subjects were included in the Hong Kong study. Thus, the representativeness might be compromised in these studies. In addition, although a bunch of environmental factors have been investigated in these studies, many other potential risk factors of NPC such as oral hygiene conditions, ENT disease, medication use, sibship structure, residential exposures, and occupational exposures have not been examined to date. Hence, there is still a large gap of the knowledge in this research topic.

2.5 ORAL FUNGAL COMMUNITIES

The human microbiome, defined as the total collection of microbes that inhabit in a whole range of the body sites, is increasingly recognized as an active part in human body functions, and has been even proposed to be an organ [113].

Bacterial communities make up over 99% of total microbial counts. Apart from the bacteriome, the mycobiome is another major component of human microbiome. Recent

studies demonstrate the importance of our commensal fungal inhabitants as critical contributors to human health and disease [114].

A large proportion of the microorganisms that constitute the human microbiome live within oral cavity, which possesses the second most diverse microbes in our body, namely, the oral microbiome [115]. It is estimated that human oral cavity harbours a complex microbiome consisting of approximately 700 bacterial species [116] and 100 fungal species in healthy individuals [117]. Therefore, the oral mycobiome, which primarily refers to various fungal communities, is a core component of the oral microbiome.

Generally, oral microorganisms form homeostatic communities that live in a mutualistic and harmonious relationship with the host. However, local ecological challenges under certain conditions may undermine the community balance and lead to a microbial dysbiosis characterized by altered microbiome profile and the potential to cause diseases [118,119].

Previous knowledge on the mycobiome was largely limited to culturable fungal communities, which cannot fully characterize the diversity and composition of the mycobiome due to its non-culturable nature. With the use of DNA sequencing technology, which studies the microbiome via detection of the genomic component of microbes, identification of both culturable and non-culturable microbes are possible. However, the low throughput and high cost of the first-generation sequencing method still limit its large-scale application in microbiome study area. Given the limitations in methodology in previous studies, it is not surprising that little progress has been made in exploring the potential role of the oral fungal communities in health and disease.

The advent of novel high-throughput sequencing, also known as next-generation sequencing (NGS), with its high performance in terms of resolution and throughput as well as the decreasing cost, has revolutionized the genomic research. With the advent of NGS techniques, the complexity of the fungal communities that reside within our oral cavity becomes clearer, and in turn offers new insights into the relationship of mycobiome and disease.

In 2010, Ghannoum et al. [117] firstly characterized the “basal” oral mycobiome profile in 20 healthy individuals using NGS, by sequencing the internal transcribed spacer (ITS) region of fungal genome in DNA extracted from oral rinse samples. Seventy-four culturable and 11 non-culturable fungal genera, in total 101 species, were identified in this report. Among the most frequent species, *Candida*, *Aureobasidium*, *Saccharomycetales*, *Aspergillus*, *Fusarium*, and *Cryptococcus*, four of them are known to be pathogenic in humans. A recent NGS study by Perera et al. [120] found the species richness and diversity of mycobiome were significantly lower in oral squamous-cell carcinomas (OSCC) cases compared with intra-oral fibro-epithelial polyps (FEP) controls. The taxonomic abundance at genera and species levels were significantly different between OSCC and FEP. Their findings suggest a dysbiosis in mycobiome is associated with OSCC risk. Another NGS study by Pranab K. et al. [121] reported oral bacterial diversity and richness as well as oral fungal richness were significantly decreased in tumor tissues compared with the adjacent normal tissues from patients with oral

tongue cancer. Abundance of seven fungal genera was significantly different between tumor tissues and normal tissues. Similar observations were also revealed in two studies that investigated the oral mycobiome in patients with head and neck squamous-cell carcinomas (HNSCC) [122,123], in which significantly reduced alpha diversity was observed in HNSCC patients. Besides, one previous study indicated that salivary fungal community composition and diversity were dramatically altered among oral lichen planus patients (OLP) compared to healthy controls, and fungal dysbiosis was associated with the aggravation of OLP [124]. Another line of evidence showed that the abundance of oral genus *Candida* was higher in subjects with periodontal disease (32%) than healthy controls (2.2%) [125].

Recently, our group found that poor oral hygiene was associated with an increased risk of NPC [126], suggesting that oral microbiome (both bacteriome and mycobiome) may be involved in the tumorigenesis. Previous studies and our recent analysis, based on 16S rRNA amplicon sequencing, have shown that the oral bacterial microbiome in NPC patients prior to radiation treatment was significantly different compared to that in control subjects [36,37]. These findings implicate a possible link between oral bacterial dysbiosis and NPC. However, whether the composition of oral fungal communities could play a role in the development of NPC is still unknown to date.

3 RESEARCH AIMS

The overall aim of this thesis is to provide more solid evidence and precise insights into the etiology of NPC, with focus on the associations of NPC with environmental, viral, and other microbial factors.

The specific study aims are:

- To investigate the magnitude and pattern of associations between lifelong residential exposures and the risk of NPC.
- To investigate the association of occupational exposure with the risk of NPC.
- To indentify potential environmental factors for EBV reactivation in a high-incidence area of NPC.
- To characterize the oral fungal profiling in NPC patients and healthy controls, and to explore the relationship between oral fungal communities and the risk of NPC.

4 MATERIALS AND METHODS

4.1 DATA SOURCES AND STUDY DESIGN

4.1.1 Study base

In 2010, we undertook a multicenter collaborative, large, population-based case-control study of NPC in southern China (the NPC Genes, Environment, and EBV [NPCGEE] study) [127]. The four constituted studies in this thesis work were carried out on the basis of the NPCGEE study.

Study base of the NPCGEE study was defined as individuals officially residing in Zhaoqing area of Guangdong Province and the Wuzhou and Guiping/Pingnan areas of Guangxi Autonomous Region in southern China from March 2010 to December 2013 for NPC cases, and from November 2010 to November 2014 for controls. The study areas include 13 cities/counties (Deqing, Fengkai, Gaoyao, Huaiji, Sihui, Zhaoqing, Guangning, Wuzhou, Cenxi, Cangwu, Tengxian, Pingnan, and Guiping) and cover a population of approximately 8 million (**Figure 4**). These 13 cities/counties were chosen because of their high incidence of NPC, their geographic contiguity, their existing opportunities for collaboration with local investigators, and their relatively stable population compared with other, more urban areas in this geographic region. Based on information from local cancer registries, the estimated total number of incident NPC cases is approximately 850 per year [7]. To ensure adequate statistical power for detection of modest associations, we initially aimed to recruit 2,600 cases and 2,600 controls. Eligible participants were aged 20-74 years, residing in the study area, without a history of malignant disease or congenital or acquired immunodeficiency, and physically and mentally able to participate in the study.

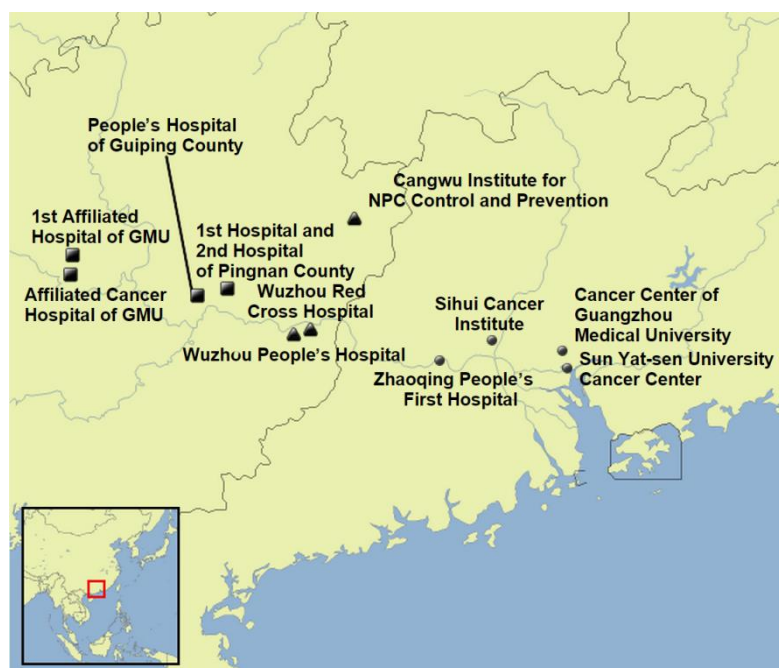


Figure 4. Geographic locations of participating hospitals and institutions. GMU, Guangxi Medical University; NPC, nasopharyngeal carcinoma. (Source: Ref. [127])

4.1.2 NPC case recruitment

Newly diagnosed and histopathologically confirmed NPC cases were recruited through a rapid case ascertainment system involving a network of physicians who diagnosed and/or treated NPC at hospitals in the study area. We hired a total of 12 full-time interviewers along with a few part-time interviewers (contact persons), and 3 full-time study coordinators. Some interviewers were retired nurses from the participating hospitals thus they could also help to identify newly diagnosed NPC cases. Contact persons notified the study personnel (3 study coordinators) as soon as a new NPC case was histopathologically confirmed, after which physician permission was sought to contact each patient. In total, we identified 3,047 eligible incident NPC cases during the study period and 2,554 of these eligible cases (84%) agreed to participate.

4.1.3 Population control recruitment

The same recruitment criteria were applied equally to controls to ensure that both cases and controls arose from the same study base. Controls without NPC frequency matched on age (within 5 years), sex, and geographic area distribution of the cases were randomly selected every 6-12 months from the total population registries covering the study areas.

To rule out the possibility that the selected controls were NPC patients, during interview, self-reported cancer diagnosis history was collected and was validated by village doctors, town hospitals, or community health centers, wherever such records were available. During the study period, two potential controls were detected with NPC, therefore were categorized into case group. We estimated the participation rate in controls to be approximately 10% lower than that of cases, therefore, we increased the number of controls appropriately during control sampling. In total, 3,202 potential controls were randomly selected from the total population of the study area. Of the selected controls, 2,648 (83%) consented to participate.

4.2 DATA COLLECTION AND SAMPLE COLLECTION

Structured, electronic questionnaire interview was implemented to subjects face-to-face by trained interviewers in the study hospitals or at home, as appropriate. A small portion of participants were interviewed by phone due to not able to participate on site. The questionnaire covers demographics, race/ethnicity, body size, residential history, occupational history, history of chronic ear, nose, and throat diseases, family history of NPC and other cancers, cigarette smoking, alcohol and tea drinking, use of Chinese herbal medicine, reproductive history, and dietary habits (i.e. 10 years ago at interview, adolescent, and childhood). Extensive efforts were made to minimize information bias and ensure the quality of questionnaire data. For instance, interviewers were trained with a manual that describes standard survey techniques to be implemented for all participants; logic checks were built into the electronic questionnaire, and interviews were audiotaped for quality control; we periodically analyzed the collected data during recruitment phase to ensure that results were within expectation. We also tried to assign approximately equal numbers of cases and control subjects to each interviewer.

With regard to data collection of residential history, participants were asked to report all residences where they had lived for at least three years from birth. For each residence, subjects were asked to specify the duration of living, housing type (building [concrete structure], cottage [clay brick structure], or boat), main fuel for cooking (wood, coal, kerosene, gas, or electricity), sizes of main rooms (bedroom, hall, and kitchen) and corresponding sizes of windows, kitchen separated from the bedroom (yes or no), exposure to smoke when cooking (a lot, some, a little, or no), burning incense (never, occasionally (during festivals), twice per month (1st and 15th), or daily), burning anti-mosquito coils in summer (yes or no), source of drinking water (tap water, wells, river, or other (natural spring, pond, or stream)), and proximity to a main road, factory, and mining area (< 300 meters, 300-1,000 meters, > 1,000 meters, or unknown).

In the content of occupational history, we assessed the ages at starting and ending full-time work (> 20 hours/week), and all occupations held for ≥ 1 year, listed from the earliest to the most recent job, including housework. For each job, participants were asked to report their job title, age at the beginning and ending of the job, and exposure to any type of occupational dust, chemical vapor, exhaust/smoke, acid, or alkali.

After the interview, biosamples including blood, saliva, hair, and finger and toe nails were also collected. Blood samples were stored at 4 °C for up to three days and then transported to laboratory. Plasma, serum, red blood cell, and buffy coat were isolated and stored at -80 °C. Participants were required not to eat or chew gum for 30 min before saliva collection. Saliva sample (2 to 4 ml) was collected into a 50-ml falcon tube containing 2.5 ml lysis buffer (50 mM Tris, 50 mM EDTA, 50 mM sucrose, 100 mM NaCl, 10% SDS, pH 8.0) and stored at -20 °C less than 3 days before being long-term stored at -80 °C. Hair and nails samples were enclosed into envelopes and stored at room temperature.

4.3 LABORATORY METHODS

4.3.1 EBV serological tests

For **study III**, antibody levels of VCA/IgA (EUROIMMUNAG, Lübeck, Germany) and EBNA1/IgA (Zhongshan Bio-Tech Company, Zhongshan, China) were measured by commercial ELISA kits following the manufactures' instructions. Serum antibody levels of VCA/IgA and EBNA1/IgA were presented as relative optical density (rOD) values, calculated as the ratio of the sample optical density to a reference control (calibrator). Across-batch coefficients of variation (CV) for a control serological sample were 9.1% for VCA/IgA and 9.2% for EBNA1/IgA. Kappa coefficients for test-retest values for approximately 10% of samples that were randomly retested were 0.88 ($P < 0.001$) for VCA/IgA and 0.85 ($P < 0.001$) for EBNA1/IgA.

In this study population of adults in southern China, where virtually all individuals undergo primary EBV infection in early childhood, elevated VCA/IgA and EBNA1/IgA were implicitly assumed to indicate EBV reactivation, as opposed to primary infection. Study subjects were classified as exhibiting serological evidence of EBV reactivation (positive,

Score ≥ 0.65) or not (negative, Score < 0.65) using an EBV-based risk score [106,128]: Score = $[e^{(-3.934 + 2.203 \times \text{VCA/IgA} + 4.797 \times \text{EBNA1/IgA})}] / [1 + e^{(-3.934 + 2.203 \times \text{VCA/IgA} + 4.797 \times \text{EBNA1/IgA})}]$. The EBV-based risk score was established to identify high-risk individuals (i.e. EBV seropositive) for NPC in endemic regions [128,129].

4.3.2 Oral fungal microbiome profiling

In **study IV**, a total of 1,081 saliva samples (542 NPC patients, 539 controls) from the Wuzhou subset of the NPCGEE study were utilized for assessing the oral fungal microbiome among NPC patients and controls.

We extracted DNA from saliva samples using a two-step protocol, which included a lysozyme lysis preprocessing (lysozyme from chicken egg white; Sigma-Aldrich) and a mechanical bead beating, and followed by using the TIANamp blood DNA kit (TIANGEN Biotech Co., Ltd, Beijing, China) as per manufacturer's instructions. Amplicon libraries were constructed by targeting the fungal Internal Transcribed Spacer (ITS)-2 region, using the following primers in the first stage PCR: ITS3ngsmixes: CANCGATGAAGAACGYRG; ITS4ngsUni: CCTSCSCTTANTDATATGC. Primers were selected based on published work illustrating that these were strong candidates for high-throughput sequencing analysis of the full ITS in fungi [130]. Second stage index PCR was performed to barcode the samples with indexes. Amplicon libraries were then purified using an Agencourt AMPure XP purification kit. Final libraries were quantified using a Qubit fluorometer, and the purity of libraries was measured on an Agilent 2100 Bioanalyzer system. Amplicon libraries were pooled at equimolar concentrations. Pooled libraries of 10 pM with 15% phiX were loaded and sequenced on Illumina MiSeq platform using MiSeq Reagent Kit (v.3, 600 cycles) following the 2×300-bp paired-end sequencing protocol. Extraction blanks, respective positive and negative controls for the various PCR steps were run in parallel.

The Illumina-generated demultiplexed paired-end fungal ITS sequences were processed using QIIME 2 (v. 2020.8.0). Primers and adaptors sequences were trimmed using Cutadapt [131]. Sequences were quality filtered and denoised using DADA2 [132] with the parameters based on the Interactive Quality Plot to filter out any phiX and chimeric sequences, and to construct amplicon sequence variants (ASVs) table. Taxonomic assignments were performed with a Naive Bayesian classifier trained against the latest UNITE database (release for QIIME, v. 8.3). Low-abundance ASVs present in less than 2 samples were removed from ASV table. Finally, quality-filtered ASV and taxonomy tables were used for further analysis in R.

4.4 STATISTICAL ANALYSES

4.4.1 Residential exposures and NPC

After exclusion of subjects with ineligible age, missing or poor-quality questionnaire data, 2,533 cases and 2,597 population controls were included in the analysis of **Study I**.

We used multivariate logistic regression models to calculate ORs and 95% CIs for the associations between residential exposures and the risk of NPC. All logistic regression models included the frequency matching variables (age, sex, and geographic area [Zhaoqing, Wuzhou, or Guiping/Pingnan]) as well as potential confounders selected based on known or suspected risk factors for NPC that are plausibly related to residential exposures. The ORs and 95% CIs were estimated for overall lifetime exposure (ever exposed vs. never exposed), and at four specific time points set at age 10 years, 18 years, 30 years, and the most recent 10 years preceding diagnosis/interview.

Linear trend tests were performed to assess potential duration-response relationship between duration of residential exposure and NPC risk. We conducted a sensitivity analysis that treated residential exposures as time-varying variables to account for changes in residential covariate values over time within individuals, and a weighted Cox regression model was used. Effect modifications by sex, age, educational level, first-degree family history of NPC, current occupation, or cigarette smoking were also evaluated.

4.4.2 Occupational exposures and NPC

In **Study II**, upon completion of all data cleaning, 2,514 cases and 2,586 controls were included in the final analysis. Unconditional logistic regression was used to estimate ORs and 95% CIs of NPC risk in relation to occupational exposures, adjusted for the frequency matching variables and potential confounders selected based on previous knowledge.

Duration of occupational exposure and age at first exposure were categorized by tertiles of the distribution among controls. We further used restricted cubic splines to characterize the relationship between risk of NPC and duration of occupational exposure. Linearity of the relationships was examined via the Wald chi-squared test [133]. We pre-set the knots at the 5th, 35th, 65th, and 95th percentiles of the distribution of duration based on the sample size of present study [134].

We conducted likelihood ratio tests to assess potential effect modifications by sex, age, education level, housing type, family history of NPC, and cigarette smoking. Joint associations between occupational exposures were also examined, and the additive interactions were evaluated by the relative excess risk due to interaction (RERI) [135].

4.4.3 Environmental factors and EBV reactivation

Given EBV is reactivated in virtually all NPC patients, we assessed the association of environmental factors with EBV reactivation only among the population-based controls, excluding the NPC cases. After data cleaning for ineligible subjects, and subjects with missing data or without blood samples, there were 1,916 subjects (74% of the total control subjects) included in the final analysis of **Study III**. No differences in age (chi-squared test, $P = 0.98$), sex ($P = 0.95$), or education ($P = 0.55$) were found between the final dataset ($N = 1,916$) and the full dataset ($N = 2,597$) of controls.

We calculated adjusted ORs and corresponding 95% CIs for associations between EBV reactivation and environmental factors using unconditional logistic regression, adjusting for age, sex, geographic area, and educational level. Linear trend tests for associations between environmental factors and EBV reactivation were performed by using the median value within each category or by treating the categorical variable as an ordinal variable, as applicable.

4.4.4 Oral fungal microbiome and NPC

In **Study IV**, we used a subset of saliva samples collected at the Wuzhou site of the NPCGEE study. Of the 1,081 sequenced samples, two had ambiguous labels and could not be matched to the metadata, and four with less than 1,000 reads after quality filtering were removed, leaving 1,075 (538 cases, 537 controls) samples in further analysis.

Sample metadata, ASV table, taxonomy table, and representative sequences were integrated into a “Phyloseq” [136] object in R. Data set was rarefied to 5,000 reads for alpha and beta diversity analyses. Observed ASVs, Simpson index, and Shannon diversity were calculated to assess alpha diversity. Bray-Curtis dissimilarity index was computed for evaluation of beta diversity. Principal coordinates analysis (PCoA) plot was used for visualization of beta diversity. Permutational multivariate analysis of variance (PERMANOVA) analyses adjusted for age, sex, and sequencing run were performed to examine whether overall fungal composition differed by disease status. Differentially abundant fungal organisms between cases and controls were identified using linear discriminant analysis (LDA) effect size (LEfSe) [137]. For between group comparisons, Wilcoxon, Mann-Whitney-U and/or Kruskal-Wallis tests with Benjamin-Hochberg correction for false discovery were performed where appropriate. Categorical data was compared using Chi-squared or Fisher exact tests as appropriate.

All analyses in this thesis were conducted using SAS (version 9.4, SAS Institute, Cary, NC) and R (version 4.1, R Foundation for Statistical Computing, Vienna, Austria). All statistical tests were two-sided, and a p-value < 0.05 was considered statistically significant.

4.5 ETHICAL CONSIDERATIONS

This thesis work is based on the NPCGEE study, which was approved by the institutional review boards of the Harvard T.H. Chan School of Public Health, the Institute for Viral Disease Control and Prevention of the Chinese Center for Disease Control and Prevention, Sun Yat-sen University Cancer Center, Guangxi Medical University, and the Regional Ethical Review Board in Stockholm, Sweden. Written or oral informed consent was obtained from all study participants during the interview. All participants provided consent of the future utility of questionnaire data and biospecimens. The questionnaire data used in this thesis is stored on the server at the Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden. The identification number of each participant was replaced by a pseudo identification number of electronic questionnaire, which is used for linkages between sections of questionnaire data by researchers. The researchers do not have the access

to the participant's identification information. The host institute secures the data through strict safety guidelines and limited access.

5 RESULTS

5.1 RESIDENTIAL EXPOSURES AND RISK OF NPC

The demographic characteristics of incident NPC cases and population-based controls are shown in **Table 1**. Compared with controls, cases were slightly younger, less educated, and more likely to have a blue-collar job, have a first-degree family history of NPC, have ever smoked, have more filled teeth, drink less tea, and have consumed salted fish at least weekly.

Table 1. Baseline characteristics of the study participants

Characteristics	Cases (N = 2,533) n (%)	Controls (N = 2,597) n (%)	P Value
Geographic area			0.31
Zhaoqing	1,286 (50.8)	1,321 (50.9)	
Wuzhou	688 (27.2)	665 (25.6)	
Guiping/Pingnan	559 (22.1)	611 (23.5)	
Sex			0.95
Male	1,860 (73.4)	1,909 (73.5)	
Female	673 (26.6)	688 (26.5)	
Age at diagnosis/referral, years			< 0.001
Mean (SD)	48.5 (10.7)	49.8 (10.9)	
Educational level, years			0.004
≤ 6	1,005 (39.7)	932 (35.9)	
7-9	1,014 (40.0)	1,040 (40.1)	
10-12	407 (16.1)	484 (18.6)	
> 12	107 (4.2)	141 (5.4)	
Current occupation			< 0.001
Farmer	855 (33.8)	984 (37.9)	
Blue-collar	1,023 (40.4)	900 (34.7)	
White-collar	350 (13.8)	416 (16.0)	
Other/unknown	227 (12.1)	201 (11.4)	
First-degree family history of NPC			< 0.001
No	2,208 (87.2)	2,483 (95.6)	
Yes	272 (10.7)	70 (2.7)	
Unknown	47 (1.9)	43 (1.7)	
Cigarette smoking			0.08
Never	1,122 (44.3)	1,213 (46.7)	
Ever	1,410 (55.7)	1,382 (53.2)	
Filled teeth			< 0.001
No	2,082 (82.2)	2,224 (85.6)	
Yes	451 (17.8)	373 (14.4)	
Tea drinking			< 0.001
Less than daily	1,618 (63.9)	1,513 (58.3)	
Daily	911 (36.0)	1,081 (41.6)	
Salted fish consumption 10 years ago			0.02
Yearly or less	1,858 (73.4)	1,902 (73.2)	
Monthly	484 (19.1)	542 (20.9)	
Weekly or more	188 (7.4)	149 (5.7)	

This table is reproduced from Y Chen et al. *Environment International* 2021. [138]

The adjusted ORs for residential exposures in association with the risk of NPC are shown in **Table 2**. Compared with those living in a building over lifetime, NPC risk was significantly increased for individuals who ever lived in a cottage (OR = 1.56; 95% CI = 1.34-1.81) or a boat (OR = 3.87; 95% CI = 2.07-7.21). The ORs (95% CIs) for those who ever used wood, coal, and kerosene as cooking fuel vs. those using only gas or electricity for cooking were 1.34 (1.03-1.75), 1.70 (1.17-2.47), and 3.58 (1.75-7.36), respectively. Individuals ever utilized well, river, or other sources (i.e. spring, pond, or stream) as water source were at 1.5 to 2-fold excess risks of NPC than those who used only tap water all the time. NPC risk was also increased in individuals living in houses with poor ventilation, exposure to cooking smoke, having a habit of burning incense, and living proximity to a factory (**Table 2**). For instance, ORs were 3.08 (2.46-3.86) for smaller-sized vs. larger bedroom windows, 1.53 (1.20-1.94) for exposure to a lot cooking smoke vs. no smoke, 1.59 (1.31-1.95) for daily vs. never/occasionally burning incense; 1.53 (1.27-1.83) for living proximity to a factory < 300 meters vs. > 1000 meters. Of note, a modest but inverse significant association with burning anti-mosquito coils in summer (OR = 0.80; 95% CI = 0.70-0.92) was observed. For most of the residential exposures, a significant exposure-response relationship with longer duration of exposure was noted (P for trend tests < 0.05).

When we assessed the associations between residential exposures and risk of NPC at four specific residential periods, we found the magnitude of the associations was stronger for housing type, cooking fuel, source of drinking water, cooking smoke, and ventilation indicators at an early age (i.e., at age 10 years) than later in life (**Table 3**).

Weighted Cox regression analysis, in which the residential exposures were treated as time-varying variables, largely corroborated the results observed in logistic regression model (data not shown). We did not observe significant effect modification in the observed associations by sex, age, educational level, family history of NPC, cigarette smoking status, or current occupation (P for heterogeneity for all above potential modifiers > 0.05) (data not shown).

Table 2. Associations of nasopharyngeal carcinoma risk with lifetime residential exposures

Variables	Cases (N = 2,533) n (%)	Controls (N = 2,597) n (%)	OR (95% CI)	<i>P</i> trend for duration
House category				
Building only	404 (15.9)	566 (21.8)	1.00 (Ref.)	
Ever cottage	2,111 (83.3)	2,025 (78.0)	1.56 (1.34-1.81)	< 0.001
Ever boat	38 (1.5)	15 (0.6)	3.87 (2.07-7.21)	< 0.001
Cooking fuel				
Gas/electricity only	106 (4.2)	145 (5.6)	1.00 (Ref.)	
Ever wood	2,406 (95.0)	2,437 (93.8)	1.34 (1.03-1.75)	0.40
Ever coal	128 (5.1)	101 (3.9)	1.70 (1.17-2.47)	0.04
Ever kerosene	33 (1.3)	12 (0.5)	3.58 (1.75-7.36)	0.01
Source of drinking water				
Tap water only	411 (16.2)	627 (24.1)	1.00 (Ref.)	
Ever wells	1,422 (56.1)	1,390 (53.5)	1.57 (1.34-1.83)	0.009
Ever river	378 (14.9)	340 (13.1)	1.80 (1.47-2.21)	< 0.001
Ever other (spring, pond, stream)	709 (28.0)	529 (20.4)	2.03 (1.70-2.41)	< 0.001
Cooking smoke				
No smoke only	151 (6.0)	260 (10.0)	1.00 (Ref.)	
Ever a little smoke	820 (32.4)	1,021 (39.3)	1.41 (1.12-1.78)	0.78
Ever some smoke	1,124 (44.4)	997 (38.4)	2.05 (1.63-2.58)	< 0.001
Ever a lot of smoke	671 (26.5)	768 (29.6)	1.53 (1.20-1.94)	0.08
Burning incense				
Never/occasionally only	1,197 (47.3)	1,357 (52.3)	1.00 (Ref.)	
Ever twice per month	1,111 (43.9)	1,088 (41.9)	1.18 (1.05-1.33)	0.02
Ever every day	306 (12.1)	212 (8.2)	1.59 (1.31-1.95)	< 0.001
Burning mosquito coils in summer				
Never	707 (27.9)	609 (23.5)	1.00 (Ref.)	
Ever	1,826 (72.1)	1,988 (76.5)	0.80 (0.70-0.92)	< 0.001
Proximity to a factory (meters)				
> 1,000 only	1,800 (71.1)	2,099 (80.8)	1.00 (Ref.)	
Ever 300-1,000, not < 300	217 (8.6)	186 (7.2)	1.29 (1.04-1.59)	0.26
Ever < 300	367 (14.5)	255 (9.8)	1.53 (1.27-1.83)	0.07
Unknown	187 (7.4)	74 (2.8)		
Bedroom windows (m ²) ^a				
size > 3	197 (7.8)	420 (16.2)	1.00 (Ref.)	
2 < size ≤ 3	378 (14.9)	509 (19.6)	1.73 (1.38-2.17)	
1 < size ≤ 2	928 (36.6)	891 (34.3)	2.56 (2.07-3.17)	
0 < size ≤ 1	766 (30.2)	666 (25.6)	3.08 (2.46-3.86)	
0	44 (1.7)	64 (2.5)	1.86 (1.19-2.91)	
Hall windows (m ²) ^a				
size > 5	278 (11.0)	495 (19.1)	1.00 (Ref.)	
4 < size ≤ 5	222 (8.8)	280 (10.8)	1.43 (1.13-1.81)	
2 < size ≤ 4	700 (27.6)	646 (24.9)	1.97 (1.63-2.39)	
0 < size ≤ 2	628 (24.8)	628 (24.2)	1.89 (1.55-2.31)	
0	323 (12.8)	427 (16.4)	1.61 (1.25-2.07)	
Kitchen windows (m ²) ^a				
size > 3	201 (7.9)	342 (13.2)	1.00 (Ref.)	
2 < size ≤ 3	365 (14.4)	452 (17.4)	1.37 (1.09-1.72)	
1 < size ≤ 2	721 (28.5)	782 (30.1)	1.57 (1.27-1.94)	
0 < size ≤ 1	735 (29.0)	764 (29.4)	1.67 (1.34-2.08)	
0	87 (3.4)	131 (5.0)	1.17 (0.83-1.64)	

This table is reproduced and adapted from Y Chen et al. *Environment International* 2021. [138]

^a A weighted average was calculated as the overall lifetime exposure level for indicators of ventilation, therefore no trend tests for duration of exposure were performed.

Table 3. Associations of nasopharyngeal carcinoma risk with residential exposures, at specific residential period

Variables	Periods of residential history			
	Age 10 years OR (95% CI)	Age 18 years OR (95% CI)	Age 30 years OR (95% CI)	Last 10 years OR (95% CI)
House category				
Building	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
Cottage	1.49 (1.29-1.72)	1.35 (1.18-1.54)	1.29 (1.14-1.46)	1.21 (1.07-1.37)
Boat	3.04 (1.40-6.61)	2.57 (1.17-5.63)	2.23 (0.92-5.43)	3.16 (1.12-8.92)
Cooking fuel				
Gas/electricity	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
Wood	1.38 (1.07-1.77)	0.95 (0.78-1.17)	0.92 (0.79-1.07)	0.87 (0.76-1.00)
Coal/kerosene	1.39 (0.74-2.61)	1.12 (0.67-1.88)	1.70 (1.11-2.59)	1.68 (1.13-2.49)
Source of drinking water				
Tap water	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
Well	1.45 (1.25-1.69)	1.26 (1.09-1.46)	1.03 (0.89-1.18)	0.95 (0.82-1.09)
River	1.72 (1.40-2.12)	1.50 (1.22,1.85)	1.20 (0.92-1.55)	1.14 (0.83-1.58)
Other (spring, pond, stream)	1.71 (1.42-2.05)	1.48 (1.24-1.77)	1.46 (1.23-1.74)	1.44 (1.22-1.71)
Cooking smoke				
No smoke	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
A little smoke	1.46 (1.11-1.93)	1.38 (1.07-1.80)	1.14 (0.92-1.40)	1.06 (0.88-1.28)
Some smoke	2.28 (1.76-2.96)	2.09 (1.63-2.68)	1.77 (1.44-2.18)	1.62 (1.33-1.97)
A lot of smoke	1.41 (1.08-1.84)	1.29 (1.00-1.66)	1.33 (1.06-1.67)	1.35 (1.07-1.70)
Burning incense				
Never/occasionally	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
Twice per month	1.11 (0.98-1.25)	1.13 (1.00-1.28)	1.09 (0.96-1.23)	1.13 (1.00-1.28)
Every day	1.34 (1.04-1.74)	1.43 (1.11-1.85)	1.59 (1.26-2.02)	1.73 (1.39-2.15)
Burning mosquito coils in summer				
No	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
Yes	0.84 (0.75-0.95)	0.79 (0.70-0.88)	0.81 (0.72-0.92)	0.77 (0.68-0.87)
Bedroom windows (m ²)				
size > 3	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
2 < size ≤ 3	1.94 (1.45-2.59)	1.71 (1.30-2.24)	1.57 (1.24-1.98)	1.31 (1.06-1.62)
1 < size ≤ 2	3.61 (2.78-4.69)	2.93 (2.29-3.76)	2.48 (2.00-3.08)	1.92 (1.58-2.33)
0 < size ≤ 1	3.73 (2.86-4.85)	3.23 (2.51-4.15)	3.16 (2.53-3.95)	2.77 (2.25-3.41)
0	3.79 (2.79-5.16)	3.24 (2.40-4.38)	2.89 (2.13-3.93)	1.74 (1.26-2.40)
Hall windows (m ²)				
size > 5	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
4 < size ≤ 5	1.32 (1.02-1.70)	1.20 (0.93-1.54)	1.21 (0.94-1.56)	1.09 (0.87-1.38)
2 < size ≤ 4	2.00 (1.63-2.46)	1.81 (1.48-2.22)	1.66 (1.37-2.00)	1.48 (1.25-1.77)
0 < size ≤ 2	2.19 (1.75-2.73)	2.11 (1.70-2.63)	2.00 (1.63-2.47)	1.91 (1.58-2.33)
0	1.83 (1.45-2.30)	1.72 (1.38-2.16)	1.88 (1.51-2.35)	1.60 (1.30-1.98)
Kitchen windows (m ²)				
size > 3	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
2 < size ≤ 3	1.73 (1.35-2.22)	1.68 (1.32-2.16)	1.19 (0.94-1.52)	1.21 (0.97-1.52)
1 < size ≤ 2	1.76 (1.40-2.20)	1.67 (1.33-2.09)	1.59 (1.28-1.97)	1.37 (1.12-1.67)
0 < size ≤ 1	2.29 (1.83-2.87)	2.31 (1.84-2.90)	2.05 (1.64-2.55)	2.04 (1.65-2.51)
0	1.26 (0.98-1.63)	1.30 (1.01-1.68)	1.41 (1.09-1.82)	1.37 (1.06-1.76)

This table is reproduced and adapted from Y Chen et al. *Environment International* 2021. [138]

5.2 OCCUPATIONAL EXPOSURES AND RISK OF NPC

After fully adjusting for confounding factors, we found individuals exposed to broad categories of occupational dusts (excluding soil dust) (OR = 1.45, 95% CI = 1.26-1.68), chemical vapors (excluding pesticides) (OR = 1.37, 95% CI = 1.17-1.61), exhausts/smokes (OR = 1.42, 95% CI = 1.25-1.60), or acids/alkalis (OR = 1.56, 95% CI = 1.30-1.89) in the workplace showed an elevated risk of NPC compared with those unexposed (**Table 4**). The risk of NPC increased significantly with experiencing multiple pollutants of the four broad categories of occupational exposures examined (P for trend < 0.001). For instance, the OR was 2.05 (95% CI = 1.36-3.10) for subjects who experienced all four occupational exposures compared to those without any exposures. Relative risk estimates for all four categories of occupational exposures appeared to linearly increase with increasing duration of exposure (P values for non-linear association were > 0.10 for all four categories of occupational exposures) (**Figure 5**). Stronger associations were seen with earlier age at first exposure for exhausts/smokes and acids/alkalis exposures, but not for dusts and chemical vapors (data not shown).

Table 4. Associations between categories of occupational exposures and the risk of nasopharyngeal carcinoma

Occupational exposures	Controls, n (%)	Cases, n (%)	OR (95% CI)
Dust exposure			
None	782 (30.4)	624 (25.2)	1.00 (ref.)
Soil dust only, ever	842 (32.7)	733 (29.6)	1.13 (0.97-1.32)
Dust except soil dust, ever	949 (36.9)	1120 (45.2)	1.45 (1.26-1.68)
Chemical vapor exposure			
None	1117 (43.5)	1009 (40.8)	1.00 (ref.)
Pesticide only, ever	1026 (40.0)	908 (36.7)	0.98 (0.85-1.12)
Chemical vapor except pesticide, ever	423 (16.5)	554 (22.4)	1.37 (1.17-1.61)
Smoke/exhaust exposure			
None	1721 (67.3)	1446 (58.9)	1.00 (ref.)
Ever	838 (32.7)	1008 (41.1)	1.42 (1.25-1.60)
Acid/alkali exposure			
None	2338 (91.5)	2123 (87.2)	1.00 (ref.)
Ever	218 (8.5)	311 (12.8)	1.56 (1.30-1.89)
Any occupational exposures			
None	369 (19.6)	245 (12.2)	1.00 (ref.)
Any	1514 (80.4)	1763 (87.8)	1.69 (1.40-2.03)
One exposure	872 (46.3)	911 (45.4)	1.54 (1.27-1.87)
Two exposures	419 (22.3)	547 (27.2)	1.92 (1.55-2.38)
Three exposures	174 (9.2)	232 (11.6)	1.88 (1.44-2.45)
Four exposures	49 (2.6)	73 (3.6)	2.05 (1.36-3.10)
P for trend			< 0.001

This table is reproduced and adapted from Y Chen et al. *Cancer* 2021. [139]

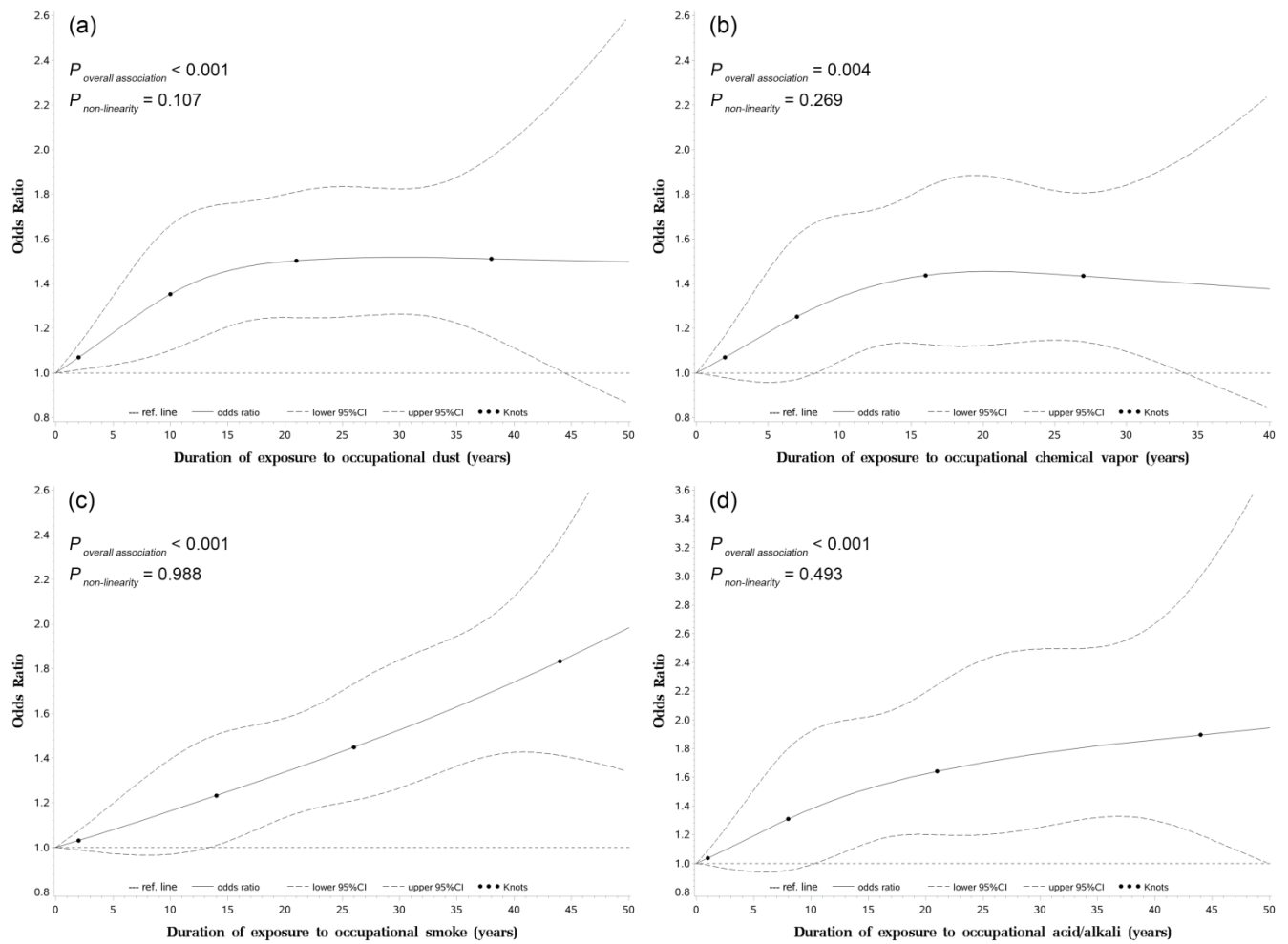


Figure 5. Duration-response relationships between occupational exposures and risk of nasopharyngeal carcinoma using restricted cubic spline function. (This figure is reproduced from Y Chen et al. *Cancer* 2021. [139])

The associations of NPC risk with subtypes within the four broad categories of occupational exposures are presented in **Figure 6**. Individuals exposed to dust from metals, textiles, cement, or coal had a 30%-61% higher risk of NPC compared with those unexposed. Regarding types of chemical vapor, occupational exposure to formaldehyde, organic solvents, or dye showed a 1.6 to 2.3-fold higher risk of NPC. Occupational exposure to exhaust or smoke, including diesel exhaust, firewood smoke, asphalt/tar smoke, vehicle exhaust, and welding smoke also conferred a significant excess risk of NPC (ORs = 1.35 to 2.12). Exposure to occupational acids (including sulfuric acid, hydrochloride, and nitric acid) or alkalis (including concentrated alkali and ammonia) was associated with a 1.6-fold excess risk of NPC. Most of the subtypes of occupational exposures showed positive trends with increasing duration of exposure or earlier age at first exposure (data not shown).

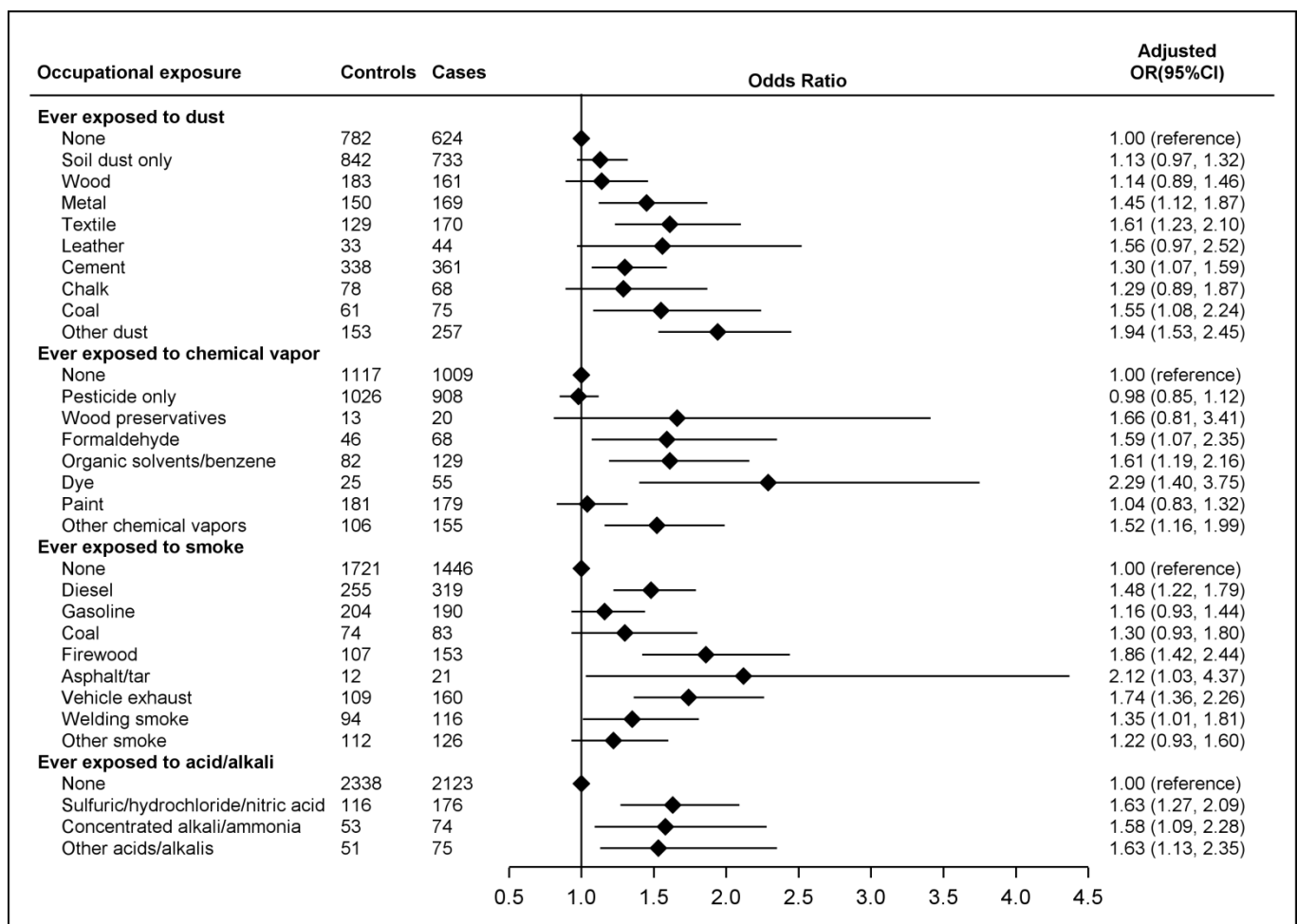


Figure 6. Associations between subtypes of occupational exposures and risk of nasopharyngeal carcinoma. (This figure is reproduced from Y Chen et al. *Cancer* 2021. [139])

5.3 ENVIRONMENTAL FACTORS AND EBV REACTIVATION

The characteristics of 1,916 population-based control subjects stratified by EBV reactivation status are presented in **Table 5**. The overall prevalence of EBV reactivation was 22.5% (431/1,916), with higher levels among older individuals and those who lived in the Zhaoqing area. EBV reactivation status did not differ by sex, educational level, first-degree family history of NPC, and body mass index (BMI) 10 years prior to the interview.

Table 5. The Characteristics of 1,916 population-based control subjects stratified by EBV reactivation status

Characteristics	EBV reactivation status ^a		P value ^b
	Negative (N = 1,485)	Positive (N = 431)	
	n (%)	n (%)	
Area			0.01
Zhaoqing	585 (74.2)	203 (25.8)	
Wuzhou	455 (78.9)	122 (21.1)	
Guiping&Pingnan	445 (80.8)	106 (19.2)	
Sex			0.33
Female	400 (79.1)	106 (20.9)	
Male	1085 (77.0)	325 (23.0)	
Age, y			< 0.001
20-29	44 (78.6)	12 (21.4)	
30-39	222 (82.5)	47 (17.5)	
40-49	527 (80.8)	125 (19.2)	
50-59	424 (76.8)	128 (23.2)	
60-74	268 (69.3)	119 (30.7)	
Educational level, y			0.06
≥ 10	501 (75.1)	166 (24.9)	
7-9	632 (80.1)	157 (19.9)	
≤ 6	352 (76.5)	108 (23.5)	
First-degree family history of NPC			0.91
No	1417 (77.5)	412 (22.5)	
Yes	43 (79.6)	11 (20.4)	
Unknown	25 (75.8)	8 (24.2)	
BMI 10 y ago (kg/m2)			0.11
< 18.5	149 (73.8)	53 (26.2)	
18.5-22.9	958 (79.0)	255 (21.0)	
23.0-27.4	334 (76.3)	104 (23.7)	
≥ 27.5	42 (68.9)	19 (31.1)	

Abbreviations: EBV, Epstein-Barr virus; NPC, nasopharyngeal carcinoma; BMI, body mass index; VCA/IgA, IgA antibodies against EBV capsid antigens; EBNA1/IgA, IgA antibodies against EBV nuclear antigen1.

^a Two EBV serological markers (VCA/IgA, EBNA1/IgA) were used to determine the status of EBV reactivation. An EBV score was calculated using a formula: $Score = [e^{(-3.934 + 2.203 \times VCA/IgA + 4.797 \times EBNA1/IgA)}] / [1 + e^{(-3.934 + 2.203 \times VCA/IgA + 4.797 \times EBNA1/IgA)}]$. $Score < 0.65$ was defined as negative, while $Score \geq 0.65$ was defined as positive.

^b P values were determined using the chi-squared test.

Generally, we found no associations between EBV reactivation and extensive lifestyle factors, including alcohol drinking, tea drinking, ENT diseases, use of medication and herbs, consumption of salted fish or preserved foods, oral hygiene conditions, and sibling structure (**Figure 7**). However, a higher risk of EBV reactivation was seen among current smokers (OR = 1.37, 95% CI = 1.02-1.83), but not in former smokers, compared to non-smokers. Among ever smokers, earlier age at smoking initiation, longer duration of smoking, more cumulative pack-years of smoking, consumption of unfiltered cigarettes, and having ever engaged in deep inhalation when smoking all exhibited significant exposure-response trends with the risk of EBV reactivation (data not shown).

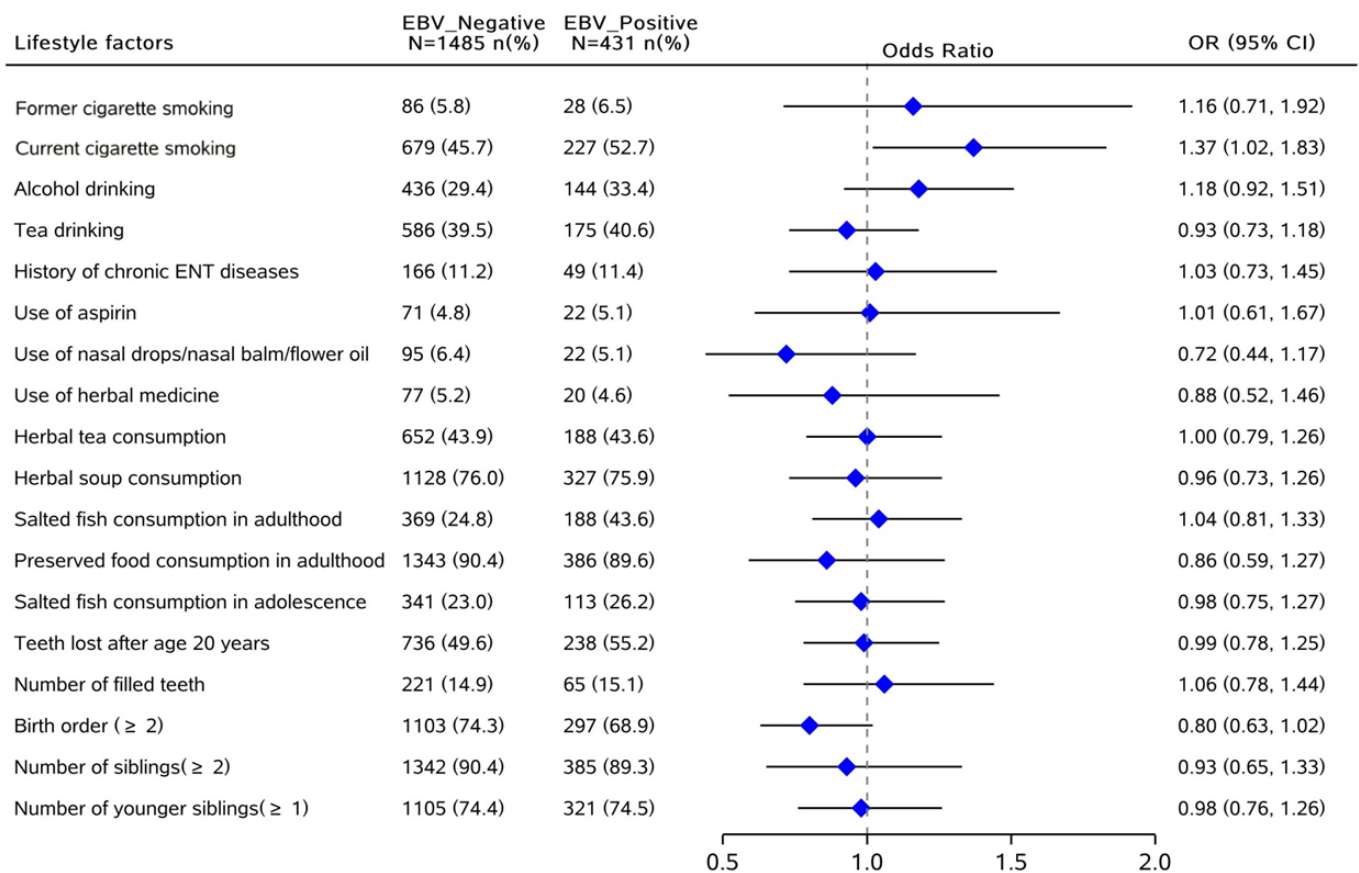


Figure 7. Associations between lifestyle factors and EBV reactivation.

Abbreviations: OR, odds ratio; CI, confidence interval; EBV, Epstein-Barr virus; ENT, ear, nose, and throat; VCA/IgA, IgA antibodies against EBV capsid antigens; EBNA1/IgA, IgA antibodies against EBV nuclear antigen1.

Two EBV serological markers (VCA/IgA, EBNA1/IgA) were used to determine the status of EBV reactivation. An EBV score was calculated using a formula: $\text{Score} = \frac{e^{(-3.934 + 2.203 \times \text{VCA/IgA} + 4.797 \times \text{EBNA1/IgA})}}{1 + e^{(-3.934 + 2.203 \times \text{VCA/IgA} + 4.797 \times \text{EBNA1/IgA})}}$. Score < 0.65 was defined as negative, while Score ≥ 0.65 was defined as positive.

OR estimates were calculated using logistic regression, adjusted for age (continuous variable), sex, geographic area, and educational level.

None of the residence-related exposures, including types of housing, cooking fuel, and water source, and factors contributing to household air pollution, and ventilation of the home, were associated with EBV reactivation (**Figure 8**). Exposure to occupational dust showed a marginally inverse association with EBV reactivation (OR = 0.78; 95% CI = 0.62-0.98), whereas no associations were observed with exposure to occupational chemical vapors, smokes/exhausts, or acids/alkalis (**Figure 8**).

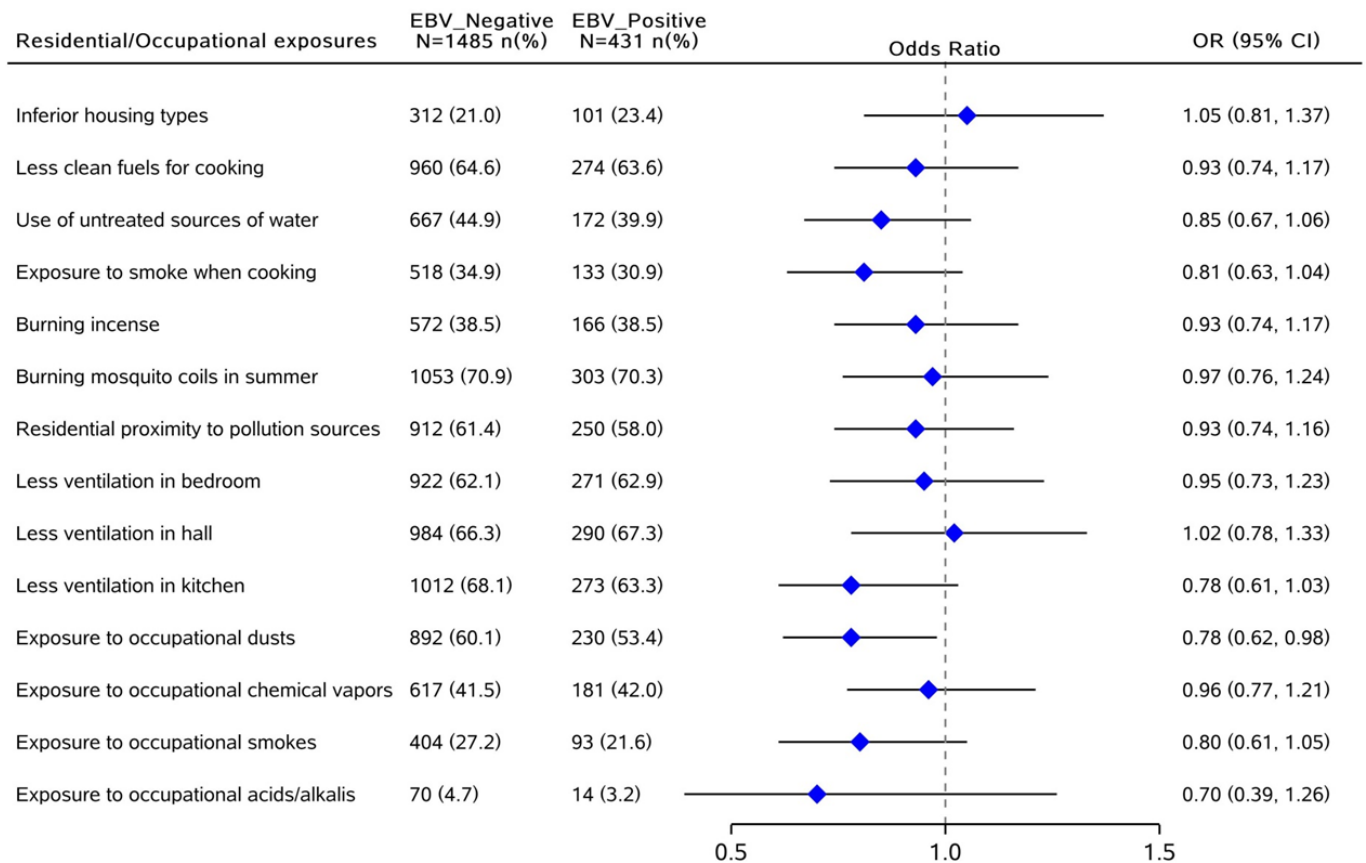


Figure 8. Associations between residential exposures, occupational exposures and EBV reactivation.

Abbreviations: OR, odds ratio; CI, confidence interval; EBV, Epstein-Barr virus; VCA/IgA, IgA antibodies against EBV capsid antigens; EBNA1/IgA, IgA antibodies against EBV nuclear antigen1. Two EBV serological markers (VCA/IgA, EBNA1/IgA) were used to determine the status of EBV reactivation. An EBV score was calculated using a formula: $\text{Score} = \frac{e^{(-3.934 + 2.203 \times \text{VCA/IgA} + 4.797 \times \text{EBNA1/IgA})}}{1 + e^{(-3.934 + 2.203 \times \text{VCA/IgA} + 4.797 \times \text{EBNA1/IgA})}}$. Score < 0.65 was defined as negative, while Score \geq 0.65 was defined as positive. OR estimates were calculated using logistic regression, adjusted for age (continuous variable), sex, geographic area, and educational level.

5.4 ORAL FUNGAL PROFILING AND THE RISK OF NPC

After denoising and quality filtering, there were 1,075 subjects (538 NPC cases, 537 controls) of whom sufficient high-quality sequences were available to be retained for analysis.

The top 2 abundant fungal phyla, Ascomycota (87.9%) and Basidiomycota (11.6%), accounted for 99.5% of the total abundance in the healthy population controls' oral samples.

The 6 most abundant fungal genera (comprising 66.5% of all genera) in controls were *Malassezia* (17.5%), *Candida* (15.1%), *Cladosporium* (13.8%), *Aspergillus* (8.5%), *Blumeria* (8.1%), and *Wallemia* (3.5%) (**Figure 9**).

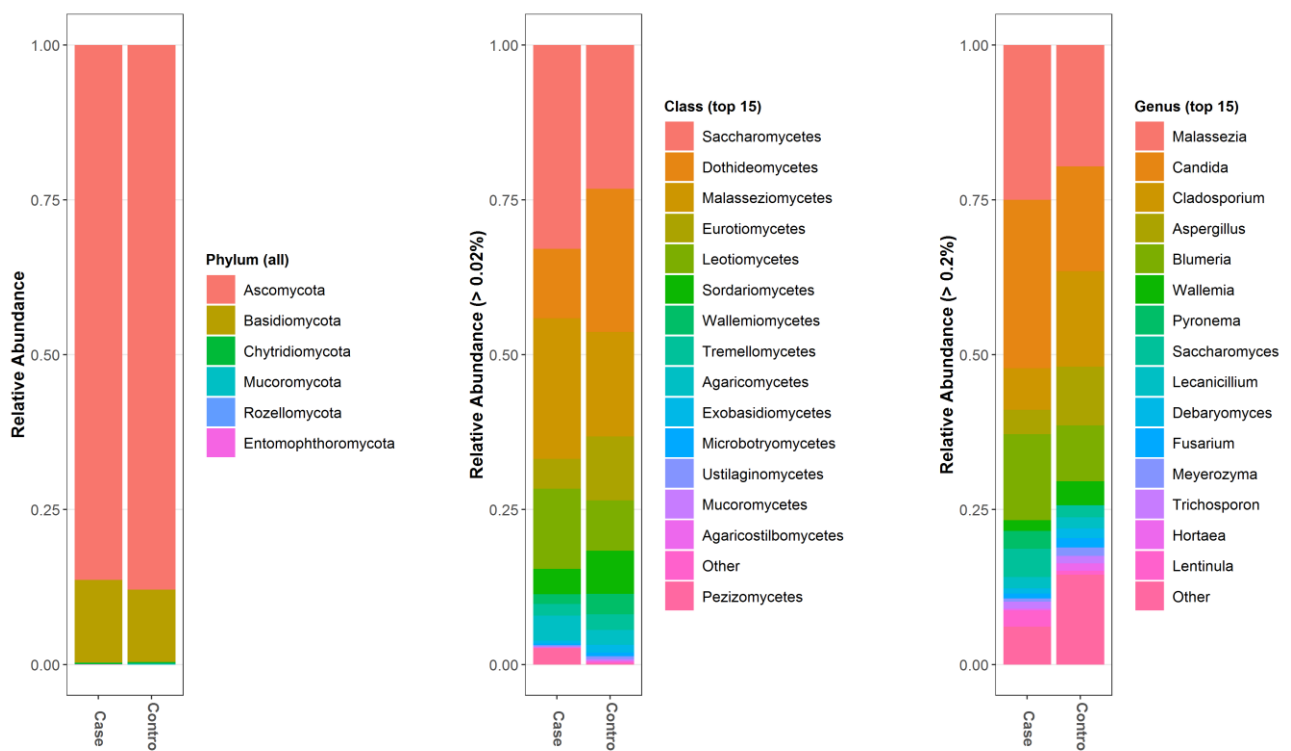


Figure 9. Relative abundances of predominant oral fungal taxa at the phylum, class and genus level in the nasopharyngeal carcinoma patients and healthy controls.

We noted significantly fewer observed ASVs, lower Simpson index and Shannon diversity in NPC cases compared to those of controls (Wilcoxon tests $P < 0.002$) (**Figure 10**). The results did not change after adjustment for multiple potential confounders, including age, sex, sequencing run, residence region, housing type, smoking, alcohol drinking, tea drinking, salted fish consumption, and history of ENT diseases using logistic regression models (data not shown). We also observed a significant difference in global fungal community patterns (beta diversity) between NPC cases and controls based on Bray-Curtis dissimilarity, using Adonis model adjusted for age, sex, and sequencing run (9,999 permutations, false discovery rate [FDR] corrected $P < 0.001$) (**Figure 11**).

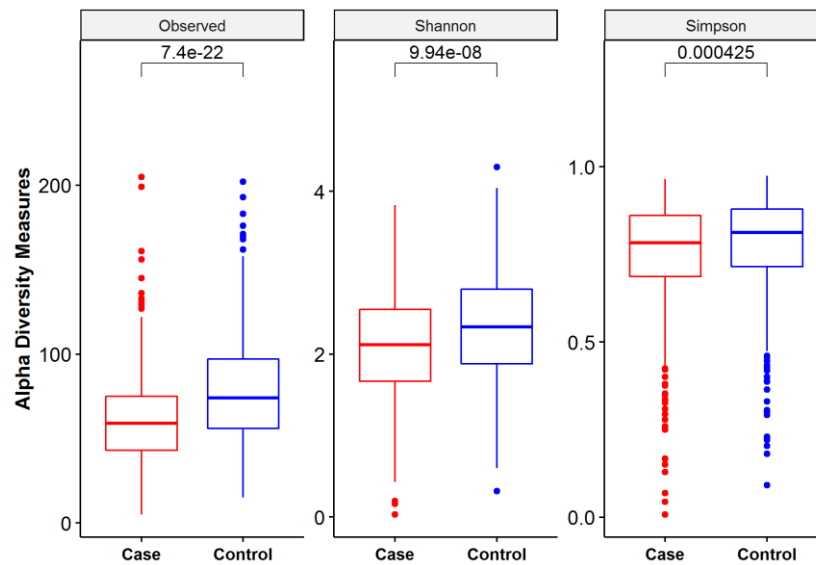


Figure 10. Box plots of alpha-diversity indices comparing healthy controls' and nasopharyngeal carcinoma patients' samples.

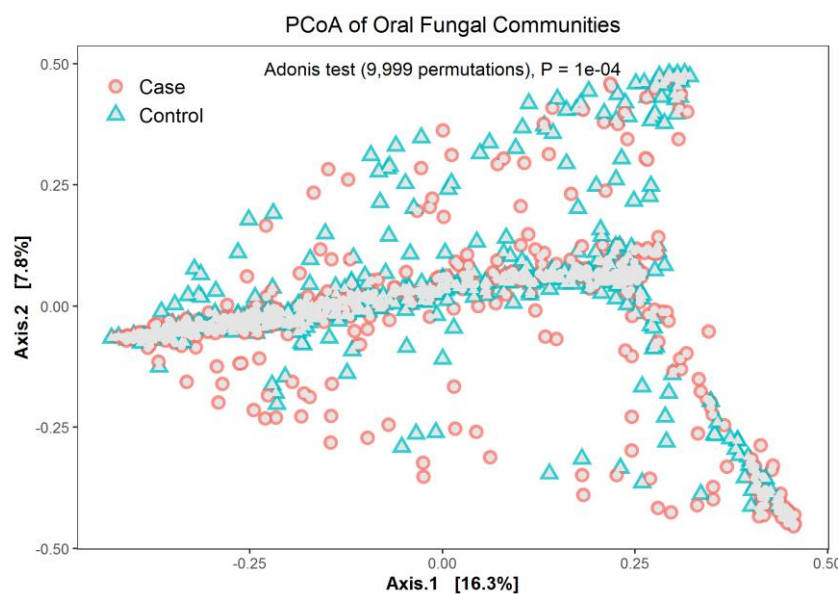


Figure 11. A Bray-Curtis distance-based principal coordinates analysis (PCoA) plot of oral fungal communities.

Based on a threshold on the logarithmic LDA score > 3.0 for discriminative features using LefSe analysis, we identified a set of organisms including 1 phylum, 16 classes, 28 orders, 23 families, 27 genera and 36 species that showed significantly differential abundance between NPC cases and controls (LDA score > 3.0 , FDR corrected $P < 0.05$). The most differentially abundant fungal organisms in NPC cases and controls at phylum, class, genus, and species levels are presented in **Figure 12**. For instance, the 5 most differentially enriched species in NPC cases were *Candida albicans*, *Saccharomyces cerevisiae*, *Pyronema domesticum*, *Lentinula edodes*, and *Candida tropicalis*. By contrast, *Cladosporium halotolerans*, *Cladosporium cladosporioides*, *Aspergillus penicillioides*, *Hortaea werneckii*, and *Debaryomyces prosopidis* were the top 5 differentially abundant species in controls.

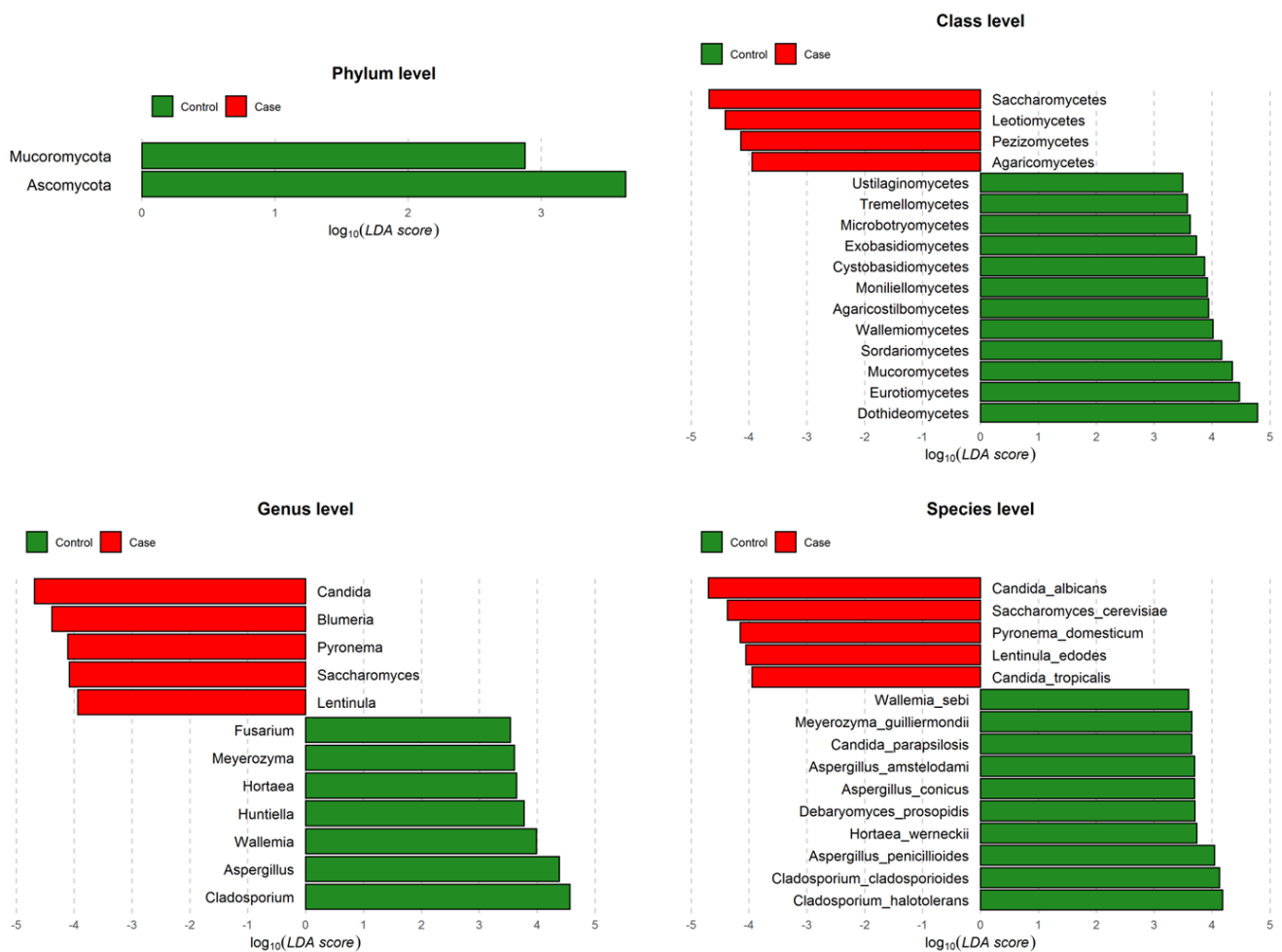


Figure 12. Differentially abundant oral fungal taxa in nasopharyngeal carcinoma patients and healthy controls.

Linear Discriminant Analysis (LDA) Effect Size (LEfSe) was used for pairwise comparison between groups. Only microbial organisms with LDA score > 3.5 are shown.

6 DISCUSSION

6.1 FINDINGS, INTERPRETATIONS AND IMPLICATIONS

On the basis of a rigorous, large population-base case-control study in a NPC endemic area, we demonstrated that living a house with poor residential conditions and occupational exposures to dusts, chemical vapors, exhausts/smokes, or acids/alkalis are associated with an increased risk of NPC in Study I and Study II. We found no associations between most of the environmental factors examined and EBV reactivation, a pivotal step in NPC carcinogenesis, in Study III. We noted a dysbiosis in oral mycobiome, characterized by a loss of fungal community diversity and increasing abundance of pathogenic fungal communities in NPC patients compared to healthy controls in study IV. The findings of this thesis work provide concrete evidence and insights for better understanding the etiology of NPC, and have public health implications for preventing or mitigating the burden of NPC in endemic areas.

6.1.1 Residential exposures and risk of NPC

By using the residential history data from the NPCGEE study in southern China, we found ever living in an inferior housing type (cottage or boat vs. building) was associated with an increased risk of NPC. We observed an almost 4-fold increased risk of NPC among individuals ever living in a boat; this finding is concordant with prior knowledge that Cantonese “boat people” in southern China have the highest incidence of NPC [6]. The high OR may be due to a much lower SES, and/or a special dietary pattern, e.g. higher consumption of salted fish in the “boat people”. Although SES strongly affects living conditions such as housing type and low SES is a risk factor for NPC [39,140], the association between inferior housing type and NPC is unlikely to be totally explained by the low SES because we included educational level and occupation to serve as proxies of SES in the multivariate regression models to control the confounding from SES. On the other hand, individuals living in inferior housing types may be more likely to be exposed to various pathogenic agents and hazardous substances, therefore increasing the risk of NPC. Moreover, a positive duration-response trend, which has not been reported previously, was seen for the associations with inferior housing types. Collectively, our findings provide robust evidence for linking inferior housing type with increased risk of NPC.

In this study we found a significant association between NPC risk and use of less clean fuels (wood, coal/charcoal, or kerosene) for cooking, exposure to cooking smokes and incense smokes in a duration-response fashion. Smokes/fumes from indoor combustion of solid fuels, cooking, and burning incense are causes of household air pollution (HAP). The oncogenic effect of HAP has been well investigated, and HAP has been linked to higher risks of lung cancer, cervical cancer, and a series of upper aero-digestive tract cancers [141]. HAP contributors have also been previously examined in NPC, but the results across studies are often conflicting, and the exposure-response relationship has been rarely reported in prior studies. Our results therefore indirectly support HAP is a risk factor for NPC through more accurate and concrete evidence.

To our knowledge, this is the first study to address the association between source of drinking water and NPC risk. Our results indicate that using untreated water sources including wells, rivers, and natural spring, pond, or stream waters increased 1.6 to 2.0-fold risk of NPC. Numerous hazardous agents e.g. metals, chemicals, and microbes contained in untreated waters [142] might contribute to the excess risk of NPC. Which constituents present in untreated waters and the mechanisms accounting for the excess risk warrant to be further studied.

In accordance with most prior evidence [52,54-56], our findings showed a positive association between lifelong poorer ventilation and NPC risk. The potential mechanism behind the association may be that ill-ventilated house can hardly release smokes, fumes or chemicals from household activities, such as combustion of solid fuels for cooking, leading to HAP hence increasing NPC risk. Our results plus previous reports indicate poorer house ventilation is a strong risk factor for NPC.

The association of residential exposures at early age with NPC risk has scarcely been reported previously. We observed a stronger association for exposures at an earlier age for most of the residential factors examined. This observation indicates that the nasopharyngeal epithelium may be more vulnerable to residential exposures at early ages and may partly explain the relatively young age of NPC occurrence in the high-incidence areas.

Taken together, Study I demonstrates that poor residential conditions and household air pollution are associated with an elevated risk of NPC. Measures that ameliorate poor residential conditions might reduce the disease burden in epidemic regions.

6.1.2 Occupational exposures and risk of NPC

The association between occupational exposures and the risk of NPC has been extensively evaluated in previous studies. The association, however, remains poorly understood and controversial findings have been reported due partly to various methodological limitations. In Study II, we investigated the associations of various occupational exposures with NPC risk using a large population-based case-control study in a high-incidence population. After adjusting for multiple confounders, we found individuals exposed to broad categories of occupational dusts, chemical vapors, exhausts/smokes, and acids/alkalis were at an elevated risk of NPC. Specifically, the associations were principally attributable to exposure to 14 subtypes of occupational agents within these broad categories. The associations of NPC with both broad categories and more specific subtypes of exposures were often stronger with longer duration of exposure. These observations may suggest a strong causal role of occupational exposures in NPC development.

Although the underlying mechanisms for the observed associations of occupational exposures with NPC remain largely speculative, potential mechanistic pathways might involve chronic physical irritation and inflammation caused by deposition of particles in the nasopharynx, carcinogenic properties of certain dusts and chemical substances, and co-exposure to heat [20].

Our findings provide strong evidence for supporting a role of occupational exposures in NPC etiology in an endemic area. Given occupational exposures are modifiable, improvement of working conditions in the workplace and wearing personal protective equipment are tangible preventive strategies to reduce NPC risk.

6.1.3 Environmental factors and EBV reactivation

Given the important role of EBV reactivation (i.e., the shift of EBV latent infection to lytic infection) in the carcinogenesis of NPC, we aim to investigate whether exposure to environmental factors is associated with EBV reactivation. Generally, we found a wide range of environmental factors, except for cigarette smoking, were not associated with EBV reactivation.

These predominantly null findings suggest that stable environmental factors are not likely to be primary determinants of EBV reactivation. Other non-environmental factors, including host and/or viral genetic variation, may have a larger role in EBV reactivation in this population. Alternatively, short-term environmental exposures not captured by our questionnaire in this study, such as temporal sources of endogenous or environmental stress, may influence EBV reactivation.

Our finding is in line with two previous hospital-based studies and a screening-based study in southern China [99,100,111], which showed a positive association between cigarette smoking and EBV seropositivity; and both our study and prior studies observed a positive exposure-response trend for the association. Besides, evidence from molecular studies demonstrates that cigarette smoke extract promotes EBV replication and enhances the expression of genes in lytic-phase [100]. Our results plus previous observations suggest cigarette smoking is involved in EBV reactivation. These findings may also suggest that cigarette smoking contributes to NPC oncogenesis not only by a direct carcinogenic effect of tobacco smoke, but also indirectly by induction of EBV reactivation.

By contrast, recent genomic analyses showed that host and viral genetic variations may affect EBV lytic reactivation [143-145]. For instance, in a large-scale whole-genome sequencing (WGS) study in southern China, Xu *et al.* identified two non-synonymous EBV variants within the *BALF2* gene, a core component of lytic viral DNA replication machinery, that were associated with a 6.1- to 8.7-fold increased risk of NPC [143].

In short, our findings suggest that stable environmental factors are unlikely to contribute to EBV reactivation although a probably association with cigarette smoking exists. Future studies into the mechanism of EBV reactivation may focus on host and viral genetic variations, or transient endogenous/exogenous stress.

6.1.4 Oral fungal profiling and risk of NPC

The potential role of oral fungal community in the development of NPC has not been examined before. To the best of our knowledge, this is the first population-based case-control

study to investigate the association between oral fungal microbiome and NPC risk, and we found NPC patients harbored a lower richness and reduced diverse oral fungal community compared to healthy controls, and the fungal community patterns significantly differed by disease status.

These findings suggest that the oral mycobiome is significantly associated with disease status (i.e., NPC patient or control). Similar observations were noted in patients with oral squamous cell carcinoma, oral tongue cancer, and overall head and neck squamous cell carcinoma, as previous studies reported a reduced alpha diversity in the oral mycobiome in cancer patients [120-123]. Our results are robust because the large sample size ensuring sufficient statistical power and the associations remained unchanged after adjustment for many known or potential confounders. Further, our results are unlikely to be treatment related, because the NPC cases were newly diagnosed ones and samples were collected prior to radiotherapy or chemotherapy treatment [146,147].

Furthermore, we identified a bunch of oral fungal organisms that are differentially present in NPC patients and controls, characterized by an increased abundance in opportunistic or pathogenic fungi and a decrease in symbiotic fungi in NPC patients. *Candida*, the most differentially abundant and enriched genus in NPC patients, has been proposed to be implicated in carcinogenesis through multiple pathways including production of mutagenic and carcinogenic substances, inducing dysplasia and colonizing existing premalignant lesions, inducing host immune responses to produce proteinases and pro-inflammatory mediators [148-150]. By contrast, *Cladosporium*, was the most differentially enriched genus in the healthy controls, in other words, the least differentially abundant genus in NPC patients. Studies have shown that many species of *Cladosporium* are capable to produce bioactive compounds such as antibiotics, p-Methylbenzoic acid, Ergosterol peroxide, and Calphostin C, which have been shown to possess broadly beneficial human health effects [151,152]. Therefore, our findings indicate that the differentially abundant oral fungal organisms identified in the analysis may have a role in NPC carcinogenesis.

In summary, our results suggest there is an oral fungal dysbiosis in NPC patients. Findings from this study expand the understanding of the etiology in NPC, and might pave a new way for NPC prevention and control.

6.2 METHODOLOGICAL CONSIDERATIONS

6.2.1 Study design

Theoretically, a cohort study design is ideal to identify risk factors for a disease because the exposure information is collected before disease occurrence. However, it may not be the case for relatively uncommon diseases, such as NPC. For example, in Sihui, a NPC high-risk area, only 125 NPC cases were identified during an average of 16.9 years of follow-up in a cohort of 18,986 individuals [103]. In a cohort study with 9,622 community-dwelling males conducted in Taiwan, a NPC intermediate-risk area, 33 incident NPC cases were diagnosed during a total of 185,587 person-years of follow-up [153]. In low-risk area, such as Sweden,

from approximately 5.3 million individuals who were alive and free of NPC on 1 January 1961 or born after 1961, and followed up until 31 December 2009, only 301 NPC cases were identified [154]. For areas without established healthcare registers, cohort design becomes logistically impractical. In addition, a cohort study design could be extremely costly, time-consuming, and inefficient in a context of rare disease. Also, information collected at baseline might be limited. It will not be able to test new hypotheses if relevant data were not collected before. Therefore, a case-control study design is an optimal alternative for studying risk factors of rare diseases.

Ideally, a case-control study should include all incident cases of the disease under study in a defined population during a specified time period; control subjects without the disease should be randomly selected directly from the source population that gives rise to the cases. However, most of prior case-control studies that addressing the association between various exposures and NPC are implemented with compromised methods, such as recruitment of prevalent cases, and use of hospital-based or neighborhood-based study design. By using these study designs, information on environmental exposures is not representative therefore introducing selection bias and information bias. Even worse, a population and person-time cannot be clearly defined in many case-control studies, eventually resulting in, if any, inconsistent findings across studies.

Considering the rarity of NPC, we conducted a strict, large population-based case-control study, namely the NPCGEE study, in southern China, where the incidence of NPC is highest in the world while established healthcare registers are not available in the study area. We clearly defined the study base (study area and study period), and recruited histopathologically confirmed, incident NPC cases, and controls were randomly selected from total population. In this thesis, all four studies were performed based on the NPCGEE study to address the research questions.

6.2.2 Sample size

Sample size is an important component of statistics. Sample size has significant influence on the internal and external validity of a study. An association generated from sample data always has some levels of uncertainty, which is determined by the potential variability of the data and the sample size. Small sample size increases uncertainty in estimated association due to lack of statistical power, therefore the association is likely to occur randomly or by chance, and eventually leading to not be able to extrapolate the statistical analysis results to the overall population. On the contrary, large sample size decreases uncertainty and gives more reliable results with greater precision and statistical power. In addition, large sample size can also provide greater power to detect small or modest differences (effect sizes). When an effect size is small, a larger sample size is required in order to detect the small difference; otherwise the effect will be masked by chance or randomness.

We initially aimed to recruit 2,600 cases and 2,600 controls in the NPCGEE study. The sample size was carefully calculated before implementation to ensure sufficient statistical power for detection of modest associations meanwhile to avoid wasting resource and time.

6.2.3 Selection bias

Selection bias is a systematic error derived from the fact that the selected subjects are not representative of the source population intended to be analyzed. Selection bias is one of the main concerns in observational studies, and it results in either overestimation or underestimation of association. Accordingly, efforts were made in the NPCGEE study to minimize selection bias. We employed the same inclusion criteria to cases and controls, ensuring that cases and controls came from the same study population. We also randomly selected controls from the total population in study area. Although we achieved relatively high participation rates among cases (84%) and controls (83%), we acknowledge that a probability of selection bias cannot be ruled out. For instance, in study I, we observed an unexpected finding of a reduced risk of NPC conferred by burning anti-mosquito coils. The participation rate was lower among urban controls than rural ones in our study, potentially causing a higher frequency of burning anti-mosquito coils among controls due to the poorer hygienic circumstances in rural areas. Therefore, the unexpected association might occur due to this selection bias. In study II, a portion of non-participating controls could not be contacted due to working outside of their hometown. Those controls usually worked in various manufacturing factories, therefore could be exposed to all sorts of occupational exposures. Therefore, in this case, selection bias leads to a low percentage of occupational exposures among controls, favoring overestimated ORs. In study III, selection bias may also occur because around 25% of controls without blood samples were excluded from the analysis. However, since the distributions of age and sex and educational level were similar among those with samples and those without, we argue that selection bias due to the unavailability of the samples, if any, should be minimal. In study IV, if the non-participating cases were more likely with a poor oral health, which is associated with reduced microbial diversity, this would also lead to an underestimated association; alternatively, if the non-participating cases were in better oral conditions, an overestimated association could occur.

6.2.4 Information bias (misclassification)

In this thesis, the cases were all histopathologically confirmed, therefore misclassification of disease status is less likely to occur. Misclassification of exposure is one of the primary concerns in the four studies, and may lead to either overestimated or underestimated association. To minimize information bias, extensive efforts were made in the study design and implementation stage. For instance, interviewers were trained with a manual that describes standard survey techniques to be implemented for all participants; logic checks were built into the structured, electronic questionnaire, and interviews were audiotaped for quality control; we periodically analyzed the collected data during recruitment phase to ensure that results were within expectation. We also tried to assign approximately equal numbers of cases and control subjects to each interviewer. Nonetheless, information bias is

inevitable for its retrospective nature in a case-control study. First, exposure information was retrospectively self-reported and may have been imprecise, particularly for distant past exposures. Second, if cases were more likely to remember or report their exposures than controls, recall bias may have occurred and result in an over-estimated association. Third, interviewers were not blinded to case-control status, despite interviewers were required to interview cases and controls in the same way, and therefore could have led to some extent of differential misclassification. In addition, in study III, because lack of standard definition of EBV reactivation, we classified EBV reactivation status based on a combination of two markers, i.e., VCA/IgA and EBNA1/IgA, the two most commonly used indicators of EBV reactivation in a screening setting; however, EBV reactivation status may change if markers such as EA/IgA, Zta/IgA, Rta/IgA, or EBV load are used. Finally, in study IV, unlike traditional observational studies, there is no straightforward analytical workflow for human microbiome studies yet [155], thus numerous biases could be introduced during the DNA-PCR-sequencing procedures, and the situation may be even more challenging for mycobiome studies [156-159].

6.2.5 Confounding

Confounding refers to a situation where the association of the exposure with the outcome is mixed with the effects of an additional factor (or set of factors) resulting in a distortion of the true association [160]. Confounders are defined as factors associated to both exposure and outcome, but not an intermediate step in the causal pathway between exposure and outcome. Confounding can be controlled at the phase of study design (matching, restriction, and randomization) or during data analysis (adjustment in regression model and stratification).

In the NPCGEE study, the controls were frequency-matched on age, sex, and geographic region of the cases to reduce the confounding of these three variables which are known risk factors for NPC. Furthermore, we employed multivariate regression models to adjust for matching variables and potential confounders based on prior knowledge and other known risk factors in the study population. In addition, stratified analyses were also performed to evaluate and minimize the influence of confounding. Nevertheless, completely ruling out residual and unmeasured confounding is virtually impossible in an observational study. For example, SES, commonly measured by income, education, or social class, is a strong confounder in the analysis of association between residential exposure and NPC, but information on income, a straightforward indicator of SES, was not available in this study. Therefore, the results may be influenced by a degree of residual confounding from SES, although educational level and occupation had been used as proxies for SES in the multivariate regression models.

6.2.6 Generalizability

Generalizability, also known as external validity, evaluates to which extent findings could be applied to other settings. The distribution of histopathological subtype of NPC is closely related to geographical distribution, with almost all undifferentiated carcinoma being

diagnosed in high-incidence areas and keratinizing squamous cell carcinoma predominating in low-incidence regions. It has been proposed that risk factors, as well as underlying oncogenic mechanisms vary among different histopathological types of NPC. Given all studies in this thesis were conducted in a high-incidence population, the findings should be able to generalize to most regions and countries with high- and intermediate incidence of NPC, including, southern China, Hong Kong, Taiwan regions and southeast Asian countries. However, the generalization of the results to low-risk population settings, such as most western countries, should be with cautions.

7 CONCLUSIONS

- Poor residential conditions and household air pollution are associated with an increased risk of NPC. Large-scale studies in other populations or longitudinal studies are warranted to further corroborate these findings. If confirmed, measures that ameliorate residential conditions might alleviate the burden of NPC in endemic areas.
- Occupational exposures to dusts, chemical vapors, exhausts/smokes, or acids/alkalis are associated with an excess risk of NPC. Large-scale studies in similar settings or longitudinal studies with more precise assessment of occupational exposures are warranted to further corroborate these results. If these results are causal, improvement of working conditions in the workplace and wearing personal protective equipment are tangible preventive strategies to reduce NPC risk in endemic areas.
- Exposure to stable environmental factors, except for cigarette smoking, is not likely to be primary determinant of EBV reactivation. Future studies into the mechanism of EBV reactivation may focus on host and/or viral genetic variations, or transient endogenous/exogenous stress.
- A dysbiosis in oral mycobiome, characterized by reduced fungal community richness and diversity, as well as an increased abundance in pathogenic fungi and a decrease in symbiotic fungi, is associated with an increased risk of NPC. This finding, if confirmed by other studies, might pave a new way for NPC prevention and control.

8 FUNDING

Studies in this thesis have received funding from:

- National Cancer Institute, United States
- Swedish Research Council, Sweden
- Swedish Foundation for International Cooperation in Research and Higher Education, Sweden
- Karolinska Institutet Distinguished Professor Award, Sweden
- Program for New Century Excellent Talents in University, China
- National Basic Research Program of China, China
- Guangxi Natural Science Foundation, China
- High-level Talents Research Start-up Project of Fujian Medical University, China
- China Scholarship Council, China

9 ACKNOWLEDGEMENTS

First of all, I would like to thank all study participants who contributed data and samples to the NPCGEE study; and thank all the fieldwork coordinators, interviewers, clinical stuffs who helped us with data and sample collection. Their contributions are the prerequisites of conducting all these studies. I would particularly like to express my gratitude to those who gave me tremendous support and help in many ways during my PhD journey.

Weimin Ye, my main supervisor, thank you for providing me the opportunity to start a doctoral training at Karolinska Institutet five years ago and for always guiding and supporting me throughout the years. Thank you for sharing your broad and in-depth knowledge, and outstanding experience in epidemiology and biostatistics. I really appreciate for your encouragement, enlightenment and trust in my PhD journey. Your critical thinking, rigorous attitude, and diligent spirit toward science will keep guiding me to become an independent researcher.

My co-supervisor, **Zhiwei Liu**, I appreciate the opportunity to work with on the NPC and EBV serology studies. Thank you for sharing your experience and point of view on the studies. I appreciate the fruitful and insightful discussions with you. Thank you also for your generous and timely help.

My co-supervisor, **Rob Knight**, thank you for sharing your profound knowledge in microbiome research, and for your always enthusiastic encouragement and support.

Zhe Zhang, my friend, teacher, and tutor, thank you so much for leading me to the wonderful World of Scientific Research. I'm sincerely grateful for your unconditional support, generous help, priceless advices, and always inspiring ideas both on researches and personal life. Thank you also for taking care of me during my stay in China.

Thank you, **Fang Fang** and **Qin Cheng**, for always being so supportive and helpful for all the things throughout the years.

I would like to thank all of collaborators in my work. I feel extremely honor to join the NPCGEE study and to work with you. Special thanks to the amazing senior collaborators: **Hans-Olov Adami, Ellen T. Chang, Yi Zeng, Yixin Zeng, Guangwu Huang**, and **Ingemar Ernberg**, for all the effort and input from conceiving to implementing this extraordinary research. I'm really grateful to **Ellen T. Chang, Hans-Olov Adami**, and **Hongwei Zhao**, for always thoughtful comments and revisions on the manuscripts, and for the insightful discussions with you. I also would like to thank **Qing Liu, Qihong Huang, Shanghang Xie, Sumei Cao, Weihua Jia, Yonglin Cai, Guomin Chen, Yuming Zheng, Jian Liao, Yancheng Li, Longde Lin, Ruimei Feng, Weihong Chen, Dongming Wang**, and **Wanxin Li**, for your contributions, valuable efforts and advices on the studies.

Many thanks to present and previous members of Weimin's group: **Joar Franzen, Amir Sohrabi, Weiwei Bian, Jingru Yu, Yun Du, Ji Zhang, Fatemeh Sadeghi, Tingting Huang**.

Justine Debelius, Donal Barrett, Isabella Ekheden, Nelson Ndegwa Gichora, Nikolaos Tertipis, Ulrika Zagai, Marie Lindén, Amelie Plymoth, Anna Berglund, Yawen Sun, Xiaorong Yang, Fei He, Song Lin, Ping Yuan, and Yvonne Nartey, for being always helpful and supportive. **Amir, Joar, and Donal**, thank you for teaching and sharing your experimental knowledge and experience and skills in lab work, I enjoyed discussing and working with you very much, and I learned a lot from you. **Joar and Donal**, thank you for inviting me and Fei to your homes, we had so much fun and joyful time with you and your families.

I want to thank all the past and present **MEBers**, including but not limited to: **Shihua Sun, Wenjiang Deng, Qingyang Xiao, Yiqiang Zhan, Ruqing Chen, Yunzhang Wang, Yinxì Wang, Qing Shen, Ruyue Zhang, Jie Song, Haomin Yang, Xu Chen, Jiayao Lei, Wei He, Jiangwei Sun, Lin Li, Tong Gong, Donghao Lu, Huan Song, Jianwei Zhu, Xinhe Mao, Erwei Zeng, Zheng Ning, Xia Li, Qian Yang, Chen Wang, Shengxin Liu, Nanbo Zhu, Anders Forss, Marica Leone, and Miriam Martini**. Thank you for being excellent friends, and colleagues, you are always so helpful and supportive.

I would like to thank all of my Chinese colleagues in **Guangwu Huang** and **Zhe Zhang**, and in **Xiaoyun Zeng**'s research groups from Guangxi Medical University. Special thanks to **Shixing Zheng, Xiaoying Zhou, Xue Xiao, Yingxi Mo, Guofei Feng, Yunliang Lu, Di Zhang**, and **Qiuli Liang**, for all the supports to my studies.

To my family, my **parents**, my **brothers**, my **uncle**, thank you for all the endless support.

Xaingfei, my beloved wife, thank you for everything. You are the pillar and strength source of my life.

10 REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209-249.
2. Bray F, Colombet M, Mery L, et al. Cancer Incidence in Five Continents, Vol. XI; 2017. Lyon: International Agency for Research on Cancer. <http://ci5.iarc.fr>. Accessed October 20, 2020.
3. Chang ET, Ye W, Zeng Y-X, et al. The Evolving Epidemiology of Nasopharyngeal Carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2021;30(6):1035-1047.
4. Chen Y-P, Chan ATC, Le Q-T, et al. Nasopharyngeal carcinoma. *The Lancet.* 2019;394(10192):64-80.
5. Cao SM, Simons MJ, Qian CN. The prevalence and prevention of nasopharyngeal carcinoma in China. *Chin J Cancer.* 2011;30(2):114-119.
6. Tao Q, Chan AT. Nasopharyngeal carcinoma: molecular pathogenesis and therapeutic developments. *Expert Rev Mol Med.* 2007;9(12):1-24.
7. Jia WH, Huang QH, Liao J, et al. Trends in incidence and mortality of nasopharyngeal carcinoma over a 20-25 year period (1978/1983-2002) in Sihui and Cangwu counties in southern China. *BMC Cancer.* 2006;6:178.
8. Tang L-L, Chen W-Q, Xue W-Q, et al. Global trends in incidence and mortality of nasopharyngeal carcinoma. *Cancer Lett.* 2016;374(1):22-30.
9. Middeldorp JM. Epstein-Barr Virus-Specific Humoral Immune Responses in Health and Disease. In: Münz C, editor. Epstein Barr Virus Volume 2. One Herpes Virus: Many Diseases. Vol. 391. Heidelberg, New York, Dordrecht, London: Springer International Publishing Switzerland; 2015. p. 289-323.
10. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394-424.
11. Chua MLK, Wee JTS, Hui EP, et al. Nasopharyngeal carcinoma. *Lancet.* 2016;387(10022):1012-1024.
12. Nicholls JM. Nasopharyngeal Carcinoma: Classification and Histologic Appearances. *Adv Anat Pathol.* 1997;4:71-84.
13. Ren ZF, Liu WS, Qin HD, et al. Effect of family history of cancers and environmental factors on risk of nasopharyngeal carcinoma in Guangdong, China. *Cancer Epidemiol.* 2010;34(4):419-424.
14. Tang LQ, Li CF, Li J, et al. Establishment and Validation of Prognostic Nomograms for Endemic Nasopharyngeal Carcinoma. *J Natl Cancer Inst.* 2016;108(1).
15. Walker HS, McLaughlin JK, Weiner JA, et al. Occupational risk factors for nasopharyngeal cancer in Sweden. *Br J Ind Med.* 1990;47(3):213-214.

16. Vaughan TL, Shapiro JA, Burt RD, et al. Nasopharyngeal cancer in a low-risk population: defining risk factors by histological type. *Cancer Epidemiol Biomarkers Prev.* 1996;5(8):587-593.
17. Lo KW, To KF, Huang DP. Focus on nasopharyngeal carcinoma. *Cancer Cell.* 2004;5(5):423-428.
18. Chua DT, Sham JS, Kwong DL, et al. Treatment outcome after radiotherapy alone for patients with Stage I-II nasopharyngeal carcinoma. *Cancer.* 2003;98(1):74-80.
19. Lee AWM, Tung SY, Ng WT, et al. A multicenter, phase 3, randomized trial of concurrent chemoradiotherapy plus adjuvant chemotherapy versus radiotherapy alone in patients with regionally advanced nasopharyngeal carcinoma: 10-year outcomes for efficacy and toxicity. *Cancer.* 2017;123(21):4147-4157.
20. Chang ET, Adami HO. The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2006;15(10):1765-1777.
21. Hildesheim A, Wang C-P. Genetic predisposition factors and nasopharyngeal carcinoma risk: A review of epidemiological association studies, 2000–2011. *Semin Cancer Biol.* 2012;22(2):107-116.
22. Xie SH, Yu IT, Tse LA, et al. Tobacco smoking, family history, and the risk of nasopharyngeal carcinoma: a case-referent study in Hong Kong Chinese. *Cancer Causes Control.* 2015;26(6):913-921.
23. Liu Z, Chang ET, Liu Q, et al. Quantification of familial risk of nasopharyngeal carcinoma in a high-incidence area. *Cancer.* 2017;123(14):2716-2725.
24. Albeck H, Bentzen J, Ockelmann HH, et al. Familial clusters of nasopharyngeal carcinoma and salivary gland carcinomas in Greenland natives. *Cancer.* 1993;72(1):196-200.
25. Ko JY, Sheen TS, Hsu MM, et al. Familial clustering of nasopharyngeal carcinoma. *Otolaryngol Head Neck Surg.* 1998;118(5):736-737.
26. Ung A, Chen CJ, Levine PH, et al. Familial and sporadic cases of nasopharyngeal carcinoma in Taiwan. *Anticancer Res.* 1999;19(1B):661-665.
27. Warnakulasuriya KA, Johnson NW, Linklater KM, et al. Cancer of mouth, pharynx and nasopharynx in Asian and Chinese immigrants resident in Thames regions. *Oral Oncol.* 1999;35(5):471-475.
28. Yu WM, Hussain SS. Incidence of nasopharyngeal carcinoma in Chinese immigrants, compared with Chinese in China and South East Asia: review. *J Laryngol Otol.* 2009;123(10):1067-1074.
29. Jeannel D, Ghnassia M, Hubert A, et al. Increased risk of nasopharyngeal carcinoma among males of French origin born in Maghreb (north Africa). *Int J Cancer.* 1993;54(4):536-539.
30. Pathmanathan R, Prasad U, Sadler R, et al. Clonal proliferations of cells infected with Epstein-Barr virus in preinvasive lesions related to nasopharyngeal carcinoma. *N Engl J Med.* 1995;333(11):693-698.
31. Wu HC, Lin YJ, Lee JJ, et al. Functional analysis of EBV in nasopharyngeal carcinoma cells. *Lab Invest.* 2003;83(6):797-812.

32. Hildesheim A, Levine PH. Etiology of nasopharyngeal carcinoma: a review. *Epidemiol Rev.* 1993;15(2):466-485.
33. Jia WH, Qin HD. Non-viral environmental risk factors for nasopharyngeal carcinoma: a systematic review. *Semin Cancer Biol.* 2012;22(2):117-126.
34. Lau HY, Leung CM, Chan YH, et al. Secular trends of salted fish consumption and nasopharyngeal carcinoma: a multi-jurisdiction ecological study in 8 regions from 3 continents. *BMC Cancer.* 2013;13:298.
35. Barrett D, Ploner A, Chang ET, et al. Past and Recent Salted Fish and Preserved Food Intakes Are Weakly Associated with Nasopharyngeal Carcinoma Risk in Adults in Southern China. *J Nutr.* 2019.
36. Zhu X-X, Yang X-J, Chao Y-L, et al. The Potential Effect of Oral Microbiota in the Prediction of Mucositis During Radiotherapy for Nasopharyngeal Carcinoma. *EBioMedicine.* 2017;18:23-31.
37. Debelius JW, Huang T, Cai Y, et al. Subspecies Niche Specialization in the Oral Microbiome Is Associated with Nasopharyngeal Carcinoma Risk. *mSystems.* 2020;5(4).
38. Stephen D. Sturton HLW, Olive G. Sturton. Etiology of cancer of the nasopharynx. *Cancer.* 1966;19.
39. Zheng YM, Tuppin P, Hubert A, et al. Environmental and dietary risk factors for nasopharyngeal carcinoma: a case-control study in Zangwu County, Guangxi, China. *Br J Cancer.* 1994;69(3):508-514.
40. Lakhanpal M, Singh LC, Rahman T, et al. Contribution of susceptibility locus at HLA class I region and environmental factors to occurrence of nasopharyngeal cancer in Northeast India. *Tumour Biol.* 2015;36(4):3061-3073.
41. Shanmugaratnam K, Tye CY, Goh EH, et al. Etiological factors in nasopharyngeal carcinoma: a hospital-based, retrospective, case-control, questionnaire study. *IARC Sci Publ.* 1978;(20):199-212.
42. Yu MC, Ho JH, Lai SH, et al. Cantonese-style salted fish as a cause of nasopharyngeal carcinoma: report of a case-control study in Hong Kong. *Cancer Res.* 1986;46(2):956-961.
43. Yu MC, Garabrant DH, Huang TB, et al. Occupational and other non-dietary risk factors for nasopharyngeal carcinoma in Guangzhou, China. *Int J Cancer.* 1990;45(6):1033-1039.
44. Yu MC, Mo CC, Chong WX, et al. Preserved foods and nasopharyngeal carcinoma: a case-control study in Guangxi, China. *Cancer Res.* 1988;48(7):1954-1959.
45. Feng BJ, Khyatti M, Ben-Ayoub W, et al. Cannabis, tobacco and domestic fumes intake are associated with nasopharyngeal carcinoma in North Africa. *Br J Cancer.* 2009;101(7):1207-1212.
46. Guo X, Johnson RC, Deng H, et al. Evaluation of nonviral risk factors for nasopharyngeal carcinoma in a high-risk population of Southern China. *Int J Cancer.* 2009;124(12):2942-2947.
47. He YQ, Xue WQ, Shen GP, et al. Household inhalants exposure and nasopharyngeal carcinoma risk: a large-scale case-control study in Guangdong, China. *BMC Cancer.* 2015;15(1):1022.

48. Armstrong RW, Kutty MK, Armstrong MJ. Self-specific environments associated with nasopharyngeal carcinoma in Selangor, Malaysia. *Soc Sci Med.* 1978;12(3D-4D):149-156.
49. Geser A, Charnay N, Day NE, et al. Environmental factors in the etiology of nasopharyngeal carcinoma: report on a case-control study in Hong Kong. *IARC Sci Publ.* 1978;(20):213-229.
50. Xie SH, Yu IT, Tse LA, et al. Domestic incense burning and nasopharyngeal carcinoma: a case-control study in Hong Kong Chinese. *Environ Mol Mutagen.* 2014;55(9):751-756.
51. Friberg JT, Yuan JM, Wang R, et al. Incense use and respiratory tract carcinomas: a prospective cohort study. *Cancer.* 2008;113(7):1676-1684.
52. Lin TM, Chen KP, Lin CC, et al. Retrospective study on nasopharyngeal carcinoma. *J Natl Cancer Inst.* 1973;51(5):1403-1408.
53. West S, Hildesheim A, Dosemeci M. Non-viral risk factors for nasopharyngeal carcinoma in the Philippines: results from a case-control study. *Int J Cancer.* 1993;55(5):722-727.
54. Jeannel D, Hubert A, de Vathaire F, et al. Diet, living conditions and nasopharyngeal carcinoma in Tunisia: a case-control study. *Int J Cancer.* 1990;46(3):421-425.
55. Chen DL, Huang TB. A case-control study of risk factors of nasopharyngeal carcinoma. *Cancer Lett.* 1997;117(1):17-22.
56. Lourembam DS, Singh AR, Sharma TD, et al. Evaluation of Risk Factors for Nasopharyngeal Carcinoma in a High-risk Area of India, the Northeastern Region. *Asian Pac J Cancer Prev.* 2015;16(12):4927-4935.
57. Henderson BE, Louie E, SooHoo Jing J, et al. Risk factors associated with nasopharyngeal carcinoma. *N Engl J Med.* 1976;295(20):1101-1106.
58. Vaughan TL, Strader C, Davis S, et al. Formaldehyde and cancers of the pharynx, sinus and nasal cavity: I. Occupational exposures. *Int J Cancer.* 1986;38(5):677-683.
59. Vaughan TL. Occupation and Squamous Cell Cancers of the Pharynx and Sinonasal Cavity. *Am J Ind Med.* 1989;16(5):493-510.
60. Vaughan TL, Davis S. Wood dust exposure and squamous cell cancers of the upper respiratory tract. *Am J Epidemiol.* 1991;133(6):560-564.
61. Mirabelli MC, Hoppin JA, Tolbert PE, et al. Occupational exposure to chlorophenol and the risk of nasal and nasopharyngeal cancers among U.S. men aged 30 to 60. *Am J Ind Med.* 2000;37(5):532-541.
62. Vaughan TL, Stewart PA, Teschke K, et al. Occupational exposure to formaldehyde and wood dust and nasopharyngeal carcinoma. *Occup Environ Med.* 2000;57(6):376-384.
63. Zhu K, Levine RS, Brann EA, et al. Case-control study evaluating the homogeneity and heterogeneity of risk factors between sinonasal and nasopharyngeal cancers. *Int J Cancer.* 2002;99(1):119-123.
64. Freeman BLE, Blair A, Lubin JH, et al. Mortality from solid tumors among workers in formaldehyde industries: an update of the NCI cohort. *Am J Ind Med.* 2013;56(9):1015-1026.

65. Mohner M, Liu Y, Marsh GM. New insights into the mortality risk from nasopharyngeal cancer in the national cancer institute formaldehyde worker cohort study. *J Occup Med Toxicol*. 2019;14:4.
66. Hardell L, Johansson B, Axelson O. Epidemiological study of nasal and nasopharyngeal cancer and their relation to phenoxy acid or chlorophenol exposure. *Am J Ind Med*. 1982;3(3):247-257.
67. Olsen JH, Asnaes S. Formaldehyde and the risk of squamous cell carcinoma of the sinonasal cavities. *Br J Ind Med*. 1986;43(11):769-774.
68. Hansen J, Olsen JH. Formaldehyde and Cancer Morbidity among Male Employees in Denmark. *Cancer Causes Control*. 1995;6(4):354-360.
69. Siew SS, Kauppinen T, Kyyronen P, et al. Occupational exposure to wood dust and formaldehyde and risk of nasal, nasopharyngeal, and lung cancer among Finnish men. *Cancer Manag Res*. 2012;4:223-232.
70. Siew SS, Martinsen JI, Kjaerheim K, et al. Occupational exposure to wood dust and risk of nasal and nasopharyngeal cancer: A case-control study among men in four nordic countries-With an emphasis on nasal adenocarcinoma. *Int J Cancer*. 2017;141(12):2430-2436.
71. Demers PA, Boffetta P, Kogevinas M, et al. Pooled reanalysis of cancer mortality among five cohorts of workers in wood-related industries. *Scand J Work Environ Health*. 1995;21(3):179-190.
72. Coggon D, Ntani G, Harris EC, et al. Upper airway cancer, myeloid leukemia, and other cancers in a cohort of British chemical workers exposed to formaldehyde. *Am J Epidemiol*. 2014;179(11):1301-1311.
73. Coggon D, Harris EC, Poole J, et al. Extended follow-up of a cohort of british chemical workers exposed to formaldehyde. *J Natl Cancer Inst*. 2003;95(21):1608-1615.
74. Reif J, Pearce N, Kawachi I, et al. Soft-tissue sarcoma, non-Hodgkin's lymphoma and other cancers in New Zealand forestry workers. *Int J Cancer*. 1989;43(1):49-54.
75. Ning JP, Yu MC, Wang QS, et al. Consumption of salted fish and other risk factors for nasopharyngeal carcinoma (NPC) in Tianjin, a low-risk region for NPC in the People's Republic of China. *J Natl Cancer Inst*. 1990;82(4):291-296.
76. Zheng W, McLaughlin JK, Gao YT, et al. Occupational risks for nasopharyngeal cancer in Shanghai. *J Occup Med*. 1992;34(10):1004-1007.
77. Li W, Ray RM, Gao DL, et al. Occupational risk factors for nasopharyngeal cancer among female textile workers in Shanghai, China. *Occup Environ Med*. 2006;63(1):39-44.
78. Armstrong RW, Armstrong MJ, Yu MC, et al. Salted Fish and Inhalants as Risk-Factors for Nasopharyngeal Carcinoma in Malaysian Chinese. *Cancer Res*. 1983;43(6):2967-2970.
79. Sriamporn S, Vatanasapt V, Pisani P, et al. Environmental risk factors for nasopharyngeal carcinoma: a case-control study in northeastern Thailand. *Cancer Epidemiol Biomarkers Prev*. 1992;1(5):345-348.

80. Armstrong RW, Imrey PB, Lye MS, et al. Nasopharyngeal carcinoma in Malaysian Chinese: occupational exposures to particles, formaldehyde and heat. *Int J Epidemiol*. 2000;29(6):991-998.
81. Ekburanawat W, Ekpanyaskul C, Brennan P, et al. Evaluation of non-viral risk factors for nasopharyngeal carcinoma in Thailand: results from a case-control study. *Asian Pac J Cancer Prev*. 2010;11(4):929-932.
82. Turkoz FP, Celenkoglu G, Dogu GG, et al. Risk factors of nasopharyngeal carcinoma in Turkey-an epidemiological survey of the Anatolian Society of Medical Oncology. *Asian Pac J Cancer Prev*. 2011;12(11):3017-3021.
83. Ekpanyaskul C, Sangrajang S, Ekburanawat W, et al. Semi-quantitative exposure assessment of occupational exposure to wood dust and nasopharyngeal cancer risk. *Asian Pac J Cancer Prev*. 2015;16(10):4339-4345.
84. Ho CK, Lo WC, Huang PH, et al. Suspected nasopharyngeal carcinoma in three workers with long-term exposure to sulphuric acid vapour. *Occup Environ Med*. 1999;56(6):426-428.
85. Hildesheim A, Dosemeci M, Chan CC, et al. Occupational exposure to wood, formaldehyde, and solvents and risk of nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev*. 2001;10(11):1145-1153.
86. Yang XR, Diehl S, Pfeiffer R, et al. Evaluation of risk factors for nasopharyngeal carcinoma in high-risk nasopharyngeal carcinoma families in Taiwan. *Cancer Epidemiol Biomarkers Prev*. 2005;14(4):900-905.
87. Zheng X, Yan L, Nilsson B, et al. Epstein-Barr virus infection, salted fish and nasopharyngeal carcinoma. A case-control study in southern China. *Acta Oncol*. 1994;33(8):867-872.
88. Xie SH, Yu IT, Tse LA, et al. Occupational risk factors for nasopharyngeal carcinoma in Hong Kong Chinese: a case-referent study. *Int Arch Occup Environ Health*. 2017;90(5):443-449.
89. Marsh GM, Morfeld P, Zimmerman SD, et al. An updated re-analysis of the mortality risk from nasopharyngeal cancer in the National Cancer Institute formaldehyde worker cohort study. *J Occup Med Toxicol*. 2016;11:8.
90. Collins JJ, Acquavella JF, Esmen NA. An updated meta-analysis of formaldehyde exposure and upper respiratory tract cancers. *J Occup Environ Med*. 1997;39(7):639-651.
91. Kwon SC, Kim I, Song J, et al. Does formaldehyde have a causal association with nasopharyngeal cancer and leukaemia? *Ann Occup Environ Med*. 2018;30:5.
92. Beigzadeh Z, Pourhassan B, Kalantary S, et al. Occupational exposure to wood dust and risk of nasopharyngeal cancer: A systematic review and meta-analysis. *Environ Res*. 2019;171:170-176.
93. E M, Yin J, Jin W, et al. Wood dust exposure and risks of nasopharyngeal carcinoma: a meta-analysis. *Eur J Public Health*. 2020.
94. Coglian VJ, Grosse Y, Baan RA, et al. Meeting report: summary of IARC monographs on formaldehyde, 2-butoxyethanol, and 1-tert-butoxy-2-propanol. *Environ Health Perspect*. 2005;113(9):1205-1208.

95. Hsu JL, Glaser SL. Epstein-barr virus-associated malignancies: epidemiologic patterns and etiologic implications. *Crit Rev Oncol Hematol*. 2000;34(1):27-53.
96. Young LS, Yap LF, Murray PG. Epstein-Barr virus: more than 50 years old and still providing surprises. *Nat Rev Cancer*. 2016;16(12):789-802.
97. Laichalk LL, Hochberg D, Babcock GJ, et al. The dispersal of mucosal memory B cells: evidence from persistent EBV infection. *Immunity*. 2002;16(5):745-754.
98. Raab-Traub N. Epstein-Barr virus in the pathogenesis of NPC. *Semin Cancer Biol*. 2002;12(6):431-441.
99. He YQ, Xue WQ, Xu FH, et al. The Relationship Between Environmental Factors and the Profile of Epstein-Barr Virus Antibodies in the Lytic and Latent Infection Periods in Healthy Populations from Endemic and Non-Endemic Nasopharyngeal Carcinoma Areas in China. *EBioMedicine*. 2018;30:184-191.
100. Xu FH, Xiong D, Xu YF, et al. An epidemiological and molecular study of the relationship between smoking, risk of nasopharyngeal carcinoma, and Epstein-Barr virus activation. *J Natl Cancer Inst*. 2012;104(18):1396-1410.
101. Ji MF, Wang DK, Yu YL, et al. Sustained elevation of Epstein-Barr virus antibody levels preceding clinical onset of nasopharyngeal carcinoma. *Br J Cancer*. 2007;96(4):623-630.
102. Chien YC, Chen JY, Liu MY, et al. Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men. *N Engl J Med*. 2001;345(26):1877-1882.
103. Cao SM, Liu Z, Jia WH, et al. Fluctuations of epstein-barr virus serological antibodies and risk for nasopharyngeal carcinoma: a prospective screening study with a 20-year follow-up. *PLoS One*. 2011;6(4):e19100.
104. Zeng Y, Zhang LG, Li HY, et al. Serological mass survey for early detection of nasopharyngeal carcinoma in Wuzhou City, China. *Int J Cancer*. 1982;29(2):139-141.
105. Yu X, Ji M, Cheng W, et al. Assessment of the Long-term Diagnostic Performance of a New Serological Screening Scheme in Large-scale Nasopharyngeal Carcinoma Screening. *J Cancer*. 2018;9(12):2093-2097.
106. Liu Z, Ji MF, Huang QH, et al. Two Epstein-Barr virus-related serologic antibody tests in nasopharyngeal carcinoma screening: results from the initial phase of a cluster randomized controlled trial in Southern China. *Am J Epidemiol*. 2013;177(3):242-250.
107. Liu Y, Huang Q, Liu W, et al. Establishment of VCA and EBNA1 IgA-based combination by enzyme-linked immunosorbent assay as preferred screening method for nasopharyngeal carcinoma: a two-stage design with a preliminary performance study and a mass screening in southern China. *Int J Cancer*. 2012;131(2):406-416.
108. Rosemarie Q, Sugden B. Epstein–Barr Virus: How Its Lytic Phase Contributes to Oncogenesis. *Microorganisms*. 2020;8(11):1824.
109. Fang CY, Lee CH, Wu CC, et al. Recurrent chemical reactivations of EBV promotes genome instability and enhances tumor progression of nasopharyngeal carcinoma cells. *Int J Cancer*. 2009;124(9):2016-2025.
110. Zeng Y, Zhong JM, Ye SQ, et al. Screening of Epstein-Barr virus early antigen expression inducers from Chinese medicinal herbs and plants. *Biomed Environ Sci*. 1994;7(1):50-55.

111. Hu T, Lin CY, Xie SH, et al. Smoking can increase nasopharyngeal carcinoma risk by repeatedly reactivating Epstein-Barr Virus: An analysis of a prospective study in southern China. *Cancer Med.* 2019;8(5):2561-2571.
112. Mai Z-M, Lin J-H, Ngan RK-C, et al. Solar Ultraviolet Radiation and Vitamin D Deficiency on Epstein-Barr Virus Reactivation: Observational and Genetic Evidence From a Nasopharyngeal Carcinoma-Endemic Population. *Open Forum Infect Dis.* 2020;7(10).
113. Baquero F, Nombela C. The microbiome as a human organ. *Clin Microbiol Infect.* 2012;18:2-4.
114. Tiew PY, Mac Aogain M, Ali N, et al. The Mycobiome in Health and Disease: Emerging Concepts, Methodologies and Challenges. *Mycopathologia.* 2020;185(2):207-231.
115. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. *Nature.* 2012;486(7402):207-214.
116. Human Oral Microbiome Database (HOMD). 2020. Available online at: <http://www.homd.org/>.
117. Ghannoum MA, Jurevic RJ, Mukherjee PK, et al. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog.* 2010;6(1):e1000713.
118. Kilian M, Chapple IL, Hannig M, et al. The oral microbiome - an update for oral healthcare professionals. *Br Dent J.* 2016;221(10):657-666.
119. Bandara HMHN, Panduwawala CP, Samaranayake LP. Biodiversity of the human oral mycobiome in health and disease. *Oral Dis.* 2019;25(2):363-371.
120. Perera M, Al-Hebshi NN, Perera I, et al. A dysbiotic mycobiome dominated by *Candida albicans* is identified within oral squamous-cell carcinomas. *J Oral Microbiol.* 2017;9(1):1385369.
121. Mukherjee PK, Wang H, Retuerto M, et al. Bacteriome and mycobiome associations in oral tongue cancer. *Oncotarget.* 2017;8(57):97273-97289.
122. Vesty A, Gear K, Biswas K, et al. Microbial and inflammatory-based salivary biomarkers of head and neck squamous cell carcinoma. *Clin Exp Dent Res.* 2018;4(6):255-262.
123. Shay E, Sangwan N, Padmanabhan R, et al. Bacteriome and mycobiome and bacteriome-mycobiome interactions in head and neck squamous cell carcinoma. *Oncotarget.* 2020;11(25):2375-2386.
124. Li Y, Wang K, Zhang B, et al. Salivary mycobiome dysbiosis and its potential impact on bacteriome shifts and host immunity in oral lichen planus. *Int J Oral Sci.* 2019;11(2):13.
125. Peters BA, Wu J, Hayes RB, et al. The oral fungal mycobiome: characteristics and relation to periodontitis in a pilot study. *BMC Microbiol.* 2017;17(1):157.
126. Liu Z, Chang ET, Liu Q, et al. Oral Hygiene and Risk of Nasopharyngeal Carcinoma- A Population-Based Case-Control Study in China. *Cancer Epidemiol Biomarkers Prev.* 2016;25(8):1201-1207.
127. Ye W, Chang ET, Liu Z, et al. Development of a population-based cancer case-control study in southern china. *Oncotarget.* 2017;8(50):87073-87085.

128. Chen GH, Liu Z, Ji MF, et al. Prospective assessment of a nasopharyngeal carcinoma risk score in a population undergoing screening. *Int J Cancer*. 2020;148(10):2398-2406.
129. Ji MF, Sheng W, Cheng WM, et al. Incidence and mortality of nasopharyngeal carcinoma: interim analysis of a cluster randomized controlled screening trial (PRO-NPC-001) in southern China. *Ann Oncol*. 2019;30(10):1630-1637.
130. Tedersoo L, Lindahl B. Fungal identification biases in microbiome projects. *Environ Microbiol Rep*. 2016;8(5):774-779.
131. Martin M. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnetjournal*. 2011;17.1:10-12.
132. Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581-583.
133. Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. *Stat Med*. 2010;29(9):1037-1057.
134. Harrell FE. Regression Modeling Strategies: with applications to linear models, logistic regression, and survival analysis. Springer-Verlag New York, Inc. New York, USA; 2010.
135. Lundberg M, Fredlund P, Hallqvist J, et al. A SAS program calculating three measures of interaction with confidence intervals. *Epidemiology*. 1996;7(6):655-656.
136. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8(4):e61217.
137. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12(6):R60.
138. Chen Y, Chang ET, Liu Z, et al. Residence characteristics and risk of nasopharyngeal carcinoma in southern China: A population-based case-control study. *Environ Int*. 2021;151:106455.
139. Chen Y, Chang ET, Liu Q, et al. Occupational exposures and risk of nasopharyngeal carcinoma in a high-risk area: A population-based case-control study. *Cancer*. 2021;127(15):2724-2735.
140. Yu MC, Yuan JM. Epidemiology of nasopharyngeal carcinoma. *Semin Cancer Biol*. 2002;12(6):421-429.
141. Josyula S, Lin J, Xue X, et al. Household air pollution and cancers other than lung: a meta-analysis. *Environ Health*. 2015;14:24.
142. Han D, Currell MJ, Cao G. Deep challenges for China's war on water pollution. *Environ Pollut*. 2016;218:1222-1233.
143. Xu M, Yao Y, Chen H, et al. Genome sequencing analysis identifies Epstein-Barr virus subtypes associated with high risk of nasopharyngeal carcinoma. *Nat Genet*. 2019;51(7):1131-1136.
144. Rueger S, Hammer C, Loetscher A, et al. The influence of human genetic variation on Epstein-Barr virus sequence diversity. *Sci Rep*. 2021;11(1):4586.

145. Yu G, Hsu W-L, Coghill AE, et al. Whole-Exome Sequencing of Nasopharyngeal Carcinoma Families Reveals Novel Variants Potentially Involved in Nasopharyngeal Carcinoma. *Sci Rep.* 2019;9(1).
146. Kumpitsch C, Moissl-Eichinger C, Pock J, et al. Preliminary insights into the impact of primary radiochemotherapy on the salivary microbiome in head and neck squamous cell carcinoma. *Sci Rep.* 2020;10(1).
147. Robinson S, Peterson CB, Sahasrabhojane P, et al. Observational Cohort Study of Oral Mycobiome and Interkingdom Interactions over the Course of Induction Therapy for Leukemia. *mSphere.* 2020;5(2).
148. Sanjaya PR, Gokul S, Gururaj Patil B, et al. Candida in oral pre-cancer and oral cancer. *Med Hypotheses.* 2011;77(6):1125-1128.
149. Ramirez-Garcia A, Rementeria A, Aguirre-Uribe JM, et al. Candida albicans and cancer: Can this yeast induce cancer development or progression? *Crit Rev Microbiol.* 2014;1-13.
150. Andrea H, Anett D, Gabor D, et al. Infectious Agents Associated with Head and Neck Carcinomas. In: Donelli G, editor. *Advances in Microbiology, Infectious Diseases and Public Health.* Vol. 1. Switzerland: Springer International Publishing; 2016. p. 63-80.
151. Gallo ML, Seldes AM, Cabrera GM. Antibiotic long-chain and α,β -unsaturated aldehydes from the culture of the marine fungus *Cladosporium* sp. *Biochem Syst Ecol.* 2004;32(6):545-551.
152. AlMatar M, Makky EA. *Cladosporium cladosporioides* from the perspectives of medical and biotechnological approaches. *3 Biotech.* 2015;6(1).
153. Hsu WL, Yu KJ, Chien YC, et al. Familial Tendency and Risk of Nasopharyngeal Carcinoma in Taiwan: Effects of Covariates on Risk. *Am J Epidemiol.* 2010;173(3):292-299.
154. Liu Z, Fang F, Chang ET, et al. Sibship size, birth order and risk of nasopharyngeal carcinoma and infectious mononucleosis: a nationwide study in Sweden. *Int J Epidemiol.* 2016;45(3):825-834.
155. Knight R, Vrbanac A, Taylor BC, et al. Best practices for analysing microbiomes. *Nat Rev Microbiol.* 2018;16(7):410-422.
156. Lindahl BD, Nilsson RH, Tedersoo L, et al. Fungal community analysis by high-throughput sequencing of amplified markers-a user's guide. *New Phytol.* 2013;199(1):288-299.
157. Diaz PI, Hong B-Y, Dupuy AK, et al. Mining the oral mycobiome: Methods, components, and meaning. *Virulence.* 2016;8(3):313-323.
158. Ali, Mac A, Morales, et al. Optimisation and Benchmarking of Targeted Amplicon Sequencing for Mycobiome Analysis of Respiratory Specimens. *Int J Mol Sci.* 2019;20(20):4991.
159. Nilsson RH, Anslan S, Bahram M, et al. Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nat Rev Microbiol.* 2018;17(2):95-109.
160. Skelly AC, Dettori JR, Brodt ED. Assessing bias: the importance of considering confounding. *Evid Based Spine Care J.* 2012;3(1):9-12.