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

PROGNOSIS OF NASOPHARYNGEAL CARCINOMA: BODY MASS INDEX, PLASMA EPSTEIN-BARR VIRUS DNA AND ORAL MICROBIOME

Yun Du

Karolinska Institutet

Stockholm 2023

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023

ISBN 978-91-8017-044-4
Cover illustration by Sisi Huang and Yun Du.

Prognosis of Nasopharyngeal Carcinoma: Body Mass Index, Plasma Epstein-Barr Virus DNA and Oral Microbiome

Thesis for Doctoral Degree (Ph.D.)

By

Yun Du

The thesis will be defended in public at Lecture Hall Atrium, Nobels väg 12B, Solna, Karolinska Institutet, 20th June 2023, 14:00.

Principal Supervisor:
Professor Weimin Ye
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Co-supervisor(s):
Statistician Li Yin
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Professor Magnus Nilsson
Karolinska Institutet
Department of Clinical Science, Intervention and Technology

Senior research specialist Amir Sohrabi
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Opponent:
Senior Investigator Anil Chatuvedi
National Cancer Institute
Division of Cancer Epidemiology & Genetics

Examination Board:
Docent Nele Brusselaers
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology
and
Antwerp University
Global Health Institute

Professor Lennart Greiff
Lund University
Department of Otorhinolaryngology

Docent Giorgio Tettamanti
Karolinska Institutet
Institute of Environmental Medicine

To time.

Abstract

Nasopharyngeal carcinoma (NPC) has a geographically skewed distribution worldwide, with high incidence rates in East and Southeast Asia. Although hospital-based studies suggest that the development of new radiotherapy techniques has contributed to improved NPC prognosis, population-based research on NPC patient survival is lacking. In addition, potential environmental prognostic factors for NPC, including body mass index (BMI) and body shape, pretreatment plasma Epstein-Barr virus (EBV) DNA, and oral microbiome, are not yet well understood. Therefore, this thesis aims to characterize population-based NPC survival patterns and identify prognostic factors for NPC in a population-based context in southern China.

In **Paper I**, we aimed to estimate population-based NPC survival in an unbiased manner by implementing a tracing strategy to ensure a high follow-up rate in a representative cohort of NPC patients. We also aimed to compare this estimate with survival results from other studies in endemic areas, and to calculate the number of avoidable deaths from earlier detection or more widespread access to advanced medical care. Based on patients with incident NPC enrolled in the population-based NPC Genes, Environment, and EBV (NPCGEE) project, we developed a passive-active-passive circle follow-up strategy that achieved a high rate (98.3%) of complete follow-up for vital status through 2018. We estimated that 5-year overall survival for NPC diagnosed at stages I, II, III, IVa, IVb, and IVc was 91.1%, 88.1%, 79.8%, 63.8%, 57.7%, and 34.4%, respectively. In general, we found that population-based NPC survival lags by approximately 10 years behind survival reported in large hospital-based cohorts. We estimated that 174 NPC deaths per 1000 patients could be avoided within five years of diagnosis if all advanced-stage cases were instead diagnosed at early stages.

In **Paper II**, we examined whether pretreatment BMI and body shape were associated with the prognosis of NPC, using the NPC patient cohort from the NPCGEE project. We found that being overweight at diagnosis, was associated with a 25% lower all-cause mortality rate whereas those with a thinner body shape had a higher all-cause mortality rate, than those with a normal weight/body shape. When we examined associations with BMI and body shape 10 years before diagnosis, similar but weaker associations existed, but for BMI and body shape at age 20 years, the associations with NPC prognosis disappeared. The lack of effect modification by stage at diagnosis, along with the detection of similar associations with BMI and body shape 10 years before diagnosis, suggests that the results were not primarily due to reverse causation.

In **Paper III,** we assessed the relationship between pretreatment plasma EBV DNA and NPC survival, using the NPC patient cohort from the NPCGEE project. We found that higher pretreatment plasma EBV DNA load was associated with increased risks of all-cause and NPC-specific mortalities, particularly in the first five years after diagnosis; cases with detectable plasma EBV DNA (compared with undetectable) had more than double the risk of all-cause and NPC-specific death. Higher pretreatment EBV DNA levels may reflect a greater tumor burden, and may signal a need for more intensive chemotherapy and/or heightened clinical surveillance.

In **Paper IV,** we estimated associations of oral microbiome with NPC prognosis, using a subcohort of patients with saliva specimens from the NPCGEE project. We showed that lower within-community diversity was associated with higher all-cause and NPC-specific mortalities, and some (albeit not all) measures of between-community diversity were also associated with all-cause and NPC-specific mortalities. None of candidate bacteria were found to be significant prognostic biomarkers, suggesting

that the observed associations resulted from global patterns rather than specific microbiota. These results indicate that microbiota may affect host immune function or contribute to the development of adverse treatment effects that in turn influence NPC prognosis.

In conclusion, using a population-based patient cohort of NPCGEE project in southern China, we estimated generalizable 5-year survival rates and investigated potential environmental prognostic factors. We found that population-based NPC survival lags behind large-hospital-based survival; overweight at diagnosis indicated a favorable long-term prognosis, whereas a thinner body shape at diagnosis is associated with worse prognosis; pretreatment plasma EBV DNA is a strong prognostic factor for NPC; and decreased within-community diversity in oral microbiome is related to increased mortality. Taken together, these findings constitute some of the first population-based evidence on NPC prognosis in southern China, and point to potential routes to improving NPC management and long-term outcomes in this NPC-endemic region.

List of scientific papers

* Equal contribution

I. **Du Y***, Feng R*, Chang E T, Yin L, Huang T, Li Y, Zhou X, Huang Y, Zhou F, Su C, Xiao X, Jia W, Zheng Y, Adami H-O, Zeng Y, Cai Y, Zhang Z, Xu M, Ye W. Population-based nasopharyngeal carcinoma survival study in southern China. (Manuscript)

II. **Du Y***, Feng R*, Chang E T, Yin L, Huang T, Li Y, Zhou X, Huang Y, Zhou F, Su C, Xiao X, Jia W, Zheng Y, Adami H-O, Zeng Y, Cai Y, Zhang Z, Xu M, Ye W. Body mass index and body shape before treatment and nasopharyngeal carcinoma prognosis: a population-based patient cohort study in southern China. International Journal of Cancer, 2023 March; 1153(2):290-301

III. **Du Y***, Feng R*, Chen Y, Chang E T, Yin L, Huang T, Huang Y, Li Y, Zhou X, Zhou F, Su C, Xiao X, Jia W, Zheng Y, Adami H-O, Zeng Y, Cai Y, Xu M, Zhang Z, Ye W. Pre-treatment plasma EBV DNA and nasopharyngeal carcinoma prognosis: a prospective population-based cohort study in southern China. (Manuscript)

IV. **Du Y***, Feng R*, Chang E. T*, Debelius J. W, Yin L, Xu M, Huang T, Zhou X, Xiao X, Li Y, Liao J, Zheng Y, Huang G, Adami H-O, Zhang Z, Cai Y, Ye W. Influence of pre-treatment saliva microbial diversity and composition on nasopharyngeal carcinoma Prognosis. Frontiers in Cellular and Infection Microbiology, 2022 March; 12:831409

Related work

(Not included in the thesis)

Du Y, Yu X, Chang ET, Lian S, Wu B, Li F, Chu B, Wei K, Zhan J, Liang X, Ye W, Ji M. Pre-diagnostic anti-EBV antibodies and primary liver cancer risk: a population-based nested case-control study in southern China. BMC Cancer. 2023; 23:250.

Du Y, Zhang W, Lei F, Yu X, Li Z, Liu X, Ni Y, Deng L, Ji M. Long-term survival after nasopharyngeal carcinoma treatment in a local prefecture-level hospital in southern China. Cancer Manag Res. 2020;12:1329-1338

Yu X, Ji M, Cheng W, Wu B, Lian S, **Du Y**, Cao S. A retrospective cohort study of nasopharyngeal carcinoma screening and hepatocellular carcinoma screening in Zhongshang city. J Cancer 2019; 10(8):1909-1914.

Contents

List of abbreviations

ACT	Adjuvant chemotherapy
ASV	Amplicon sequences variants
BMI	Body mass index
CCRT	Concurrent chemoradiotherapy
NPC	Nasopharyngeal carcinoma
EBNA1-IgA	Immunoglobulin A of EBV nuclear antigen 1
EBV	Epstein-Barr virus
Faith's PD	Faith's phylogenetic diversity,
2DRT	Two-dimensional radiotherapy
3DRT	Three-dimensional radiotherapy
FDR	False discovery rate
HR	Hazard ratio
ICT	Induction chemotherapy
IMRT	Intensity-modulated radiotherapy
KPS	Karnofsky Performance Scale
NPCGEE	NPC Genes, Environment, and EBV project
PCoA	Principal coordinates analysis
qPCR	Quantitative polymerase chain reaction
PCs	Principal coordinates
RPCA	Robust Aitchison principal-component analysis
SEER	Surveillance, Epidemiology, and End Results Program
VCA-IgA	Immunoglobulin A of EBV viral capsid antigen

1 Introduction

Nasopharyngeal carcinoma (NPC) is an uncommon cancer worldwide but is prevalent in populations in Southeast Asia, Middle East and North Africa [1]. Due to the histology of undifferentiated type and hidden anatomical location, radiotherapy together with chemotherapy, rather than surgery plays an essential role in treatment [2]. The tumor location in the deep nasopharynx and asymptomatic feature brings difficulties in early diagnosis, thus, most patients were diagnosed at advanced stages [3]. The population-based 5-year survival probabilities of NPC were 79.3% and 65.2% in Hong Kong and Taiwan, respectively [4,5]. However, data on population-based survival probabilities remain largely unknown in China. Further, the investigation of potential environmental prognostic factors in a population-based setting has been limited.

This thesis outlines the long-term population-based survival probabilities of NPC, compares the estimates with those from other studies in endemic areas, and calculates the avoidable deaths from earlier detection or more widespread access to advanced medical care. Further, this thesis investigates the associations of pretreatment body mass index (BMI) and body shape, pretreatment plasma Epstein-Barr virus (EBV) DNA and saliva microbiota, with the prognosis of NPC in a population-based context.

2 Literature review

2.1 General background on nasopharyngeal carcinoma

2.1.1 Descriptive epidemiology

Nasopharyngeal carcinoma (NPC) is a highly malignant cancer that originates in the nasopharynx (**Figure 2.1**). NPC is relatively rare worldwide, with an age-standardized incidence of less than one per 100,000 person-years [1]. Age-standardized incidences in southern China, however, showed as high as 25 per 100,000 person-years (**Figure 2.2**). Based on global cancer statistics, the estimated number of incident NPC cases in 2020 was 133,354 globally and ranked 23rd among all cancers. By region, 113,659 (85.2%) were diagnosed in Asia, 10,041 (8.8%) in Africa and 5,204 (4.6%) in Europe in 2020. In China, the number of estimated incident cases was 62,444 in 2020 [6] and the distribution of NPC cases varied geographically. In northern China, the age-standardized incidence is lower than 2/100,000 person-years whereas in southern China, it may exceed 25/100,000 person-years [7]. In southeast Asia (i.e., Singapore and Malaysia), high NPC incidence seems to be associated with social and racial mixture with populations from southern China. However, in Japanese populations, which primarily interact with populations from northern China, NPC incidence is low [7]. The incidence in males is twice or triple, that in females [8]. It has declined gradually worldwide since 1970 [7,9], probably due to economic growth and its implications.

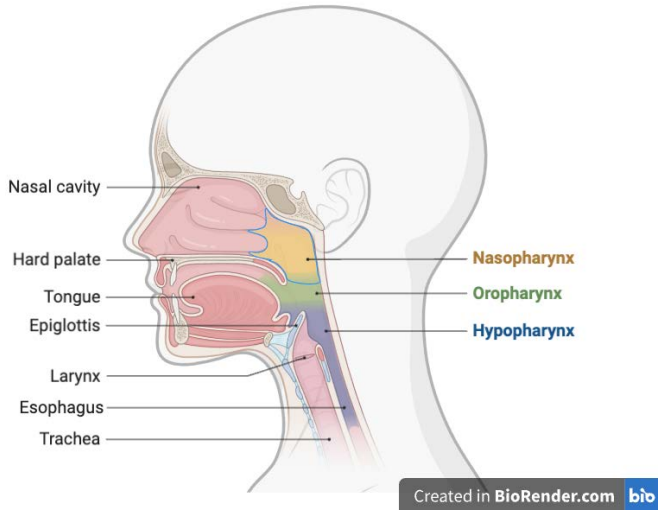


Figure 2.1 The location of the nasopharynx (Created by BioRender: https://www.biorender.com/)

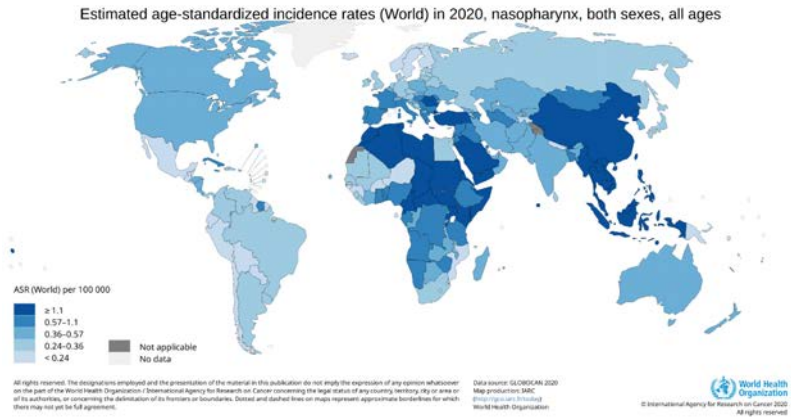


Figure 2.2 Estimated age-standardized incidence rate (World) of nasopharyngeal carcinoma in 2020 worldwide.

2.1.2 Risk factors

NPC occurrence concerns the interplay between genetic and environmental factors [1]. Epidemiological studies on migration patterns offer some hints of this interplay. The incidence of NPC among immigrants from high-risk regions to low-risk regions has declined, albeit has remained higher than that among local natives [10,11] and it is higher among their offspring [12,13]. The risk of NPC among populations emigrating from low-risk regions to high-risk areas has increased [14,15]. The phenomenon of family clustering of NPC also strengthens the importance of gene-environment interaction, as a segregation study of familial NPC explained [16].

2.1.2.1 Genetic risk factors

Numerous genetic risk factors have been identified and linked to the pathogenesis of NPC (**Table 2.1**). The most investigated genes include human leukocyte antigen (HLA)-A/B/C, which are involved in the immune system's ability to recognize and eliminate abnormal cells, based on large-scale genome-wide association studies [17–20]. The macrophage-stimulating 1 receptor (MST1R) gene plays an essential role in defense against viral infection according to a whole-exosome sequencing study [21]; integrin-α 9 (ITGA9) and RAD51L1 were reported to be associated with increased NPC risk [22,23], related to cell-cell mediation and DNA damage repair. In addition, many statistical associations have been observed but further validation studies across diverse populations are needed.

Table 2.1 Genetic alterations associated with NPC risk.

Studies	Genes	Function	Location	Direction of associations
Bei et al. [17]	Human leukocyte antigen (HLA)-A/B/C	Immune response	6p21.3	Increased/decreased risk
Dai et al. [21]	macrophage-stimulating 1 receptor (MST1R)	Host defense against viral infection	3p21.3	Decreased risk
Ng et al. [22]	Integrin-α 9 (ITGA9)	Mediate cell–cell and cell–matrix adhesion	3p21	Increased risk
Qin et al. [23]	RAD51L1	DNA damage repair		Increased risk

Abbreviations: NPC, nasopharyngeal carcinoma.

2.1.2.2 Environmental risk and preventive factors

The causes of NPC remain inconclusive, although certain genetic risk factors have been identified. There is no single causal risk factor responsible for the occurrence of NPC. A comprehensive gene-environment interaction has been postulated. In addition to the genetic factors, some environmental risk factors are well-confirmed, including older age, Cantonese background, EBV infection, family history of NPC, tobacco use, male sex and consumption of Chinese-style salted fish in early life (**Table 2.2**). In high-incidence areas, NPC incidence peaks at approximately 45-60 years of age. About 83% of NPC cases are associated with EBV infection [24]. Of note, specific EBV strains may be related to NPC. For instance, based on a large-scale genome sequencing study, two EBV variants within BALF2 genes, encoding a major binding protein, were found to be strongly correlated with NPC risk with an odds ratio of 8.69 [25]. Anti-EBV antibodies (i.e., immunoglobulin A of EBV nuclear antigen 1, EBNA1-IgA and immunoglobulin A of EBV viral capsid antigen, VCA-IgA) have been used for NPC screening due to their strong associations [26]. Tobacco is one confirmed cause of NPC according to the International Agency for Research on Cancer [27]. Traditional Asian pickled vegetables are defined as possible carcinogens. Our population-based cohort study found a modest association between hard salted fish (directly salted and afterwards dried) and NPC risk [28].

Table 2.2 Environmental risk and preventive factors for NPC[a].

Factors	Odds ratio/Relative risk	Direction of associations
Well-confirmed		
Older age (up to ~60 years in high-incidence areas)		↑↑
Cantonese		↑↑
EBV infection (anti-EBV IgA serology)		↑↑
First degree family history of NPC [29]	4-20	↑↑
Tobacco [30]	1.3-1.6	↑
Male	2-3	↑
Chinese-style salted fish in early life	1.1-1.5	↑
Possible		
EBV variation (BALF2 gene) [25]	8.69	↑↑
Indoor air pollution	1.1-3.5	↑
Chronic respiratory tract infection		↑
Occupational wood dust/smoke		↑
Animal-based diet [31]	2.2	↑
Plant-based diet [31]	0.5	↓
Inconclusive		
Traditional herbal medicine [32,33]		↓/↑/null
Alcohol [34]		↑/null
Tea [34]		↓/null
Occupational formaldehyde		↑/null

[a] Adapted from the article by Chang et al. [1] and other studies cited in the table. Abbreviations: NPC, nasopharyngeal carcinoma; EBV, Epstein-Barr virus.

2.1.3 Treatment

The histological subtypes of NPC include type I (keratinizing squamous cell carcinoma), type II (nonkeratinizing carcinoma) and type III (basaloid squamous cell carcinoma) [35]. The majority of NPC cases are nonkeratinizing undifferentiated carcinoma, which are sensitive to radiation. Radiotherapy plays a key role in NPC treatment, combined with/without chemotherapy because of its deep and hidden location (**Figure 2.1**) and histology. For patients with cancer stage I (according to the 8th version of the American Joint Committee on Cancer, AJCC) who are plasma EBV DNA negative, definitive radiotherapy is recommended. For patients with cancer stages T0-2 (EBV DNA positive),

N1 and M0 or T3, N0 and M0, concurrent radio-chemotherapy with/without induction/adjuvant chemotherapy is suggested. For patients with cancer stages T3-4, N1-3 and M0 or any T, N2-3 and M0, clinical trials or concurrent systematic therapy with induction/adjuvant chemotherapy is recommended. For patients with metastatic disease, the treatment principal is dependent on the conditions of the individuals, although clinical trials are strongly recommended (**Table 2.3**) [36].

Table 2.3 NPC treatment guideline based on national comprehensive cancer network[a]

Clinical stage[b]		Treatment
T1, N0, M0		Definitive RT.
T2, N0, M0		Definitive RT ± concurrent therapy if it is bulky tumor size or EBV+.
T0 (EBV+)-2, N1, M0 Or T3, N0, M0		Concurrent therapy ± induction/adjuvant chemotherapy if it is bulky tumor size or EBV+.
T3-4, N1-3, M0 Or any T, N2-3, M0		Clinical trials, or concurrent therapy + induction/adjuvant.
Any T, any N, M1	Oligometastatic	Induction chemotherapy + radiotherapy/concurrent radiotherapy
	Widely metastatic and good performance status	Systematic therapy
	Widely metastatic and poor performance status	Supportive care

[a] Adapted from national comprehensive cancer network [36].
[b] Clinical stage is the 8th Version of TNM clinical stage by American Joint Committee on Cancer.
Abbreviations: NPC: nasopharyngeal carcinoma; T, primary tumor stage; N, regional lymph nodes stage; M, metastasis stage; EBV, Epstein-Barr virus; RT, radiotherapy.

2.2 Prognosis

In general, the mortality rate for NPC is relatively low compared to other types of cancer. It accounted for approximately 1% of all cancer deaths, and was ranked as the 21st most frequent cause of cancer death worldwide in 2020 [6]. The age-standardized mortality rate declined from 3.9/100,000 in 1990 to 2.2/100,000 person-years in 2019 for Chinese males, and for Chinese females and populations worldwide, it followed a similar pattern [37–40].

According to the American Cancer Society, the 5-year overall survival probability of NPC is approximately 64%. However, the mortality rate for advanced stages of NPC is much higher, with a 5-year survival rate of less than 50% [41]. The well-established prognostic factors, which encompass the tumor characteristics, host factors, and treatment modalities, have been compiled and presented in **Table 2.4.**

Table 2.4 Well-established prognostic factors of NPC

Prognostic factors	Survival probability
Tumor	
Histology	Non-keratinizing squamous cell carcinoma better than (>) keratinizing squamous cell carcinoma [42]
Clinical cancer stage	Early stage better than (>) late stage
Host	
Age	Younger age better than (>) elder age
Sex	Female better than (>) male [43]
Karnofsky Performance Score (KPS)	Higher KPS better than (>) lower KPS
Smoking	Never smokers better than (>) smokers
Nutrition	Malnutrition with worse survival [44]
Treatment	
Radiotherapy	IMRT better than (>) 2DRT/3DRT [45]

Abbreviations: NPC, nasopharyngeal carcinoma; EBV, Epstein-Barr virus; IMRT, intensity-modulated radiation therapy; 2DRT, two-dimensional conformal radiotherapy; 3DRT, three-dimensional conformal radiotherapy.

In addition to the well-established prognostic factors described above, numerous unclear and novel factors are emerging. This thesis concentrates on three inconclusive prognostic factors within a population-based cohort. Specifically, the latest updates on associations between pretreatment body mass index (BMI) and plasma EBV DNA, as well as the microbiome, are highlighted in the following sections.

2.2.1 Population-based NPC survival probabilities worldwide.

Population-based studies investigate specific populations to gain insight into health, disease or social issues. Given the data collected from a representative sample of individuals within a particular region or demographic group, the findings are generalizable to the target population according to the study hypothesis, not only to individuals included in the study. Population-based cancer survival probabilities can be a good indicator for public health organizations and guide research and financial support for decision makers, as well as reflect treatment outcomes of the disease.

Generally, NPC survival probabilities have been increasing with the development of radiotherapy techniques, from two-dimensional radiotherapy (2DRT) era to intensity-modulated radiotherapy (IMRT) era [45]. One large-scale register-based study calculated dynamic age-standardized 5-year relative survival rates in China and observed a slight increase from 43.8% during the period of 2003-2005 to 45.4% during the period of 2012-2015 [46] , despite incomplete information on tumor morphology and cancer stage. A population-based survival study in Taiwan, China, retrospectively identified 13,407 cases of NPC between 2002 and 2010, using population-based registries and all cases recorded in the medical library, and reported an overall 5-year survival probability of 65.2%. However, the study did not include cancer stage information[4].

Beyond China, stable increases in 5-year survival rates were observed in the Americas and Europe. In Ontario Canada, it was 52.9% between 1984 and 1986 and increased to 60.2% between 1999 and 2001, whereas in Surveillance, Epidemiology, and End Results Program (SEER), it was 44.7% between 1984 and 1986, and 56% between 1999 and 2001 [47]. The 5-year overall survival rates in Europe rose from 39.0% between 1990 and 1994 to 46.0% between 2000 and 2007 [48].

These studies showed that the 5-year overall survival probabilities of NPC patients range from 40.0% to 78.2%, depending on the region and period, yet there are several limitations. First, few studies presented cancer stage distribution, one of the most important prognostic factors. One population-based NPC screening cohort identified 153 NPC cases with a 5-year survival of 77.6%, however the proportion of early-stage cancer was 45.9%, which was much higher than that in other population-based studies [49]. Second, limited studies have considered loss to follow-up, which would bias the estimates of 5-year survival rates [50]. Third, these studies encountered coverage issues, which may have impacted the generalizability and validity of the findings. For instance, the SEER program covers 50% of the American population, and most participants live in metropolitan areas [51].

2.2.2 BMI and body shape and NPC prognosis.

Overweight and obesity refer to an atypical or excessive buildup of body fat that poses a health hazard. BMI, which is determined by dividing the weight in kilograms (kg) by square of their height in meters, provides a useful approximation for determining overall body fat levels. For adults, overweight is characterized by a BMI ranging from 25.0 to 29.9, while obesity is indicated by a BMI

of 30 or higher based on the World Health Organization (WHO) [52]. Four more cutoff points specific to Asians at risk have also been established (23.0, 27.5, 32.5 and 37.5) [53]. According to estimates from 2014, the age-standardized prevalence of obesity among men and women worldwide, was 10.4% and 14.4%, respectively [54]. Between 1975 and 2016, there was a notable rise in the prevalence of obesity worldwide, with the number of cases almost tripling during this period. In 2016, 39% of individuals aged 18 years and over were classified as overweight, and 13% were categorized as obese [54,55].

In addition to BMI, other anthropometric measures are employed to access body fatness, such as weight, body shape, waist circumference, and waist-to-hip ratio. These measurements provide additional information, allowing for a more thorough evaluation of an individual's body composition and health status [56]. By considering multiple anthropometric measures, healthcare professionals can obtain a more comprehensive understanding of an individual's overall health. For instance, body shape, reflects how fat is distributed in the body.

The impact of obesity on cancer patient survival remains inconclusive [57] , although there is convincing evidence linking excess body fat, i.e., as measured by BMI, to an increased risk of various types of cancer [58,59]. Obesity paradox is used to describe the condition that a higher BMI is related to better survival outcomes for certain cancers, even though obesity is a recognized risk factor for multiple cancer types [55]. This paradox has been observed in several cancers, including breast, prostate, and lung cancer, and has stimulated considerable debate in the scientific community [60–63]. Proposed explanations for this paradox include the possibility that individuals with a higher BMI may have greater energy reserves to draw upon during cancer treatment, or that the inflammatory response associated with obesity may have a protective effect against certain cancers [64]. The obesity paradox does not represent a true causal relationship. Confounding factors such as age, sex, comorbidities and treatment pattern may influence the observed associations [65].

The risk of NPC is lower among individuals who were underweight or had a normal weight at age 20 years and 10 years before diagnosis, than those who were overweight or obese [66]. However, there is inconsistent evidence regarding the association between BMI and mortality after NPC diagnosis [67–69]. Some studies have reported that patients with a higher BMI showed a lower risk of NPC mortality compared to those who were underweight or normal weight [68,70,71]. However, concerns have been raised regarding the validity of these findings due to potential reverse causality and residual confounding [72]. Reverse causality may exist when low BMI is the result of cancer progression and cachexia, especially among patients diagnosed at advanced stages [73]. Additionally, residual confounding may occur when studies only control for certain factors such as sex, age, alcohol consumption, smoking, clinical cancer stage, and treatment modality, but no other relevant confounders such as education and occupation.

There is evidence suggesting that body weight in early life can impact health outcomes later in life [74]. However, there is a lack of research on the prognostic implications of BMI and body shape measured prior to the diagnosis of NPC. To address these gaps, we prospectively analyzed a population-based NPC patient cohort in southern China [75]. We examined the impact of BMI and body shape at age 20 years, as well as 10 years prior to diagnosis and at the time of diagnosis, on overall and NPC-specific survival rates to reduce reverse causation and residual confounding.

2.2.3 Plasma EBV DNA and NPC prognosis

Plasma EBV DNA has been considered marker for guiding treatment, predicting prognosis, and evaluating the response to therapy [36]. Circulating EBV DNA may originate from either apoptotic host cells or destroyed virus particles. Polymerase chain reaction (PCR) techniques targeting BamH1-W, LMP, EBNA1, and BZLF1 have been used to detect EBV DNA [76,77]. The BamH1-W sequence showed the highest sensitivity in quantitative PCR (qPCR). Assays using plasma had higher sensitivity (69%-99%) and specificity (87%-100%) compared to serum detection (sensitivity: 31%-87%, specificity: 83%-100%) [78,79].

The release of EBV DNA into the bloodstream, which likely originates from NPC tumor cells, may be correlated with tumor burden. The immune function of NPC cases is compromised, with decreased CD4+ and CD8+ T cells, disturbing surveilling and controlling EBV-host cells that harbor and replicate the virus [80]. Lo et al. [81,82] demonstrated that the median plasma EBV DNA copy number was significantly higher in advanced NPC cases than in early-stage NPC cases. Two meta-analyses reported that high pretreatment plasma EBV DNA levels, detectable midtreatment plasma EBV DNA, detectable posttreatment plasma EBV DNA, and slow clearance rates of plasma EBV DNA half-life were significantly linked to worse overall survival [83,84].

Pretreatment plasma EBV DNA concentration has been investigated as a predictor for both short-term and long-term survival in early- and late-stage of NPC [85–87]. Leung et al. [88] studied 90 early stages (I-II) NPC cases and found that higher pretreatment plasma EBV DNA concentration was associated with higher distant failure. Lo et al. [89] investigated the role of pretreatment serum/plasma EBV DNA in predicting early clinical events after treatment completion. Of 91 NPC patients with complete radiotherapy, those with local recurrence or distant metastasis had a median plasma EBV DNA concentration of 41,756 copies/ml, much higher than those without the corresponding events (median concentration, 5,807 copies/ml). The risk ratio of each tenfold increase in plasma EBV DNA was 3.8. Jin et al. [90] performed a retrospective study of 1036 NPC cases of stage III-IVb without distant metastasis and suggested an independent prognostic value of pretreatment plasma EBV DNA on long-term survival outcomes.

However, plasma EBV DNA for monitoring NPC prognosis remains uncertain in population-based validation studies [36]. To address the knowledge gaps, we utilized a large population-based cohort of NPC patients with collected plasma before treatment [75] to examine the associations between pretreatment plasma EBV DNA and NPC prognosis, with a focus on time-dependent and dose-response associations.

2.2.4 Microbiome and NPC

Human microbiota consists of bacteria, viruses and fungi, that colonize in different locations. Commensal bacteria are most frequently studied. The composition is influenced by age, diet, antibiotic use, environmental exposure, and geographic location [91–93]. New evidence has emerged indicating that these microbes could potentially increase the likelihood of developing specific types of cancer and affect the response to treatments.

Inadequate oral hygiene and related diseases have been associated with NPC risk [94,95]. Studies have shown that the microbiota in NPC patients was less abundant or rich in community diversity than in

healthy controls [96,97]. Additionally, the microbiota might interact with the known risk factor, EBV. Liao et al. [98] conducted a cross-sectional study of 186 NPC patients and 153 healthy controls from a physical examination center, tested saliva microbiota by 16S rRNA, and found that *Streptococcus sanguinis* was more abundant in NPC patients and was positively correlated with anti-VCA IgA, an established biomarker of NPC risk and EBV reactivation.

During radiation therapy, the composition of bacterial community undergoes gradual changes, accompanied by a significant rise in the proportion of certain gram-negative bacteria [99]. One longitudinal study collected nasopharyngeal swabs biweekly during radiotherapy. Significant beta diversity changes were observed between early and late responders. The study suggests that microbiome changes during radiotherapy could potentially serve as a short-term predictor of therapeutic response [100]. Another longitudinal study showed that the abundance of *Streptococcus* and *Actinobacillus* was higher among those with severe mucositis than those without severe mucositis during radiotherapy [101]. In a cohort study of 802 NPC patients [102], intratumoral bacterial load was found to be negatively associated with T-lymphocyte infiltration and linked to poor survival outcomes. Nevertheless, limited research has investigated the relationship between the long-term survival of NPC patients and the oral microbiota.

3 Research aims

The primary aim was to employ an efficient follow-up strategy to achieve a low loss to follow-up rate, measure the population-based NPC survival in southern China and investigate the potential prognostic factors regarding lifestyle, an EBV-related biomarker, and oral microbiota.

The specific research aims are illustrated in **Figure 3.1**:

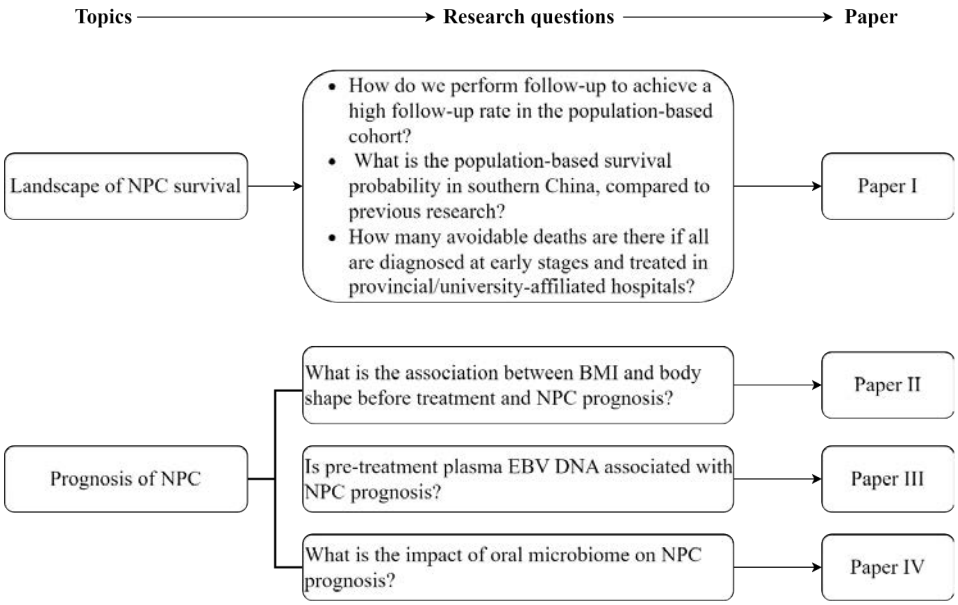


Figure 3.1 Illustration of research questions for each paper.
Abbreviations: NPC, Nasopharyngeal carcinoma; EBV, Epstein-Barr virus; BMI, body mass index.

4 Study design and methodology

4.1 Study design and study population

4.1.1 Population-based cohort study design

The population-based cohort in this thesis comprised the patient cohort of a project entitled 'NPC Genes, Environment, and EBV (NPCGEE)'. The case participants were newly diagnosed between 2010 and 2013 from the study base of 13 cities/counties with 8 million people in the Zhaoqing area of Guangdong Province and the Wuzhou and Guiping/Pingnan areas of Guangxi Autonomous Region in southern China [75]. The participants were eligible for the study if they currently resided in the study areas and were aged between 20 and 74 years, fluent in Cantonese, and able to participate in the study interview without prevalent malignancies or inherited or acquired immune deficiency. In total, we identified 3047 newly diagnosed NPC patients within the study areas, out of which 2553 (83.8%) were included based on eligibility criteria. All the participants completed the questionnaire through face-to-face interviews and 2350 patients donated saliva samples at diagnosis. Between 2018 and 2023, we performed an additional medical record review and found 24 cases that were not incident NPC cases, therefore, we excluded them, leaving 2529 participants in the patient cohort (**Figure 4.1**).

4.1.2 Study I

We aimed to explore a follow-up strategy to obtain a low rate of loss to follow-up, calculate the survival probabilities for this population-based patient cohort, compare our study with other studies in NPC-endemic areas, and compute avoidable deaths.

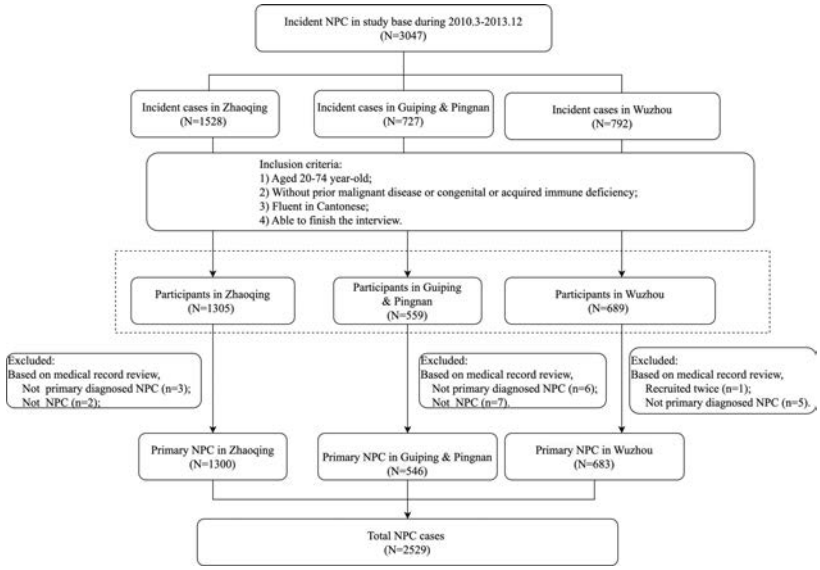


Figure 4.1 Study population for Study I. NPC, nasopharyngeal carcinoma.

4.1.3 Study II

We aimed to assess the associations between BMI and body shape, measured at age 20, 10 years before diagnosis, and at diagnosis, and all-cause and NPC-specific mortality among NPC patients,

accounting for reverse causation and residual confounding. After excluding three cases that lacked information on prediagnosis BMI and/or body shape, the final analysis included 2526 patients (**Figure 4.2**).

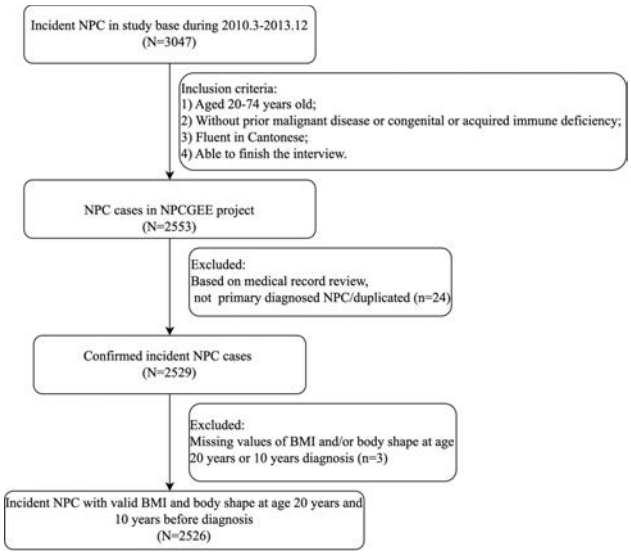


Figure 4.2 Study population for Study II. NPC, nasopharyngeal carcinoma.

4.1.4 Study III

A total of 1854 cases were included in the final analysis after excluding 386 cases who lacked measurements for the plasma EBV DNA load and 529 cases with samples collected after therapy (**Figure 4.3**).

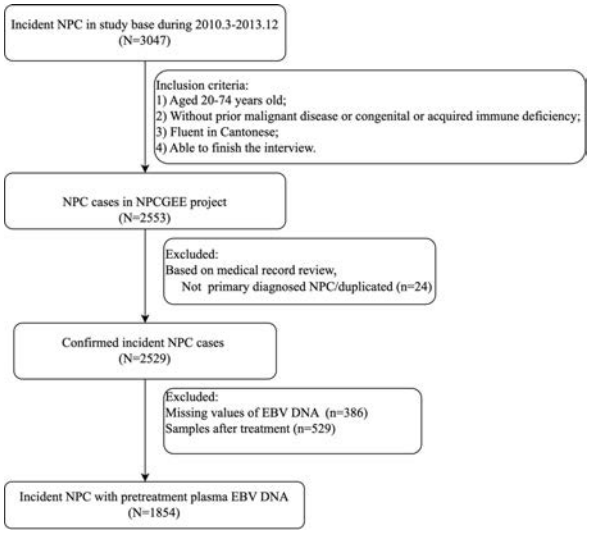


Figure 4.3 Study population for Study III. NPC, nasopharyngeal carcinoma.

4.1.5 Study IV

We aimed to detect potential associations between oral microbiota patterns and mortality of NPC patients, utilizing saliva bacterial profiles obtained through 16S rRNA sequencing. The study population comprised NPC cases in the Wuzhou area, a subcohort of NPCGEE. Eighty-nine were excluded because they refused to provide saliva specimens, and 58 had their saliva DNA extracted using varied procedures. Fifteen cases were excluded due to failed library preparation, low sequence counts, ambiguous sequencing identifiers, duplicated enrollment, or not having primary NPC. Thirty-two former smokers and 13 cases with missing values of covariates were further excluded. In total, 482 cases remained for the final analysis (**Figure 4.4**).

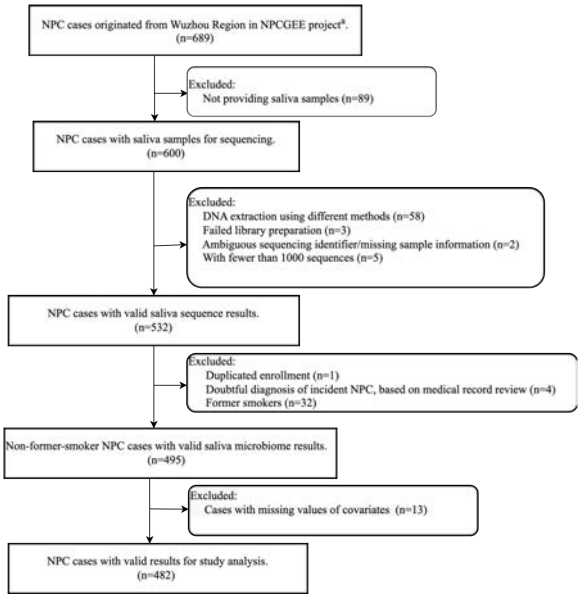


Figure 4.4 Study population for Study IV. NPC, nasopharyngeal carcinoma.

4.2 Main measurements

In the NPCGEE project, the electronic lifestyle questionnaires were composed of the following variables: basic personal characteristics, residential history, occupational history, medical history (oral hygiene, family medical history etc), smoking history, alcohol, and tea consumption history, dietary history (10 years ago and teenage), reproductive history and use of Chinese herbal medicine [75]. Biospecimens were collected, including blood, saliva, hair, fingernail, and toenail. Blood samples were refrigerated at 4°C within three days, delivered to the laboratory, processed to isolate plasma, serum, red blood cells, and buffy coat, and frozen at -80°C [75].

Between 2018 and 2023, an additional medical record review of NPC cases was launched. We made a user-customized follow-up sheet including information on histology, clinical cancer stage, and treatment. The medical records were reviewed by five clinical medical students and a junior oncologist. To ensure accuracy, a senior oncologist randomly checked 10% of the documents. The information from the follow-up sheets was entered into the *Epidata* (3.1 version) database by one person and then verified by another to ensure accuracy.

4.2.1 Exposures

*4.2.1.1 Pretreatment BMI (**Study II**)*

BMI and body shape at age 20 years, 10 years prior to diagnosis, and at diagnosis were obtained through lifestyle questionnaires or medical record reviews.

BMI was classified into four groups following WHO guidelines for Asian populations: underweight (<18.5 kg/m2), normal weight (18.5-22.9 kg/m2), overweight (23.0-27.4 kg/m2), and obese (≥ 27.5 kg/m2) [53]. Body shape was assessed using the revised Stunkard's Figure Rating Scale, which had seven male and nine female figures [103]. The reference groups were body shape 3 and normal weight due to their larger sample sizes, with some adjacent groups merged for subgroup analyses due to limited sample size.

*4.2.1.2 Pretreatment plasma EBV DNA (**Study III**)*

Pretreatment plasma EBV DNA load was quantified using real-time qPCR [81]. Initially, blood samples were centrifuged at 1500 rpm for 15 minutes, aliquoted into microtubes, and stored at -80 °C. Total nucleic acid was extracted from 400 µl plasma using a QIAamp DNA Blood Mini Kit, and eluted from the column using a final elution volume of 50 µL. Then qPCR was performed using primers and probe specific to the BamHI-W region (HONG KONG TECH DRAGON LTD). Five microliters of plasma EBV DNA was distributed and amplified in an ABI 7300 Real-Time PCR system together with a negative control (purified nuclease-free water) and positive controls of the clonal synthesis of the BamHI-W region diluted into six different concentrations. To calculate the target concentration of EBV DNA in plasma (copies/mL), the standard curve and the equation below were used.

$$C = Q \times \frac{V_{DNA}}{V_{PCR}} \times \frac{1}{V_{ext}}$$

C represents for the concentration in targeted plasma (copies/ml), Q for the targeted quantity (copies) based on the PCR detector, V_{DNA} for the total volume of DNA yielded after extraction (50 µL in our case), V_{PCR} for the volume of DNA used for PCR (5 µL in our case), and V_{ext} for the volume of plasma used for extraction (400 µL in our case). For analysis, we classified detectable plasma EBV DNA load into tertiles (low, medium, and high).

*4.2.1.3 Oral microbiota profiling (**Study IV**)*

16S rRNA sequencing

For DNA extraction, twenty-three saliva samples and one autoclaved blank Eppendorf tube with nuclease-free water as a negative control were used each round. The process involved mixing 1 mL of saliva sample with 100 µL lysozyme lysis buffer and incubating it at 37°C for 60 minutes. After adding 0.5-mm-diameter and 0.1-mm-diameter beads (BioSpec, Bartlesville, OK) to each sample, they were blended at top speed for 10 minutes to physically disrupt the microbial cells. The resulting clean liquid was then transferred to a new 2 mL Eppendorf tube for each sample.

The isolation and purification of total DNA was performed using the TIANGEN TIANamp Blood DNA Kit. The DNA concentration was measured using a Qubit 3.0 Fluorometer, Thermo Fisher Scientific. The 16S rRNA amplicon library was generated using 341F/805R primers

(CTACGGGNGGCWGCAG,GACTACHVGGGTATCTAATCC), amplified for 20 cycles (30 seconds at 98°C for melting, 30 seconds at 60°C, and 30 seconds at 72°C), and then barcoded and cleaned up in a second PCR step. The DNA purity and volume were evaluated using an Agilent 2100 Bioanalyzer system and real-time PCR. Finally, sequencing was performed at Beijing Genome Institute on an Illumina MiSeq using a 2 × 300 bp paired-end strategy. In every batch, two blank controls containing nuclease-free water and one control containing *E. coli*-positive single organisms were incorporated.

Data preprocessing

Sample sequencing data underwent demultiplexing, adaptor trimming, and joining paired-end sequences using *VSEARCH* (v.2.7) [104]. Then, it was uploaded to QIIME2 software (November 2018 release) to preprocessing [105]. The *deblur* (v. 1.0.4; *q2-dublur*) was then used for quality filtering (*q2-quality-filter*) and denoising to produce an amplicon sequences variant (ASV) table with default parameters [106,107]. Additionally, a phylogenetic tree was constructed by inserting fragments into the Greengenes 99% identity tree backbone using *q2-fragment-insertion* [108]. Finally, ASVs were assigned a taxonomy using a pretrained reference with a naive Bayesian classifier (*q2-feature-classifier*). ASVs were referred using the first letter of their lowest clearly assigned taxonomic level, the first five letters of their lowest taxonomic assignment, and the first six characters of an MD5 hash of the sequences [109].

Diversity analyses were conducted on rarefied samples of 6,500 sequences using *q2-diversity* in QIIME 2, which calculated alpha diversity measures including Faith's phylogenetic diversity (Faith's PD), observed ASVs and Shannon diversity [110,111] and beta diversity including unweighted UniFrac, weighted UniFrac, and Bray-Curtis metrics [112–114].

Exposure metrics

The exposures of **Study IV** were alpha diversity and beta diversity.

Alpha diversity describes the microbiota diversity within a sample [115]. Faith's PD provides a measure of richness (i.e., the count of different sequence variants in one sample) weighted by the phylogeny (i.e., shared evolutionary history between organisms) [111] , while observed ASVs indicate richness in the sample. The Shannon diversity index is a measure of both richness and abundance (i.e., the count of each sequence variant in one sample) [110]. Alpha diversity was treated as a continuous variable or categorized into tertiles (low, medium, and high diversity).

Beta diversity emphasizes the dissimilarity between samples [115,116]. The Bray-Curtis distance measures dissimilarities in relative abundance. The weighted UniFrac distance considers relative abundance and phylogeny, with emphasis on dominant organisms, and the unweighted UniFrac distance focuses on presence/absence and phylogeny, highlighting the uncommon microbiota.

4.2.2 Outcomes

The outcome was survival probability rate, all-cause mortality and/or NPC-specific mortality. NPC patients were monitored for their vital status, causes of death, and migration out of the study areas using a passive-active-passive circle strategy, which was adapted according to the practical situations of the regions (**Figure 4.5**). The circle strategy began with passive follow-up, which involved linking to databases such as the Cancer Registry, Medical Records, Causes of Death Registry, Medical

insurance system, and Population registry. If patients were lost to follow-up or had passed away without a known cause, we utilized active follow-up methods, such as contacting patients' relatives or the patients themselves, or conducting home visits by village doctors. Finally, we performed passive and active follow-ups repeatedly to complete the information of remaining cases (**Figure 4.5**).

The follow-up period started from diagnosis until either the 31st of December 2018, migration, or death, whichever occurred first. For two cases of death without a death date, we used the median survival time of cases of death with corresponding areas, age and cancer stage. The follow-up time for seven lost-to-follow-up patients (i.e., follow-up time=0) was set to half a day.

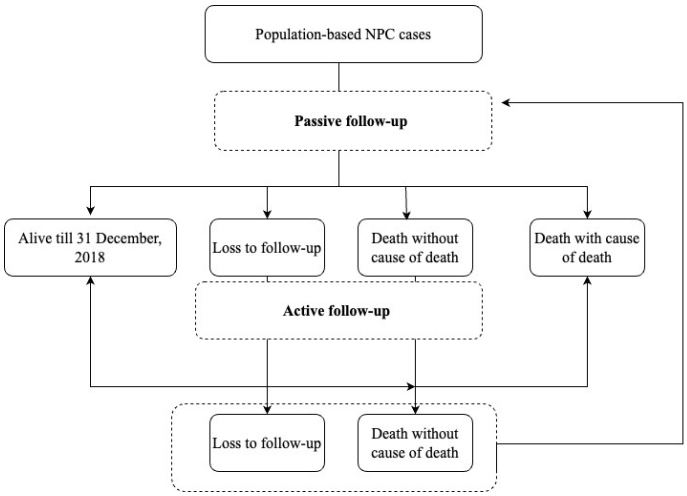


Figure 4.5 Strategy of NPCGEE follow-up. NPC, nasopharyngeal carcinoma.

4.2.3 Covariates

Covariate information was collected through questionnaires or medical records.

Education level at diagnosis was classified into four groups: illiterate/primary school, middle school, high school, and vocational or technical college/university and above [117]. Tobacco use and alcohol consumption were categorized as never, former, or current. Occupation at diagnosis was classified as farmer, blue collar, white collar, unemployed, or unknown/others.

NPC cases were treated in more than 20 hospitals in southern China. We grouped the hospitals into two categories: medical university-affiliated or province-level hospitals and prefecture-level hospitals. Karnofsky Performance Scale (KPS) was categorized into two levels (<90 and ⩾90), which determined the patient's ability to carry out regular activities [118]. The clinical TNM cancer stage was defined based on the 7th version of the AJCC staging system [119], while the pathological classification included non-keratinizing carcinoma and others according to the 2005 WHO tumour classification [35]. NPC therapy is radiotherapy-based, combined with or without chemotherapy; therefore, we categorized the treatment pattern into six groups according to clinical routine: concurrent chemoradiotherapy (CCRT), CCRT with adjuvant chemotherapy (ACT) and/or induction chemotherapy (ICT), only radiotherapy, only chemotherapy, neither radiotherapy nor chemotherapy and missing. Radiotherapy techniques included 2DRT, 3DRT and IMRT.

The selection of covariates to be included in the models followed common principles. First, we selected covariates with prior knowledge and identified covariates by examining their statistically significant univariate associations with mortality. To ensure there was no multicollinearity, we calculated a variance inflation factor and excluded variables that had a variance inflation factor greater than 10 [120].

4.3 Statistical analysis

4.3.1 Cox regression

Cox regression is a semi-parametric model for analyzing survival data without assuming any specific distribution of survival times. The basic assumption is proportional hazards over time, which is testable using the cumulative hazards plots or Schoenfeld's residuals[121,122]. In **Studies II, III, and IV**, we employed Cox regression models to estimate hazards ratios (HRs) for all-cause and NPC-specific mortality in relation to various exposures. To assess potential nonlinear dose-response associations, we used restricted cubic spline functions to visualize the HRs along the continuous exposure (**Studies II and III**) [123].

4.3.2 Flexible parametric model

Flexible parametric survival models utilize a smooth function to represent the transformation of survival. The Royston-Parmar model was used in this thesis [124,125]. In **Study I**, we performed flexible parametric models to calculate the avoidable deaths. In **Study III**, we utilized flexible parametric models to evaluate the time-dependent associations between plasma EBV DNA and mortality.

4.3.3 Permutational multivariate analysis of variance

Permutational multivariate analysis of variance (PERMANOVA) is a statistical method that uses distance matrices to fit linear models and partition distance matrix among sources of variation. PERMANOVA does not assume distributions of the variables or the dissimilarities in the matrix [126]. In **Study IV**, beta diversity was compared by PERMANOVA between groups, controlling for sex, age and sequencing plates, with 999 permutations.

4.3.4 Principal coordinates analysis

Principal coordinates analysis (PCoA) is an ordination technique used for analyzing and visualizing the similarities and differences among a set of objects. It involves transforming complex data, i.e., high-dimensional data, into a simpler format by creating a set of new variables, called principal coordinates (PCs), which represent the most important sources of variation in the data [127]. In **Study IV**, PCoA was employed to visualize the subjects and the top three PCs corresponding to microbiome pattern were included as covariates in adjusted Cox regression model.

4.3.5 Robust Aichison principal component analysis

Robust Aitchison principal component analysis (RPCA) is a statistical method used in microbiome studies to analyze and compare the composition of microbial communities across different samples. It accounts for the sparse composition and zero-flat of microbiota data. It contains two procedures, i.e.,

robust centered log-ratio transformation of absolute abundance of features and matrix completion of treating all zeros as missing values and imputation [128]. In **Study IV**, we used RPCA to depict and illustrate beta diversity. The top three PCs were incorporated into Cox regression models to evaluate the association between beta diversity and mortality.

4.4 Ethical considerations

This thesis relied on the NPCGEE project, approved by institutional review boards from various organizations, including Harvard T.H. Chan School of Public Health, 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. The ethical application for the follow-up of NPC patients enrolled in the NPCGEE study was further approved by the Regional Ethical Review Board in Stockholm, Sweden, Guangxi Medical University, Wuzhou Red Cross Hospital and Sun Yat-sen University Cancer Center, China. All participants provided written or oral informed consent during the interview.

5 Results

5.1 Population-based survival in southern China

Study I comprised 2529 newly diagnosed patients with NPC, with 1300 from Zhaoqing, 546 cases from Guiping & Pingnan, and 683 cases from Wuzhou. The average duration of follow-up was 5.50 years, and only 1.7% of the patients were lost to follow-up. A total of 11.3% of the patients in this cohort had early-stage (I-II) NPC. Among the patients, the majority (72.1%) were treated at prefecture-level hospitals (**Table 5.1**).

Table 5.1 Characteristics of primary NPC cases by geographic areas during 2010-2013 in Southern China

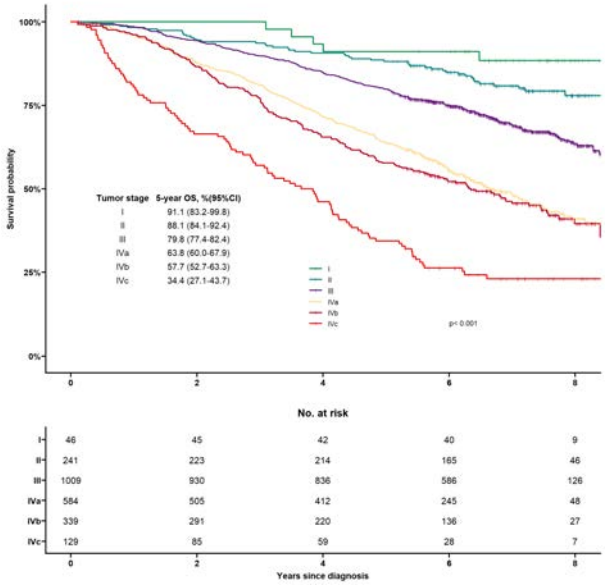
Characteristics	Zhaoqing (N=1300)	Guiping&Pingnan (N=546)	Wuzhou (N=683)	Total (N=2529)
Follow-up years				
Mean (SD)	5.67 (2.32)	5.19 (2.20)	5.42 (2.23)	5.50 (2.27)
Age at cancer diagnosis				
Mean (SD)	47.9 (10.7)	49.7 (10.9)	48.8 (10.4)	48.5 (10.7)
Vital status at end of follow-up, n (%)				
Alive	786 (60.5%)	309 (56.6%)	371 (54.3%)	1466 (58.0%)
Deceased	481 (37.0%)	228 (41.8%)	310 (45.4%)	1019 (40.3%)
Lost to follow-up	33 (2.5%)	9 (1.6%)	2 (0.3%)	44 (1.7%)
Sex, n (%)				
Female	349 (26.8%)	142 (26.0%)	182 (26.6%)	673 (26.6%)
Male	951 (73.2%)	404 (74.0%)	501 (73.4%)	1856 (73.4%)
Marital status at diagnosis, n (%)				
Married	1221 (93.9%)	514 (94.1%)	649 (95.0%)	2384 (94.3%)
Not married	79 (6.1%)	32 (5.9%)	34 (5.0%)	145 (5.7%)
Educational attainment, n (%)				
Illiterate/Primary school	478 (36.8%)	241 (44.1%)	286 (41.9%)	1005 (39.7%)
Middle school	558 (42.9%)	210 (38.5%)	243 (35.6%)	1011 (40.0%)
High school	204 (15.7%)	76 (13.9%)	127 (18.6%)	407 (16.1%)
Vocational or technical college/University and above	60 (4.6%)	19 (3.5%)	27 (4.0%)	106 (4.2%)
Occupation at diagnosis, n (%)				
Farmer	388 (29.8%)	246 (45.1%)	218 (31.9%)	852 (33.7%)
Blue collar	570 (43.8%)	179 (32.8%)	273 (40.0%)	1022 (40.4%)
White collar	195 (15.0%)	52 (9.5%)	103 (15.1%)	350 (13.8%)
Unemployed	43 (3.3%)	16 (2.9%)	18 (2.6%)	77 (3.0%)
Unknown/other	104 (8.0%)	53 (9.7%)	71 (10.4%)	228 (9.0%)
Smoking history at diagnosis, n (%)				
Never	545 (41.9%)	254 (46.5%)	317 (46.4%)	1116 (44.1%)
Former	111 (8.5%)	30 (5.5%)	38 (5.6%)	179 (7.1%)
Current	638 (49.1%)	262 (48.0%)	328 (48.0%)	1228 (48.6%)
Missing	6 (0.5%)	0 (0%)	0 (0%)	6 (0.2%)
Treatment hospitals, n (%)				
Medical university-affiliated/province-level	183 (14.1%)	315 (57.7%)	35 (5.1%)	533 (21.1%)
Prefecture-level	1026 (78.9%)	168 (30.8%)	629 (92.1%)	1823 (72.1%)
Missing	91 (7.0%)	63 (11.5%)	19 (2.8%)	173 (6.8%)
BMI before treatment, n (%)				
Normal weight	616 (47.4%)	233 (42.7%)	352 (51.5%)	1201 (47.5%)
Underweight	175 (13.5%)	45 (8.2%)	98 (14.3%)	318 (12.6%)
Overweight	335 (25.8%)	138 (25.3%)	181 (26.5%)	654 (25.9%)
Obese	63 (4.8%)	27 (4.9%)	30 (4.4%)	120 (4.7%)
Missing	111 (8.5%)	103 (18.9%)	22 (3.2%)	236 (9.3%)
KPS before treatment, n (%)				
< 90	170 (13.1%)	46 (8.4%)	96 (14.1%)	312 (12.3%)
≥ 90	1029 (79.2%)	392 (71.8%)	560 (82.0%)	1981 (78.3%)
Missing	101 (7.8%)	108 (19.8%)	27 (4.0%)	236 (9.3%)
Histological type, n (%)				
Others	39 (3.0%)	30 (5.5%)	22 (3.2%)	91 (3.6%)
Non-keratinizing carcinoma	1171 (90.1%)	457 (83.7%)	635 (93.0%)	2263 (89.5%)
Missing	90 (6.9%)	59 (10.8%)	26 (3.8%)	175 (6.9%)
Cancer stage, n (%)				

Characteristics	Zhaoqing (N=1300)	Guiping&Pingnan (N=546)	Wuzhou (N=683)	Total (N=2529)
I	19 (1.5%)	14 (2.6%)	13 (1.9%)	46 (1.8%)
II	154 (11.8%)	44 (8.1%)	43 (6.3%)	241 (9.5%)
III	530 (40.8%)	205 (37.5%)	274 (40.1%)	1009 (39.9%)
IVa	288 (22.2%)	113 (20.7%)	183 (26.8%)	584 (23.1%)
IVb	154 (11.8%)	70 (12.8%)	115 (16.8%)	339 (13.4%)
IVc	62 (4.8%)	33 (6.0%)	34 (5.0%)	129 (5.1%)
Missing	93 (7.2%)	67 (12.3%)	21 (3.1%)	181 (7.2%)
Treatment pattern, n (%)				
CCRT	484 (37.2%)	204 (37.4%)	351 (51.4%)	1039 (41.1%)
CCRT+ICT/ACT	561 (43.2%)	181 (33.2%)	215 (31.5%)	957 (37.8%)
RT only	96 (7.4%)	47 (8.6%)	68 (10.0%)	211 (8.3%)
Chemo only	45 (3.5%)	13 (2.4%)	19 (2.8%)	77 (3.0%)
Neither RT nor CT	26 (2.0%)	15 (2.7%)	11 (1.6%)	52 (2.1%)
RT+ICT/ACT	0 (0%)	2 (0.4%)	0 (0%)	2 (0.1%)
Missing	88 (6.8%)	84 (15.4%)	19 (2.8%)	191 (7.6%)
RT technique, n (%)				
2DRT	729 (56.1%)	159 (29.1%)	396 (58.0%)	1284 (50.8%)
3DRT	107 (8.2%)	32 (5.9%)	1 (0.1%)	140 (5.5%)
IMRT	300 (23.1%)	248 (45.4%)	236 (34.6%)	784 (31.0%)
No RT	71 (5.5%)	28 (5.1%)	30 (4.4%)	129 (5.1%)
Unknown technique	5 (0.4%)	18 (3.3%)	1 (0.1%)	24 (0.9%)
Missing	88 (6.8%)	61 (11.2%)	19 (2.8%)	168 (6.6%)
Nasopharyngeal radiation dose, n (%)				
< 70 Gy	211 (16.2%)	55 (10.1%)	175 (25.6%)	441 (17.4%)
≥ 70 Gy	909 (69.9%)	372 (68.1%)	453 (66.3%)	1734 (68.6%)
No radiotherapy	71 (5.5%)	28 (5.1%)	30 (4.4%)	129 (5.1%)
Missing	109 (8.4%)	91 (16.7%)	25 (3.7%)	225 (8.9%)

Abbreviations: EBV, Epstein-Barr virus; SD, standard deviation; BMI, body mass index; KPS, Karnofsky performance scale; CCRT, concurrent chemoradiotherapy; ICT, induction chemotherapy; ACT, adjuvant chemotherapy; RT, radiotherapy; Chemo, chemotherapy; 2DRT, two-dimensional radiotherapy; 3DRT, three-dimensional radiotherapy; IMRT, intensity-modulated radiotherapy.

The 5-year overall survival probability was 70.1% (95% CI, 68.4%-72.0%). The 5-year overall survival was 74.1% for patients from Zhaoqing, 67.0% from Guiping & Pingnan, and 65.2% from Wuzhou. By stage at diagnosis, it was 91.1% of stage I, 88.1% of stage II, 79.8% of stage III, 63.8% of stage IVa, 57.7% of stage IVb, and 34.4% of stage IVc (**Figure 5.1A**). Patients treated at medical university-affiliated/province-level hospitals had a 5-year overall survival probability of 77.2%, while for those treated at prefecture-level hospitals, it was 69.4% (p<0.001) (**Figure 5.1B**).

A



B

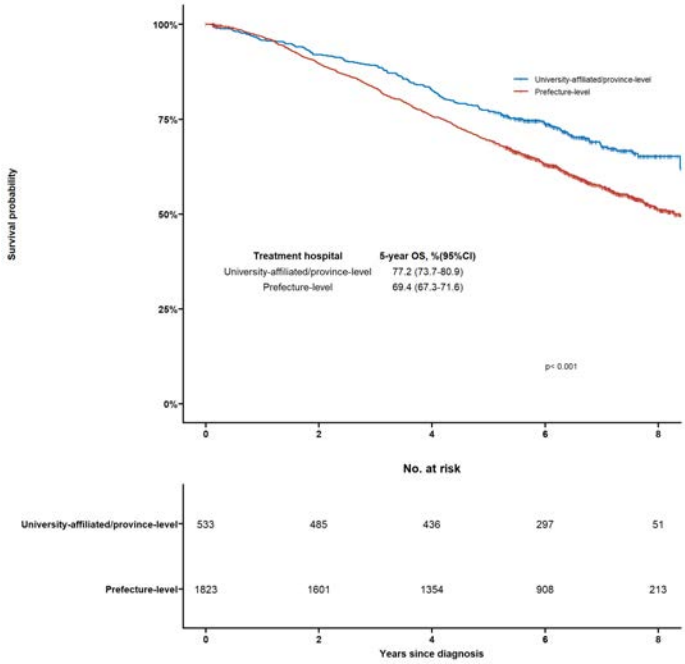


Figure 5.1 Overall survival curves of NPC cases estimated by Kaplan-Meier methods. Panel A, survival by clinical stage at diagnosis; Panel B, survival by treatment hospital type.

The most directly comparable study was performed on a hospital-based NPC cohort diagnosed between 1990 and 2012 at Sun Yat-Sen University Cancer Center, a medical university-affiliated hospital in Guangdong Province (the same province as Zhaoqing, one of our study regions) and with the largest sample size of NPC cases (20,305) ever reported in China [3]. There was a lag of approximately 10 years in the survival of our population-based NPC patients compared to those hospital-based patients. (**Figure 5.2**).

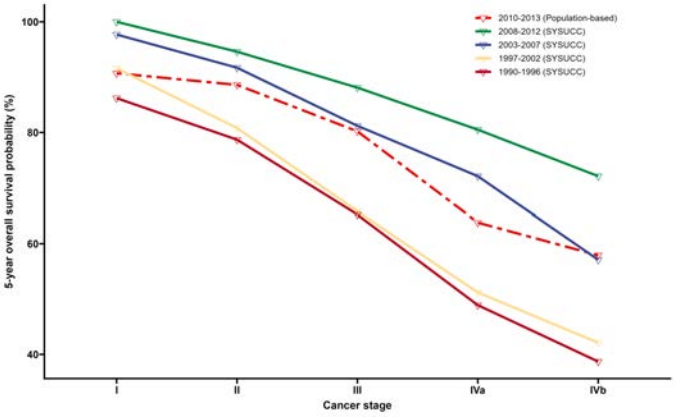


Figure 5.2 Comparison of 5-year overall survival by stage across calendar period of a large hospital-based NPC cohort in Sun Yet-Sen University Cancer Center (SYSUCC), with the results of NPCGEE.

If early diagnosis (stage I-II) was possible for all advanced-stage NPC cases (III-IVc), the standardized avoidable deaths within five years of diagnosis would be 174 (95% CI: 128-221) per

1000 patients (**Figure 5.3A**). Similarly, if all patients received treatment at medical university affiliated or province-level hospitals instead of prefecture-level hospitals, the standardized avoidable deaths within five years would be 58 (95% CI: 8-108) per 1000 patients (**Figure 5.3B**).

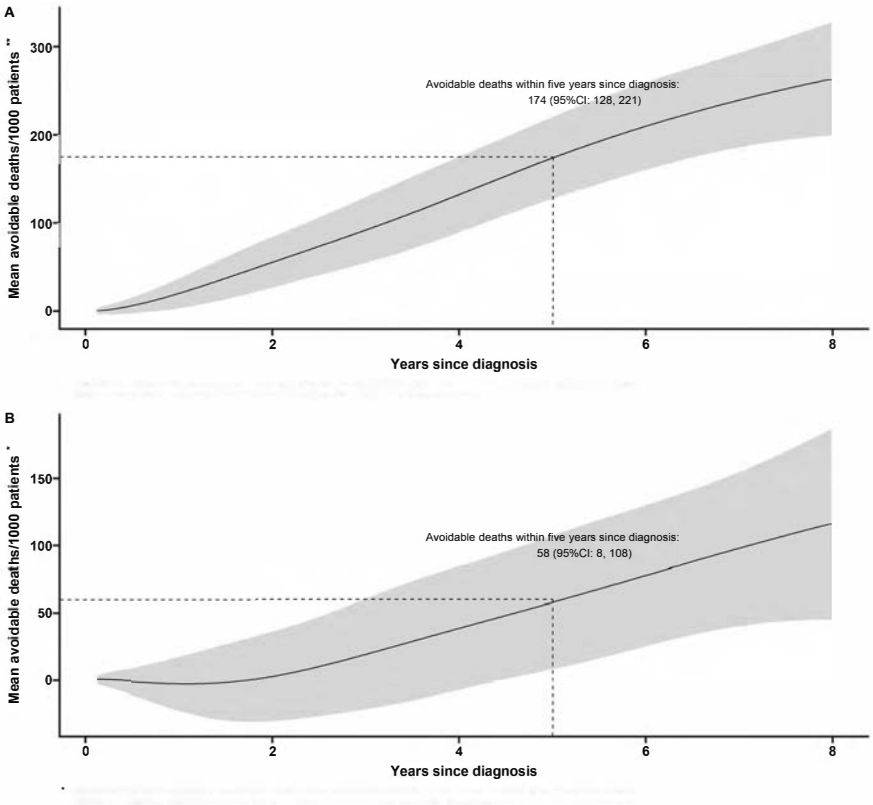


Figure 5.3 Mean avoidable deaths along years since diagnosis under different scenarios. Panel A: Mean avoidable deaths by year since diagnosis, comparing scenarios if 1000 patients were all diagnosed at early stages (I-II) vs. if all were diagnosed at advanced stages (III-IV). Panel B: Mean avoidable deaths by year since diagnosis, comparing scenarios if 1000 patients were all treated in university-affiliated or province-level hospitals vs. if all patients were treated in prefecture-level hospital.

5.2 BMI and body shape before treatment and NPC prognosis

In **Study II**, we included 2526 NPC patients, of which 316 (12.5%) were underweight, 1180 (46.7%) were normal weight, 638 (25.3%) were overweight, 115 (4.6%) were obesity and 277 (11.0%) had missing values at diagnosis.

At diagnosis, overweight and obese patients with NPC had a 25% (adjusted HR 0.75, 95% CI: 0.64-0.89) and a 28% (adjusted HR 0.72, 95% CI: 0.51-1.03) lower all-cause mortality than normal weight patients, respectively. Patients with thinner body shapes 1 or 2 at diagnosis had higher mortality rates than those with body shape 3. Both BMI and body shape at diagnosis showed a monotonic trend with all-cause mortality (**Table 5.2**).

To investigate presumable reverse causality, we conducted a stratified analysis by clinical cancer stage and observed no significant interactions. Of note, the effect size of lower BMI and thinner body shape

with higher mortality was stronger for early-stage NPC than for locally advanced or metastatic NPC. This argued against reverse causality. Patterns of associations with NPC-specific mortality were comparable to those for all-cause mortality (**Tables 5.2 and 5.3**).

Table 5.2 Hazard ratios for body mass index and body shape in association with mortality among 2526 NPC cases

	All-cause mortality				NPC-specific mortality			
	Events	Person-years	Crude HR (95%CI)	Adjusted HR (95%CI)†	Events	Person-years	Crude HR	Adjusted HR (95%CI)†
BMI at age 20 years								
Underweight	161	2234	0.96(0.81-1.14)	0.95(0.80-1.12)	111	2234	0.99(0.80-1.21)	0.97(0.79-1.19)
Normal	745	9953	ref	ref	502	9953	ref	ref
Overweight/obese	112	1701	0.87(0.72-1.07)	0.87(0.71-1.06)	88	1701	1.02(0.81-1.28)	1.02(0.81-1.28)
P for trend			0.523	0.627			0.806	0.733
BMI 10 years before diagnosis								
Underweight	102	1397	0.94(0.76-1.16)	0.93(0.76-1.15)	74	1397	1.01(0.79-1.30)	1.01(0.79-1.30)
Normal	672	8685	ref	ref	452	8685	ref	ref
Overweight	218	3287	**0.85(0.73-0.99)**	0.87(0.75-1.02)	160	3287	0.93(0.78-1.11)	0.97(0.81-1.17)
Obese	26	519	**0.63(0.43-0.94)**	0.75(0.50-1.11)	15	519	**0.55(0.33-0.91)**	0.64(0.38-1.07)
P for trend			**0.022**	0.131			0.060	0.256
BMI at diagnosis								
Underweight	152	1623	**1.23(1.03-1.48)**	1.07(0.89-1.29)	104	1623	1.21(0.97-1.51)	1.03(0.83-1.29)
Normal	490	6411	ref	ref	340	6411	ref	ref
Overweight	208	3737	**0.72(0.61-0.85)**	**0.75(0.64-0.89)**	142	3737	**0.71(0.58-0.86)**	**0.75(0.61-0.91)**
Obese	34	706	**0.62(0.44-0.88)**	0.72(0.51-1.03)	24	706	**0.63(0.42-0.96)**	0.75(0.49-1.14)
Missing	134	1412	**1.25(1.03-1.52)**	0.78(0.47-1.31)	91	1412	1.22(0.97-1.54)	0.81(0.44-1.49)
P for trend††			**< 0.001**	**< 0.001**			**< 0.001**	**0.002**
Body shape at age 20 years								
Shape 1 (thinnest)	35	462	1.06(0.75-1.49)	1.16(0.82-1.66)	22	462	1.00(0.65-1.55)	0.97(0.62-1.51)
Shape 2	379	5112	1.04(0.91-1.19)	1.08(0.94-1.24)	265	5112	1.09(0.93-1.29)	1.12(0.95-1.32)
Shape 3	448	6233	ref	ref	297	6233	ref	ref
Shape 4	125	1738	1.00(0.82-1.22)	0.93(0.76-1.14)	94	1738	1.14(0.90-1.43)	1.01(0.79-1.27)
Shape 5-9	31	343	1.26(0.87-1.81)	1.22(0.85-1.77)	23	343	1.41(0.92-2.16)	1.23(0.80-1.89)
P for trend			0.980	0.275			0.565	0.641
Body shape 10 years before diagnosis								
Shape 1 (thinnest)	27	342	1.15(0.78-1.70)	1.38(0.93-2.06)	21	342	1.30(0.84-2.03)	1.40(0.89-2.20)
Shape 2	282	3355	**1.24(1.06-1.43)**	**1.18(1.02-1.37)**	190	3355	**1.21(1.01-1.45)**	1.15(0.96-1.39)
Shape 3	447	6502	ref	ref	307	6502	ref	ref
Shape 4	181	2677	0.98(0.83-1.17)	0.92(0.78-1.10)	127	2677	1.01(0.82-1.24)	0.92(0.75-1.14)
Shape 5-9	81	1012	1.16(0.92-1.47)	1.06(0.84-1.36)	56	1012	1.17(0.88-1.55)	1.00(0.75-1.34)
P for trend			0.121	**0.025**			0.238	**0.048**
Body shape at diagnosis								
Shape 1 (thinnest)	59	476	**1.87(1.42-2.47)**	**1.68(1.26-2.23)**	45	476	**2.01(1.47-2.76)**	**1.65(1.19-2.29)**
Shape 2	314	3634	**1.29(1.11-1.51)**	**1.23(1.06-1.44)**	206	3634	**1.20(1.00-1.44)**	1.12(0.93-1.35)
Shape 3	363	5388	ref	ref	257	5388	ref	ref
Shape 4	203	3162	0.95(0.80-1.13)	0.94(0.79-1.12)	136	3162	0.90(0.73-1.11)	0.88(0.71-1.08)
Shape 5-9	79	1227	0.95(0.75-1.21)	0.96(0.75-1.24)	57	1227	0.97(0.73-1.29)	0.93(0.69-1.24)
P for trend			**< 0.001**	**< 0.001**			**< 0.001**	**0.002**

This table is adapted from Du et al [129].

Abbreviations: HR, hazard ratio; BMI, body mass index; CI, confidence interval; NPC, nasopharyngeal carcinoma. †HRs were adjusted by age, residential areas, sex, educational attainment, current occupation, smoking history, alcohol consumption, treatment hospital and KPS before treatment, and baseline hazards were stratified by cancer stage. ††Except for missing group.

Table 5.3 Hazard ratios for body mass index and body shape at diagnosis in association with all-cause and NPC-specific mortality stratified by cancer stage, among nasopharyngeal carcinoma cases during 2010-2013 in southern China

	Early				Locally advanced				Metastatic				
	Events	Person-years	Crude HR (95%CI)	Adjusted HR (95%CI)†	Events	Person-years	Crude HR (95%CI)	Adjusted HR (95%CI)†	Events	Person-years	Crude HR (95%CI)	Adjusted HR (95%CI)†	*P**
All-cause mortality													
BMI													0.466
Underweight	5	145	1.18 (0.45-3.08)	1.61 (0.56-4.59)	124	1364	1.18 (0.96-1.44)	1.09 (0.88-1.34)	23	101	1.00 (0.61-1.64)	1.16 (0.66-2.04)	
Normal	25	858	ref	ref	413	5319	ref	ref	52	222	ref	ref	
Overweight/obese	18	768	0.80 (0.44-1.47)	1.05 (0.54-2.05)	204	3546	**0.73 (0.62-0.86)**	**0.75 (0.63-0.89)**	20	123	0.70 (0.42-1.17)	0.60 (0.34-1.06)	
Body shape													0.398
Shapes 1-2 (thinnest)	16	455	1.41 (0.71-2.79)	1.45 (0.69-3.07)	253	3024	**1.25 (1.05-1.48)**	**1.28 (1.07-1.52)**	47	184	1.31 (0.83-2.06)	1.36 (0.80-2.28)	
Shape 3	17	671	ref	ref	282	4170	ref	ref	32	157	ref	ref	
Shape 4-9	15	662	0.89 (0.44-1.78)	0.98 (0.46-2.08)	218	3208	1.00 (0.84-1.19)	1.00 (0.84-1.20)	18	131	0.68 (0.38-1.21)	**0.46 (0.24-0.89)**	
NPC-specific mortality													
BMI													0.408
Underweight	2	145	0.66 (0.15-2.83)	0.78 (0.17-3.60)	83	1364	1.13 (0.88-1.44)	1.04 (0.81-1.33)	19	101	1.28 (0.73-2.25)	1.51 (0.80-2.87)	
Normal	18	858	ref	ref	288	5319	ref	ref	34	222	ref	ref	
Overweight/obese	13	768	0.81 (0.39-1.65)	1.00 (0.45-2.21)	135	3546	**0.69 (0.57-0.85)**	**0.70 (0.57-0.87)**	18	123	0.97 (0.55-1.73)	0.93 (0.50-1.74)	
Body shape													0.821
Shapes 1-2 (thinnest)	10	455	1.32 (0.56-3.12)	1.37 (0.53-3.53)	166	3024	1.13 (0.92-1.39)	1.15 (0.93-1.42)	35	184	1.35 (0.80-2.29)	1.46 (0.80-2.66)	
Shape 3	11	671	ref	ref	204	4170	ref	ref	23	157	ref	ref	
Shape 4-9	12	662	1.10 (0.49-2.50)	1.12 (0.46-2.74)	145	3208	0.92 (0.74-1.14)	0.89 (0.72-1.11)	14	131	0.74 (0.38-1.44)	0.57 (0.27-1.21)	

This table is adapted from Du et al [129].

Abbreviations: HR, hazard ratio; BMI, body mass index; CI, confidence interval; KPS, Karnofsky performance scale.

†HR was adjusted by age at cancer diagnosis, residential areas, sex, educational attainment, current occupation, smoking history, alcohol consumption, treatment hospital and KPS before treatment.

**P* for interaction

5.3 Pretreatment plasma EBV DNA and NPC prognosis

Of the 1854 NPC cases in **Study III**, 1673 (90.2%) had detectable plasma EBV DNA at diagnosis. Those individuals with detectable plasma EBV DNA were more likely to have lower education levels, be underweight, and have been diagnosed with advanced cancer stages.

Comparing those with detectable plasma EBV DNA to those with undetectable EBV DNA, the multivariate adjusted HRs for all-cause mortality and NPC-specific mortality were 2.16 (95% CI: 1.54-3.02) and 2.03 (95% CI: 1.37-3.02), respectively (**Table 5.4**).

Table 5.4 Hazard ratios of pretreatment plasma EBV DNA on mortality among 1854 nasopharyngeal carcinoma cases during 2010-2013 in southern China

Pretreatment Plasma EBV DNA	Person-years	All-cause mortality			NPC-specific mortality		
		Events	Crude HR (95%CI)	Adjusted HR (95%CI)†	Events	Crude HR (95%CI)	Adjusted HR (95%CI)†
Undetectable	1119	37	ref	ref	27	ref	ref
Detectable	8851	744	**2.59(1.86-3.61)**	**2.16(1.54-3.02)**	506	**2.41(1.63-3.55)**	**2.03(1.37-3.02)**
Low	3317	167	**1.53(1.07-2.19)**	**1.45(1.01-2.09)**	114	1.43(0.94-2.18)	1.40(0.91-2.14)
Medium	2987	242	**2.50(1.77-3.54)**	**2.22(1.56-3.17)**	158	**2.23(1.48-3.35)**	**1.99(1.31-3.03)**
High	2547	335	**4.16(2.96-5.84)**	**3.04(2.14-4.32)**	234	**3.95(2.65-5.88)**	**2.88(1.91-4.35)**

Abbreviations: NPC, nasopharyngeal carcinoma; HR, hazard ratio; BMI, body mass index; CI, confidence interval; EBV, Epstein-Barr virus.
†HR was adjusted by continuous age at cancer diagnosis, residential areas, sex, educational attainment, occupation at recruitment, smoking history, alcohol consumption, KPS before treatment, treatment hospital, pathological type and cancer stage, and baseline hazards were stratified by treatment pattern.

Treating plasma EBV DNA as a log-transformed continuous variable, the adjusted HRs per unit of increase for all-cause and NPC-specific mortality were 1.14 (95% CI: 1.11-1.17) and 1.14 (95% CI: 1.10-1.18), respectively. Furthermore, the restricted cubic spline models demonstrated that there were strong monotonic dose-response associations between pretreatment plasma EBV DNA levels and mortality (**Figure 5.4**).

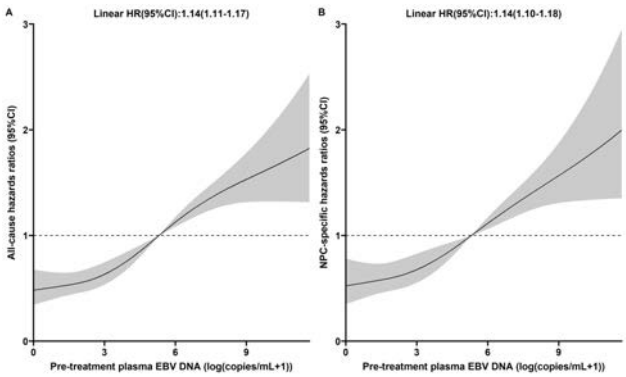


Figure 5.4 Dose-response associations between pretreatment plasma EBV DNA load with logarithm-transformation and all-cause mortality (Panel A) and NPC-specific mortality (Panel B).
Hazard ratios were adjusted for continuous age at cancer diagnosis, residential area, sex, educational attainment, current occupation, smoking history, alcohol consumption, BMI at diagnosis, KPS before treatment, pathological type and cancer stage at diagnosis, and baseline hazards were stratified by treatment pattern.
Abbreviations: BMI, body mass index; NPC, nasopharyngeal carcinoma; HR, hazard ratio; CI, confidence interval.

Five years after diagnosis, the relationships between plasma EBV DNA and mortality became weaker (**Figure 5.5**).

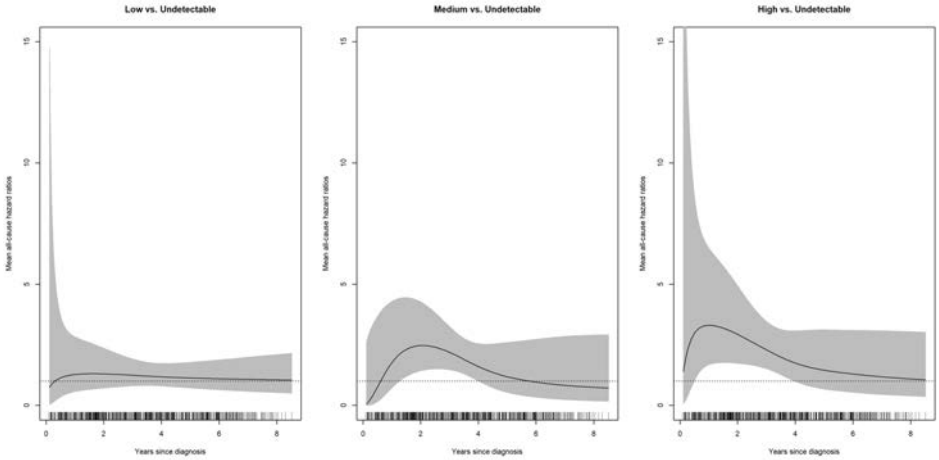


Figure 5.5 Time-dependent associations of plasma EBV DNA on mean all-cause mortality.
Hazard ratios were adjusted for continuous age at cancer diagnosis, residential area, sex, educational attainment, current occupation, smoking history, alcohol consumption, BMI at diagnosis, KPS before treatment, and cancer stage at diagnosis, considering effect modification of treatment pattern, with four degrees of freedoms.
Abbreviation: NPC, nasopharyngeal carcinoma.

5.4 Saliva microbiota and NPC prognosis

5.4.1 Characteristics of the study population

Study IV involved 482 patients with NPC. Among them, 71% were male, and 93% were diagnosed at stage III/IV. The average follow-up time was 5.29 years. With respect to the risk factors for NPC, 78% had less than 9 years of education, 66% brushed their teeth once or less per day, and 52% were current smokers at the time of the interview (**Table 5.5**).

Table 5.5 Characteristics of NPC cases and univariate associations between covariates and mortality

Characteristics	Total, n(%)[a]	Deaths, n(%)[a]	All-cause HRs, (95%CI)	Deaths from NPC, n(%)[a]	NPC-specific HRs, (95%CI)
Number of cases	482 (100.0)	210 (43.6)		181 (37.6)	
Mean follow-up yeas (SD)	5.29 (2.07)				
Mean age at diagnosis (SD)	48.45 (10.55)		**1.03 (1.01,1.04)**		**1.02 (1.00,1.03)**
Sex					
Male	342 (71.0)	165 (78.6)	ref	144 (79.6)	ref
Female	140 (29.0)	45 (21.4)	**0.58 (0.42,0.81)**	37 (20.4)	**0.55 (0.38,0.79)**
Residential community					
Wuzhou	108 (22.4)	41 (19.5)	ref	36 (19.9)	ref
Cangwu	115 (23.9)	51 (24.3)	1.26 (0.83,1.89)	46 (25.4)	1.29 (0.83,1.99)
Cenxi	165 (34.2)	79 (37.6)	1.45 (0.99,2.11)	63 (34.8)	1.31 (0.87,1.97)
Tengxian	94 (19.5)	39 (18.6)	1.17 (0.76,1.82)	36 (19.9)	1.23 (0.78,1.96)
Educational attainment					
≤ 6 years	204 (42.3)	95 (45.2)	ref	75 (41.4)	ref
7-9 years	170 (35.3)	77 (36.7)	0.96 (0.71,1.30)	70 (38.7)	1.11 (0.80,1.54)
≥ 10 years	108 (22.4)	38 (18.1)	0.72 (0.49,1.04)	36 (19.9)	0.86 (0.58,1.28)
Tobacco use					
Never	230 (47.7)	83 (39.5)	ref	67 (37.0)	ref
Current	252 (52.3)	127 (60.5)	**1.56 (1.18,2.06)**	114 (63.0)	**1.73 (1.28,2.34)**
Diagnosis calendar year					
2011	224 (46.5)	115 (54.8)	ref	96 (53.0)	ref
2012	149 (30.9)	55 (26.2)	0.75 (0.54,1.04)	53 (29.3)	0.87 (0.62,1.23)
2013	109 (22.6)	40 (19.0)	0.88 (0.61,1.26)	32 (17.7)	0.83 (0.55,1.25)
Season of saliva sampling					
Winter	114 (23.7)	51 (24.3)	ref	43 (23.8)	ref
Spring	141 (29.3)	60 (28.6)	0.98 (0.67,1.42)	50 (27.6)	0.97 (0.64,1.46)
Summer	101 (21.0)	45 (21.4)	1.03 (0.69,1.54)	40 (22.1)	1.09 (0.71,1.68)
Autumn	126 (26.1)	54 (25.7)	0.99 (0.67,1.45)	48 (26.5)	1.04 (0.69,1.58)
Tooth brushing frequency					
≤ 1/day	316 (65.6)	140 (66.7)	ref	119 (65.7)	ref
≥ 2/day	166 (34.4)	70 (33.3)	0.89 (0.67,1.18)	62 (34.3)	0.93 (0.68,1.26)
Missing or filled teeth					
0	209 (43.4)	77 (36.7)	ref	70 (38.7)	ref
1	61 (12.7)	25 (11.9)	1.15 (0.74,1.81)	22 (12.2)	1.12 (0.69,1.80)
2	52 (10.8)	21 (10.0)	1.13 (0.70,1.83)	16 (8.8)	0.95 (0.55,1.63)
3-5	79 (16.4)	39 (18.6)	**1.48 (1.00,2.17)**	36 (19.9)	**1.49 (1.00,2.23)**
6+	81 (16.8)	48 (22.9)	**1.87 (1.31,2.69)**	37 (20.4)	**1.58 (1.06,2.35)**
Cancer stage					
I-II	35 (7.3)	5 (2.4)	ref	4 (2.2)	ref
III	209 (43.4)	65 (31.0)	**2.51 (1.01,6.23)**	54 (29.8)	2.60 (0.94,7.19)
IV	238 (49.4)	140 (66.7)	**6.06 (2.48,14.81)**	123 (68.0)	**6.61 (2.44,17.92)**
Treatment regimen					
CCRT	250 (51.9)	103 (49.0)	ref	88 (48.6)	ref
CCRT+ICT/ACT	150 (31.1)	66 (31.4)	1.14 (0.84,1.56)	60 (33.1)	1.21 (0.87,1.69)
RT only	60 (12.4)	28 (13.3)	1.22 (0.80,1.85)	20 (11.0)	1.02 (0.63,1.66)
No RT	22 (4.6)	13 (6.2)	**2.01 (1.12,3.57)**	13 (7.2)	**2.33 (1.30,4.17)**
BMI before treatment					
Normal weight	257 (53.3)	130 (61.9)	ref	114 (63.0)	ref
Underweight	62 (12.9)	26 (12.4)	0.78 (0.51,1.19)	23 (12.7)	0.79 (0.50,1.24)
Overweight	93 (19.3)	36 (17.1)	**0.65 (0.45,0.95)**	29 (16.0)	**0.60 (0.40,0.90)**
Obese	70 (14.5)	18 (8.6)	**0.43 (0.26,0.70)**	15 (8.3)	**0.41 (0.24,0.70)**
History of alcohol use					
Never	330 (68.5)	134 (63.8)	ref	114 (63.0)	ref
Former	20 (4.1)	13 (6.2)	**1.97 (1.11,3.48)**	10 (5.5)	1.77 (0.93,3.39)
Current	132 (27.4)	63 (30.0)	1.24 (0.92,1.68)	57 (31.5)	1.32 (0.96,1.82)
Radiotherapy technique					
No radiotherapy	22 (4.6)	13 (6.2)	ref	13 (7.2)	ref
2DRT/3DRT	266 (55.2)	141 (67.1)	0.67 (0.38,1.19)	119 (65.7)	0.57 (0.32,1.02)
IMRT	194 (40.2)	56 (26.7)	**0.36 (0.19,0.65)**	49 (27.1)	**0.31 (0.17,0.58)**
NP radiation dose					
< 70 Gy	144 (30.1)	41 (19.8)	ref	38 (21.2)	ref
≥ 70 Gy	312 (65.3)	153 (73.9)	**1.73 (1.22,2.44)**	128 (71.5)	**1.56 (1.09,2.24)**
No radiotherapy	22 (4.6)	13 (6.3)	**2.83 (1.52,5.29)**	13 (7.3)	**3.03 (1.61,5.69)**

This table is adapted from Du et al [130].

Abbreviations: HRs, hazard ratios; SD, standard deviation; BMI, body mass index; CCRT, concurrent chemoradiotherapy; ICT, induction chemotherapy; ACT, adjuvant chemotherapy; RT, radiotherapy; IMRT, intensity-modulated radiation therapy; 2DRT, conventional 2D radiotherapy; 3DRT, conventional 3D radiotherapy, NP, Nasopharyngeal.

[a]Percentages may not be 100 because of rounding.

5.4.2 Alpha diversity and NPC prognosis

Patients with a lower Faith's PD, had significantly higher all-cause and NPC-specific mortality in unadjusted and adjusted models, compared to the medium diversity group. The adjusted hazard ratios were 1.52 (95% CI, 1.06–2.17) and 1.57 (95% CI, 1.07–2.29) for all-cause and NPC-specific mortality, respectively. The observed ASVs demonstrated a comparable relationship with mortality as compared to Faith's PD, although the associations were not statistically significant in adjusted models. There was no correlation found between Shannon diversity and either outcome (**Table 5.6**).

Table 5.6 Hazard ratios (HRs) for mortality of NPC cases in relation to alpha diversity, Cox regression models

			All-cause HRs (95%CI)			NPC-specific HRs (95%CI)	
Alpha diversity	Cases (n=482)	Deaths (n=210)	Crude	Adjusted[a]	Deaths from NPC (n=181)	Crude	Adjusted[a]
Faith's PD							
Low diversity	161	83	**1.62 (1.16,2.27)**	**1.52 (1.06,2.17)**	74	**1.64 (1.15,2.33)**	**1.57 (1.07,2.29)**
Medium diversity	161	59	ref	ref	52	ref	ref
High diversity	160	68	1.24 (0.88,1.76)	1.18 (0.82,1.72)	55	1.14 (0.78,1.66)	1.10 (0.73,1.64)
Observed ASVs							
Low diversity	161	77	**1.44 (1.03,2.02)**	**1.45 (1.01,2.10)**	68	**1.47 (1.02,2.10)**	1.44 (0.97,2.12)
Medium diversity	161	61	ref	ref	53	ref	ref
High diversity	160	72	1.30 (0.92,1.83)	1.27 (0.88,1.84)	60	1.25 (0.86,1.80)	1.24 (0.83,1.83)
Shannon							
Low diversity	161	73	1.05 (0.75,1.45)	1.07 (0.75,1.52)	65	1.08 (0.76,1.53)	1.14 (0.78,1.66)
Medium diversity	161	71	ref	ref	61	ref	ref
High diversity	160	66	0.94 (0.67,1.31)	0.96 (0.68,1.36)	55	0.91 (0.63,1.31)	0.94 (0.64,1.37)

This table is adapted from Du et al [130].
Abbreviations: Faith's PD, Faith's phylogenetic diversity.
aHRs were adjusted for age, sex, smoking history, BMI before treatment, cancer stage, treatment pattern, alcohol consumption, the number of missing or filled teeth, sequence running number, residential community and season of saliva sampling.

5.4.3 Beta diversity and NPC prognosis

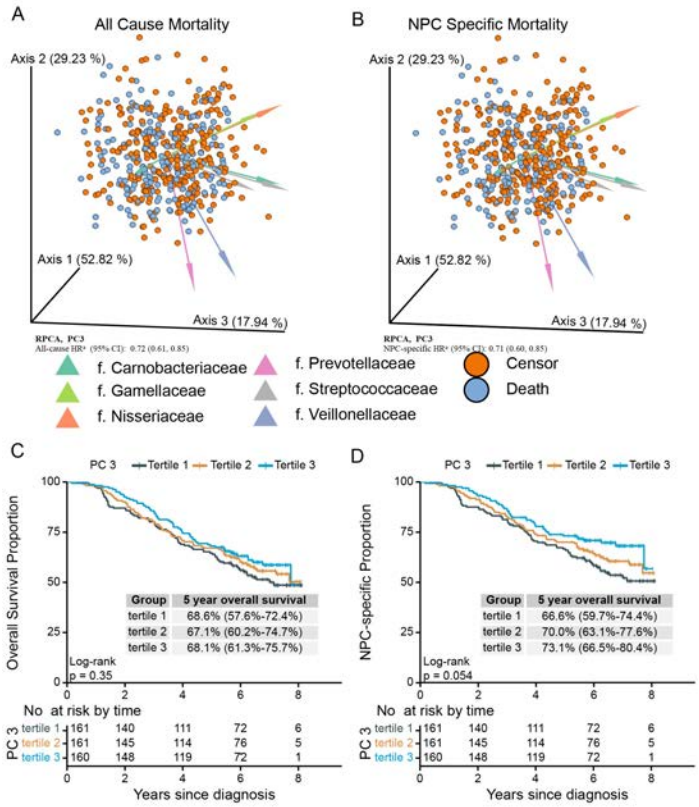
In the context of comparing nested models with and without PCs derived from PCoA, it was observed that nested models utilizing PC3 based on Bray-Curtis distance and PC1 based on weighted UniFrac distance demonstrated marginal significant associations with all-cause mortality (**Table 5.7**).

Table 5.7 Analysis of deviance table for nested with versus without top three PCoA coordinates separately

			All-cause mortality		NPC-specific mortality	
Beta diversity	PCs	Proportion of variation %	p^a	p (alpha)[b]	p^a	p (alpha)[b]
Bray curtis	PC1	10.85	0.456	0.468	0.681	0.562
	PC2	8.15	0.544	0.698	0.427	0.531
	PC3	5.17	0.061	**0.041**	0.132	0.178
Unweighted UniFrac	PC1	20.98	0.104	0.442	0.049	0.545
	PC2	5.99	0.681	0.683	0.984	0.586
	PC3	3.99	0.382	0.430	0.416	0.492
Weighted UniFrac	PC1	32.83	0.102	**0.053**	0.171	0.186
	PC2	24.89	0.434	0.407	0.940	0.580
	PC3	6.97	0.294	0.530	0.205	0.486

This table is adapted from Du et al [130].
p values were generated using likelihood ratio tests to compare nested models.
Abbreviations: PCs, principal coordinates; PCoA coordinates; Faith's PD, Faith's phylogenetic diversity.
[a]Adjusted for age, sex, sequencing plates, tobacco use, the number of missing or repaired tooth, cancer stage, treatment pattern, saliva sampling season, BMI before treatments, alcohol use, diagnosis calendar year and residential community.
[b] additionally adjusted for adjusted for Faith's PD.

As Bray-Curtis and weighted UniFrac mainly indicate relative abundance, we employed RPCA, which considers the abundance of taxa, to generate alternative beta diversity metrics. Normalized PC3 obtained through RPCA exhibited a significant correlation with both all-cause mortality (adjusted HR, 0.72, 95% CI: 0.61–0.85) and mortality specific to NPC (adjusted HR, 0.71, 95% CI: 0.60-0.85). However, we did not observe a significant correlation between PC1 or PC2 derived from RPCA and either of the outcomes. To strengthen the reliability of the associations between these PCs and the outcomes, we classified them into tertiles, plotted Kaplan-Meier curves, and developed multivariate Cox models. The analysis indicated that tertile 3 of PC3 had a 47% lower all-cause mortality and 51% lower NPC-specific mortality than tertile 1. Conversely, the associations with PC1 and PC2 remained statistically nonsignificant after adjusting for confounding factors (**Figure 5.6 and Table 5.8**).



This figure is adapted from Du et al [130].

Figure 5.6 Biplots of RPCA by survival status (A) and NPC-specific survival status (B) and Kaplan–Meier curves of overall (C) and NPC-specific survival (D) proportion between tertiled PC3 groups generated from RPCA. Arrows in (A) and (B) were top 8 taxa influencing the principal component axis. Axis1, axis2, and axis3 were equal to PC1, PC2, and PC3. The axes were labeled with the variation proportion that PCs explain. Sample loadings PC3 were z-normalized in Cox models. HRs were adjusted for age at diagnosis, sex, sequencing plates, tobacco use, the number of missing or filled tooth, cancer stage, treatment pattern, saliva sampling season, BMI before treatments, alcohol use, diagnosis calendar year, residential community and Faith's PD. RPCA, robust Aitchison principal-component analysis; NPC, nasopharyngeal carcinoma; Faith's PD, Faith's phylogenetic diversity.

Table 5.8 Hazard ratios (HRs) for mortality of NPC cases in relation to tertiled PCs from RPCA, Cox regression models

			All-cause HRs			NPC-specific HRs	
PCs[a]	Cases (n=482)	Deaths (n=210)	Crude	Adjusted[a]	Deaths of NPC (n=181)	Crude	Adjusted[b]
PC1							
tertile 1	161	63	ref	ref	51	ref	ref
tertile 2	161	81	**1.44 (1.04,2.00)**	1.21 (0.84,1.74)	73	**1.60 (1.12,2.29)**	1.31 (0.89,1.94)
tertile 3	160	66	1.13 (0.80,1.59)	1.30 (0.88,1.93)	57	1.20 (0.82,1.75)	1.32 (0.86,2.02)
PC2							
tertile 1	161	69	ref	ref	64	ref	ref
tertile 2	161	75	1.15 (0.83,1.60)	1.07 (0.75,1.54)	61	1.01 (0.71,1.43)	1.01 (0.68,1.48)
tertile 3	160	66	0.99 (0.71,1.39)	1.14 (0.78,1.67)	56	0.90 (0.63,1.29)	1.02 (0.68,1.52)
PC3							
tertile 1	161	77	ref	ref	72	ref	ref
tertile 2	161	70	0.88 (0.64,1.22)	**0.66 (0.46,0.96)**	61	0.82 (0.59,1.16)	**0.68 (0.46,1.00)**
tertile 3	160	63	0.78 (0.56,1.09)	**0.53 (0.36,0.80)**	48	**0.64 (0.44,0.92)**	**0.49 (0.32,0.76)**

This table is adapted from Du et al [130].

Abbreviation: RPCA, robust Aitchison principal-component analysis. PC, principle component.

[a] Sample loading of PCs were grouped into three tertiles.

[b]Adjusted for age at diagnosis, sex, sequencing running number, tobacco use, the number of missing or filled tooth, cancer stage, BMI before treatments, alcohol use, diagnosis calendar year, treatment pattern, saliva sampling season, residential community and Faith's phylogenetic diversity.

6 Discussion

6.1 Interpretation of findings and implications

6.1.1 What is population-based NPC survival in China?

Study I aimed to establish an effective follow-up strategy in a population-based setting in southern China. It would enable the accurate estimation of NPC survival probabilities. We also aimed to compare the results with those from other NPC-endemic areas, and to calculate avoidable deaths through early detection or treatment in advanced hospitals. We achieved a high follow-up rate of 98.3% using a passive-active-passive circle strategy that combined data linkages with direct outreach to ascertain vital status. Our population-based cohort consisted mostly of advanced-stage patients treated in prefecture-level hospitals between 2010 and 2013, and we found a 5-year overall survival rate of 70.1%, similar to that reported a decade earlier in a large cancer center in southern China. We estimated that early diagnosis could prevent 174 deaths per 1000 patients within five years, while treatment at university-affiliated or province-level hospitals could save 58 lives per 1000 patients.

During the late 20th and early 21st centuries, the development of linear accelerator technology in China enabled the use of 3DRT and IMRT, providing significant therapeutic advantages for NPC patients [131]. Our study found that 36.5% of the population received treatment with 3DRT and IMRT. Due to insufficient information on radiotherapy techniques provided in prior population-based studies [4,132], it is not possible to make direct comparisons with our findings. However, we hypothesize that the observed improvement in survival rates over time in those studies was likely influenced by the development of radiation treatment facilities, particularly around the early 21st century. Additionally, within our study population, the survival rate in Zhaoqing was significantly higher than those in the other two regions, likely attributable to its more prosperous economy [133].

Our study presents a successful approach to achieve complete follow-up for cancer patients in southern China, offering a reliable depiction of NPC survival over the past decade. Our findings reveal that NPC survival rates are lower in the general population than in hospital-based cohorts, which was an expected result but had not been previously established. These results are likely to be applicable to other regions or countries where NPC is endemic, and suggest that the actual benefits of newer therapies may be less than what is observed in clinical trials. To reduce disparities in NPC mortality rates nationally and globally, it is necessary to allocate more medical resources to enable widespread cancer screening and improve access to advanced radiation therapy facilities in remote areas.

6.1.2 Pretreatment BMI and body shape and NPC prognosis

Study II, a prospective cohort study on NPC in southern China examined the associations of BMI and body shape at different time points with all-cause and NPC-specific mortality, taking measures to address concerns of reverse causation. The study found that overweight individuals at diagnosis had a lower mortality rate of approximately 25% compared to those with normal weight. Additionally, individuals with the thinnest body shape at diagnosis had a 68% higher mortality rate than those with a normal body shape. These associations were not modified by cancer stage. BMI and body shape at age 20 did not have any correlation with overall or NPC-specific mortality. The study's lack of heterogeneity by tumor stage at diagnosis and the detection of similar associations with BMI and body shape 10 years before diagnosis suggest that the results were not primarily due to reverse causation.

Our study has implications for both biology and clinical practice. NPC is typically treated with radiation, with or without chemotherapy. However, radiation therapy can cause painful side effects on the oral and pharyngeal mucosa, which can negatively impact a patient's nutritional balance [134]. This can lead to patients' being reluctant to eat, which can have a significant impact on their overall health. The average radiation course for NPC patients lasts for about 45 days. Individuals with a lower BMI and thinner body shape, indicating poor nutrition at baseline, may be less tolerant of the side effects of radiation therapy, leading to discontinuation of treatment and poorer survival outcome [135,136]. Given our hypothesis that the observed associations between BMI and body shape, and mortality are related to adherence to radiotherapy, our findings suggest that greater clinical attention should be directed toward improving the nutrition and treatment tolerance to achieve better survival outcomes of this patient group.

6.1.3 Pretreatment plasma EBV DNA and NPC prognosis

Study III, a prospective population-based cohort study in southern China, revealed that individuals with detectable plasma EBV DNA (observed in over 90% of cases) had more than twice the risk of all-cause and NPC-specific mortality compared to those with undetectable EBV DNA in plasma. The study also demonstrated strong positive dose-response associations between log-transformed plasma EBV DNA and all-cause and NPC-specific mortality, as evidenced by both linear and spline models.

Higher pretreatment EBV DNA levels may reflect a more extensive tumor burden, which could contribute to a higher risk of mortality. Reportedly, EBV can facilitate metastasis, invasion, and recurrence of malignant cells. EBV can target cell adhesion molecules such as cadherin and integrin, which can trigger a mesenchymal-like phenotype in host malignant cells [137]. This phenomenon is an essential developmental program that can promote metastasis, drug resistance, and tumor recurrence. There are novel treatment trials targeting on EBV. For instance, EBV-specific tumor-infiltrating lymphocytes were taken from the patient, expanded in vitro, and then transferred back to the same patient [138,139]. One phase II clinical trial showed long-term benefits of adoptive transfer of EBV-specific T cells among 23 nonmetastatic NPC patients [140].

Our findings reveal that the correlation between plasma EBV DNA and mortality exhibited attenuation beyond the 5-year mark following diagnosis. Notably, prior investigations have not undertaken an exploration of whether this correlation undergoes temporal variation. In a broader context, patients with NPC who surpass the five-year survival threshold may be regarded as being statistically "cured", denoting a mortality pattern similar to that observed in the general population [141].

The present study's results suggest that NPC patients with high pretreatment plasma EBV DNA loads may benefit from heightened clinical management and surveillance for up to five years postdiagnosis. This is particularly relevant given that NPC recurrence is often the result of tumor cells invading the circulation, and pretreatment EBV DNA levels are indicative of tumor burden. Therefore, a high EBV DNA load may signal the need for more intensive chemotherapy to reduce the risk of recurrence. Conducting randomized controlled trials that stratify patients based on pretreatment plasma EBV DNA levels may be a useful avenue for further exploring this issue.

6.1.4 Oral microbiota and NPC prognosis

Study IV indicated that certain measures of reduced enrichment and phylogenetic within-community diversity were linked to increased disease-specific and overall mortality, and some measures of

between-community diversity and composition were associated with mortality. However, the results were not entirely conclusive since certain measures of alpha and beta diversities did not exhibit any association with either overall or NPC-specific mortality.

The concept that diverse and well-balanced microbiome plays a crucial role in the immune function of human oral mucosa is widely accepted [142]. The commensal microbiota has the potential to safeguard hosts from colonization by exogenous pathogens and overgrowth by indigenous pathobionts. The mechanism underlying the observed relationship between the oral microbiota and NPC prognosis remains unclear. A potential explanation is that the microbiota may impact host immune function or contribute to the development of severe mucositis, which is a common side effect of radiotherapy in NPC patients. A study conducted by Hou et al. [143], revealed that two specific microbial taxa in the retropharyngeal wall were significantly associated with the progression of mucositis. This complication can lead to discontinuation of chemotherapy or radiotherapy, as well as malnutrition, which may contribute to poorer prognosis among NPC patients. Another longitudinal study showed that the abundance of *Streptococcus* and *Actinobacillus* was higher among patients with severe mucositis than those without severe mucositis during radiotherapy [101]. In a cohort study of 802 NPC cases [102], intratumoral bacterial load was found to be negatively associated with T-lymphocyte infiltration and linked to poor survival outcomes.

Mechanisms underlying the microbiota and survival of patients with other cancers revealed that commensal microbiota was associated with the efficacy of immunotherapy [144,145]. A study demonstrated that an abnormal gut microbiome composition in mouse models can cause primary resistance to immune checkpoint inhibitors for those with advanced melanoma, non-small cell lung cancer, or renal cell carcinoma [146]. Antibiotics use could inhibit the clinical benefit of checkpoint inhibitors in patients with advanced cancer. Additionally, metagenomics analysis of patient stool samples at diagnosis revealed associations between clinical responses and fecal microbiota transplantation [146].

Our research provides insight into the connection between microbiota and NPC survival. As the relationship between gut microbiota and cancer immunotherapy response has been explored, we are inspired to investigate the potential impact of microbiota on NPC treatment response.

6.2 Methodological considerations

6.2.1 Strengths

To our knowledge, our study represents the only population-based investigation of NPC survival within an NPC-endemic region, featuring a notably high enrollment rate and an exceptional degree of follow-up completeness. Our research also encompasses comprehensive data collection regarding unified clinical stage, detailed radiation therapy modalities, and demographic and environmental risk factors, enabling us to adjust for numerous potential confounders. Consequently, our study yielded valid, representative, and generalizable insights regarding NPC survival in southern China, which is burdened with the greatest incidence of NPC.

6.2.2 Selection bias

Selection bias refers to a potential source of distortion in estimates of disease occurrence or of effects arising from procedures that influence initial, ongoing study participation or loss to follow-up. In the

NPCGEE study, we achieved relatively high participation rates among cases (84%), therefore, the selection bias that comes from initial participation should be limited. However, time-to-event studies can be prone to another form of selection bias that arises from incomplete follow-up correlated with both the exposure and outcome variables. Survival analysis typically assumes that individuals lost to follow-up are comparable to those who remain in the study. It is important to consider the possibility that certain factors may affect an individual's likelihood of dropping out, such as worsening symptoms or a belief that staying in touch with medical professionals will result in better care. As a result, the assumption of similarity between those who are lost to follow-up and those who remain in the study may not always hold true [147]. The high participation rates and near-complete follow-up suggest that selection bias may be minimal.

6.2.3 Information bias

Systematic error in a study can arise because the information on exposures, covariates and outcome collected about the study subjects is erroneous, for example, measurement error for continuous variables and misclassification for discrete variables. Recall bias, a common type of information bias, is one concern for **Study II**. Body size and shape at age 20 years and 10 years before diagnosis were assessed by self-report, and are therefore prone to error, although this error is likely to be non-differential due to the prospective data collection. Another potential misclassification is the cause of deaths in **Studies I, II, III and IV**, which are not always easily classified even in hospitals. However, due to the nondifferential misclassification of binary causes of death, it will not bias the direction of rate ratio estimate [148].

6.2.4 Confounding

Confounding is one systematic error and a common concern in nearly all non-randomized studies. Critics may argue that a third variable could potentially negate the relationship between the exposure and outcome. There are methods to control for confounding at study design stage, i.e., randomization, restriction, and matching, and at analytic stage, i.e., standardization, stratification, regression adjustment, and weighting [148].

To alleviate the concern regarding confounding, we adjusted for numerous potential confounders, including age, sex, education, occupation, sampling season, or cancer stage in **Studies I, II, III and IV**. However, there is always residual confounding unknown or unmeasured. In **Study IV**, we lacked information on antibiotic use and anti-inflammatory therapy prior to saliva collection, which could affect both microbiome diversity and survival outcomes and cause bias, although previous work has reported that oral microbiome is more stable to antibiotics than fecal microbiome [149].

6.2.5 External validity

External validity, also referred as generalizability, is a crucial aspect of research design that pertains to the extent to which the findings of a study can be applied to other situations, populations, and period. This thesis was conducted in a population-based NPC-endemic area where 84% of the cases were enrolled during the study period. Hence, the findings of our study have high external validity and may be applied to a broader population in southern China. The survival patterns of NPC patients from **Study I** can be generalized to NPC cases diagnosed with NPC between 2010 and 2013 in southern China. Similarly, the conclusions drawn from **Studies II and III** have high external validity in southern China and could potentially extend to other high-incidence NPC areas such as Hong Kong,

Taiwan, and southeast Asia, although the size of the association may be different due to the sensitivity and specificity of the qPCR test by amplified fragments and methods [79,150] and the different cut-off values of categorical BMI as well as population differences in the distribution of plasma EBV DNA load [83] and overweight. The result of **Study IV** was strictly generalizable to NPC in endemic Guangxi Autonomous Region in southern China, as the composition of the oral microbiome is highly influenced by dietary patterns and geographic location [151,152].

7 Conclusions

Based on the papers included in this thesis, several conclusions can be drawn.

- Using a passive-active-passive circle follow-up strategy, high complete follow-up in a population-based setting in southern China was achievable. Population-based NPC survival lags behind large-hospital-based survival. Earlier diagnosis could contribute to substantial reductions in NPC mortality (**Study I**).

- Being overweight or obese at NPC diagnosis was linked to lower all-cause and NPC-specific mortality, while having a leaner body shape was associated with higher mortality compared to a normal weight/body shape (**Study II**).

- Robust evidence existed indicating a correlation between pretreatment plasma EBV DNA levels and NPC prognosis, with a positive dose-response relationship. However, the associations were weakened beyond five-year postdiagnosis mark (**Study III**).

- Certain measures of oral microbiome diversity and NPC prognosis were correlated. Reduced within-community diversity was associated with increased mortality risk. The observed associations may result from global patterns, rather than specific microbiota (**Study IV**).

8 Points of perspective

This thesis not only addressed original research questions but also sparked further inquiry and inspired us to tackle unresolved NPC issues related to clinical management, as discussed in the following section.

In **Study I,** in addition to overall survival for NPC, we calculated net survival, i.e., NPC-specific survival, in southern China. This metric could be used to compare disease-specific survival outcomes between different populations or time periods, or provide insights into the effectiveness of healthcare interventions or treatment protocols. However, due to the incomplete data on cause of death, as well as variability in methods of classifying causes of death for those with available data, analyses of net survival are prone to bias. As an alternative, relative survival can be used as a proxy for net survival when the disease of interest is the primary cause of death in the population under study. To estimate relative survival in the NPCGEE patient cohort, we will try to obtain life tables based on the general population's age- and sex-specific mortality rates, thereby enabling us to perform relative survival analysis. Second, as we observed from the Kaplan-Meier curves, the overall survival probability declined slightly beyond five years of diagnosis. This observation prompts further inquiry to examine differences in survival rates and causes of death after the five-year mark, to provide potentially valuable evidence for patient monitoring. Third, our findings indicated that approximately one-fifth of deaths could have been avoided if all cases had been diagnosed at an early stage. These results motivate us to further explore early detection methods for NPC. Both serum EBV antibody and plasma EBV DNA levels have demonstrated favorable performance characteristics and cost-effectiveness as biomarkers for early NPC detection in high-risk populations, yet their predictive positive value (PPV) ranged from only 4% to 11% [153]. Our previous work showed that combining epidemiological characteristics, serum EBV antibody levels and human and EBV single-nucleotide polymorphisms (SNPs) increased the PPV to 32% [154]. Yet, few randomized controlled trials of NPC screening have been conducted, and more such studies are warranted to determine whether these tools, perhaps in combination with other biomarkers, are valid basis for effective NPC screening to reduce NPC mortality.

In **Study II,** we found associations of BMI and body shape, including 10 years before diagnosis with NPC prognosis. Yet, potential misclassification of BMI and body shape based on memory is a concern. We could eliminate this bias in a natural longitudinal cohort by capturing objective measurements of anthropometric index, whether through medical records or in-person measurements. Such a study design would also enable us to evaluate associations between post-diagnosis changes in BMI and treatment response, providing insight into how we can perform nutrition management in clinical practice. Although any associations between post-diagnosis BMI and NPC outcomes would be susceptible to reverse causation (i.e., an effect of disease severity on BMI, as opposed to an effect of BMI on NPC prognosis), it would be worthwhile to understand these associations both to improve NPC prognosis and to maintain a favorable BMI, given the importance of BMI as a mediator of other health outcomes and health-related quality of life.

In **Study III,** we observed a dose-response and time-dependent association between pre-treatment plasma EBV DNA and mortality among NPC patients in population-based setting. Our results suggest that a higher EBV DNA load may indicate higher tumour burden, which may in turn denote a requirement for more intensive chemotherapy. To further investigate this issue, randomized controlled

trials of chemotherapy regimens stratified by pre-treatment plasma EBV DNA levels could provide valuable insights. Other questions prompted by our findings include, how pre-treatment plasma EBV DNA could be incorporated into the current staging framework, and whether longitudinal monitoring of plasma EBV DNA could help with risk stratification and adaptation of therapeutic protocols. For example, there are some ongoing clinical trials (i.e., NCT02135042, NCT05772208) of applying individualized treatment based on consideration of both cancer stage and EBV DNA load. In addition, due to a lack of quantitative agreements among laboratories internationally, clinically relevant thresholds have not been standardized [150]. One possibility has been raised by Miller et al., who developed laboratory BamHI-W droplet digital PCR (dPCR) to define prognostic thresholds [155,156]. Further studies are needed to integrate dPCR or other EBV-DNA-based metrics with cancer stage to guide individualized treatment.

In **Study IV,** we uncovered associations between oral microbiota as a community and NPC prognosis. In light of our findings that oral microbiome diversity is associated with NPC mortality, the next step is to explore underlying mechanism. First, since diet impacts the gut microbiome and in turn, gut microbiota take part in the process of digestion, we could explore potential triangular associations among diet, microbiota and NPC prognosis. Second, alteration of the immune response (reflecting either an effect of the gut microbiome on the host immune system, or an impact of immunological changes on gut microbiome composition) is another potential explanation for the observed associations [102]. To explore this possibility, we plan to collect data on host immune status (e.g., as measured by C-reactive protein, immunoglobulins G, M and A, complete blood count, peripheral blood lymphocytes using epigenetic cell count measures [157] and tumor infiltrating lymphocytes phenotyping). Third, gut microbiota dysbiosis might indicate resistance to radiotherapy, chemotherapy and immunotherapy [158,159], thereby indirectly affecting NPC prognosis. To further investigate this potential mechanism, it may be worth considering incorporation of bacteria into NPC treatment protocols. Fourth, microbiota may have effects on toxicity of chemoradiotherapy, as some observational studies and animal experiments reported [160,161]. It is worth exploring associations of the interactions or modifications.

9 Funding sources

Studies in this thesis have received financial supports from:

- National Cancer Institute, United States
- Swedish Research Council, Sweden
- Swedish Cancer Society
- 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
- National Natural Science Foundation of China
- China Scholarship Council

10 Acknowledgements

Finally, it is time to write the acknowledgement section. Filled with gratitude, thankfulness, and relief, I am starting to look back in the past four years. Some people came across my world and left for next journey; some stepped into my life and moved forward together. There are happiness and suffering, laughters and tears, gratefulness and frustration, and destinations and obstacles. All of these make the whole picture of life. I'd like to write us and our stories down, to keep them everlasting.

My dear principal supervisor, **Weimin Ye**: Words cannot express my gratitude. You have been like a strict and knowledgeable parent to me throughout my PhD journey. You introduce me to the science world, teach me research skills, and guide me in life. Your rigorous approach to research, passion, and scientific talent have inspired me and shown me the direction to valuable scientific work. I am deeply grateful to have had you by my side during this important journey.

Li Yin, my dear co-supervisor and biostatistician, thanks for always responding to me promptly and kindly every time I've came across questions in statistics. I enjoyed every meeting and talk with you, either online during pandemic or onsite at MEB.

Amir Sohrabi, my co-supervisor, thank you for sharing your solid knowledge of lab experience with me, and managing the lab. I appreciated discussions with you in solving the bugs occurred.

Magnus Nilsson, my dear co-supervisor, thank you for providing me the chance to participate in gastric cancer project and sharing insights on molecular classification of stomach cancer.

Ellen Chang, my dear co-author, many thanks for improving my writing skills and epidemiological knowledge. Every 'baby article' I sent to you, it returned to me as a 'wise elder' with a clear structure and well-reasoned arguments. You have supported me a lot in the past four years, although we live in separate corners of the world. I'm looking forward to visiting you and meeting you soon.

Hans-Olov Adami, my respectful co-author, thank you for sharing your comments and suggestions on our NPCGEE articles. Your profound knowledge in research astonishes me and motivates me to go deeper.

Fang Fang, thank you. I really enjoyed talking with you. You offered me the thoughts of keeping the direction in mind and try your best approaching it, and the recommendation for daily routine to know better in one research field. **Wei He,** thank you for being my mentor and support.

I'd like to express my gratitude to my opponent **Anil Chatuvedi** and committee members **Nele Brusselaers, Lennart Greiff and Giorgio Tettamanti**. Thank you for your generosity to spare time, share your expertise and comments in my defense season. Thank **Zheng Chang** for being the chairperson of my defense.

I'd like to express my thanks to **Ellen Chang, Shifeng Lian, Yufeng Chen and Yu Zhou,** for helping me with proofreading, suggestions, and feedback on my thesis. Warm thanks to **Aleksandra Kanina, Fen Yang, Yufeng Chen, Jiangwei Sun and Kejia Hu,** for being my pre-dissertation committee and precious comments.

Many thanks to my co-authors and collaborators: **Justine Debelius**, **Ruimei Feng, Yufeng Chen, Ellen T. Chang, Li Yin, Tingting Huang, Yi Huang, Guangwu Huang, Yixin Zeng, Yancheng Li,**

Xiang Zhou, Feng Zhou, Canqiong Su, Xue Xiao, Weihua Jia, Yuming Zheng, Hans-Olov Adami, Yixin Zeng, Yonglin Cai, Miao Xu, Zhe Zhang and Weimin Ye. Thank you all for your help, input, support and excellent expertise in our research. So nice and lucky to work with you on the NPCGE project. **Ruimei**, you are sister-like to me. You always encourage me and discuss with me to find solutions when I come across difficulties. **Yonglin Cai, Zhe Zhang, Yuming Zheng, and Miao Xu,** thank you for supporting me when I travelled to your place to do the fieldwork. Without you, my thesis cannot go smoothly. Thank the friends or teachers, **Shaohua Su, Yanping Li, Qin Cao, Qing Liu, Lu Zhou, Yuqiang Lu, Yunliang Lu, Zihao Wang, Yunyun Lan, Yanping Yang and Xuemin Zhong**, I met and offered me support in Wuzhou, Nanning, Guangzhou and Zhangqing during the follow-up process.

To friends and colleagues from weimin's group: **Yufeng Chen, Amir Sohrabi, Joar Franzen, Anna Berglund, Justine Debelius, Ji Zhang, Jingru Yu, Tingting Huang, Fatemeh Sadeghi, Hui Xu, Weiwei Bian, Fei He, Yanhong Duo, Song Lin, Isabella Ekheden, Ulrika Zagai and Yawen Sun.** Dear **Justine:** big big thanks to you. You are such a nice friend and excellent instructor in microbiota field. You led me into the microbiota world and contributed a lot to my PhD study. **Yufeng**, thank you for sharing information and experience of the NPCGEE study. Also thank *Feifei*, we had a nice trip to Uppsala and relaxing dinners. **Fatemeh, Amir and Joar,** thanks so much for the support in the lab. It's struggling but also precious memories that we finally get the lab work done. The summer picnics, Kubb, and casual meals besides work are super happy and relaxing time with you guys, as well as *Matilda* and *Wollmar*. **Tingting,** thank you for being my fadder and helping me adapt to life here. Thank **Jingru, Ji, Weiwei, Fei, Hui, Yanhong and Yawen** for the fika time and mealtime together.

To my friends and colleagues at MEB, I want to express my heartfelt gratitude for the opportunity to meet you, build friendships with you, share experiences, and engage in discussions on both academic and casual topics. Your presence has enriched my life in countless ways, and I am truly thankful for the connections we have formed.

Marie Reilly, Rino Bellocco, Mark Clements, Alexander Ploner, **Arvid Sjölander** et al., thank you for your exceptional course. I am also grateful to **Paul Lichtenstein and Mark Clements** for their guidance and assistance as directors of doctoral studies. Thank you, **Alessandra Nanni** and **Gunilla Nilsson Roos** for support us on courses, half time and PhD dissertation. Your contributions have been invaluable to my academic growth. I would like to extend my deep appreciation to MEB **administrators, IT** group, **HR** staff, **PhD board** and environment maintainers.

Thanks to the past and current MEBers, included but not exclude to: **Cen Chen, Nanbo Zhu, Honghui Yao, Yuqi Zhang, Jet Termorshuizen, Alessandro Bosi, Erwei Zeng, Xinhe Mao, Jiayao Lei, Shuang Hao, Shengxin Liu, Zhengan Lu, Ge Bai, Malin Rådström, Qingyang Xiao, Mengping Zhou, Haoming Yang, Xiaoying Kang, Le Zhang, Yinxi Wang, Yanbo Feng, Zihan Dong, Hilda Björk Danielsdottir, Miriam Martini, Lin Li, Yuying Li, Mujin Nie, Ying Xiong, Wenjiang Deng, Hong Xu, Bowen Tang, Philippe Weitz, Peter Lind, Yuliya Leontyeva, Chenxi Qin, Shadi Azam, Shihua Sun, Weiyao Yin, Xia Li, Yangjun Liu, Yingying Yang, Yan Chen, Zheng Ning, Lu Pan, Ruyue Zhang** and **Pui San Tan.** *Cen, Ge, Nanbo, Honghui, Yuqi, Shengxin, Erwei, Xiaoying, and Xinhe*, thanks for all the greetings, activities, and gathering. *Nanbo*, thanks for valuable discussions on the epidemiological and statistics. *Jiayao*, thanks for the encouragement, Christmas dinner and your excellent thesis as my reference book. *Jet,* I'm always inspired by your passion. *Alessandro,* thanks for the fitness and morning coffee time.

Many thanks to my office mates: **Aleksandra Kanina, Soran Rabin Bozorg, Hui Xu, Yuanhang Yang and Qing Shen**, for the accompany, talking and having lunch together.

Thanks to friends and colleagues I've met at **TRACK (KFC)** in Novum: **Rui He, Wenyi Zheng, Qiang Wang, Xiuming Liang & Qing Wang, Moustapha Hassan, Kathrin Reiser, Ying Zhao, Qingyang Zhang, Zheyu Niu and Poomy Pandey.** The Christmas lunch we had in the kitchen, BBQ and home dinner, were enjoyable and unforgettable.

I am immensely grateful for my dear friends in Sweden. This includes, but is not limited to: **Shifeng Lian, Kejia Hu & Gang Wang, Can Cui, Yanhong Su, Chenyan Wu, Shengyuan Zeng, Xinyuan Liu, Qianwei Liu, Xi Chen, Jinghua Wu, Mu Nie, Yanping Yang, Jingru Fu, Hui Wan, Dang Wei, Huan Cai, Jianjian Gao & Fen Yang, Lihua Luo &Jiacheng Zhu, Xiaomeng Zhang, Yanhong Bao & Tuli Sun, Yuanyuan Gu & Guannan Zhou, Xijie Yang, Kaile Chen, Yuezhen Huang, Xiangfu Zhong, Yihan Hu, Yunbing Shen, Letian Zhang, Zhao Cheng and Duoduo.** There are nice, enjoyable, and relaxing moments that we laughed together and also stressful and down moments we signed and cried together. All of these memories have been beautifully captured and woven together to create the story of our journey today. *Kejia*, you are super considerate friend to me, both in research and life. Especially during the pandemic time, so grateful that you were beside me, hanging out, workout, shopping, eating and sharing ideas together. *Yanhong, Chenyan, Can, Xinyuan, Jinghua* and *Mu,* my close friends, dearest sisters and neighbors, I feel so lucky to have you in the community. *Xi, Xiaoxiannv, and Huan,* I cherished the impressive journey in Gotland with you guys.

Many thanks to my master supervisor **Mingfang Ji**. You provide me with an excellent platform, instruct me, care about me and pave the way for my further research. Big thanks to the **Ji&Tang** family: **Shifeng Lian, Xia Yu, Li Deng, Wen Quan, Xiaodong Chen, Yulong Xie, Chaxiang Teng, Jialiang Wu, Yaqiong Zhang, Qiuyu Luo** et al.

Thanks to my friends outside of Sweden, including but not limited to: **Fan Yang & Sisi Huang, Yanan Jin, Wen Wen, Mengyuan Qiang, Qingnan Tang, Shasha He, Yao Luo and Mengling Li.** *Fan and Sisi,* I'm grateful that you are so patient and kind to knock the door, encourage me to step outside, and embrace the unknown with open arms.

Yu Zhou, I'm grateful for your presence in my life. Thank you for being here and supportive.

亲爱的爸爸妈妈，感谢您们的爱和支持，放手让我去尝试不同的生活。达哥和老细，还有我的亲人们，衷心地感谢您们陪伴和照顾我的父母，谢谢您们。

Thank you for being a part of my journey and for helping me to grow and become the person I am today.

11 References

1. Chang ET, Ye W, Zeng YX, Adami HO. The evolving epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiology Biomarkers and Prevention*. 2021;30(6):1035-1047. doi:10.1158/1055-9965.EPI-20-1702

2. Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. *The Lancet*. 2019;394(10192):64-80. doi:10.1016/S0140-6736(19)30956-0

3. Sun XSS, Liu SLL, Luo MJJ, et al. The Association Between the Development of Radiation Therapy, Image Technology, and Chemotherapy, and the Survival of Patients With Nasopharyngeal Carcinoma: A Cohort Study From 1990 to 2012. *Int J Radiat Oncol Biol Phys*. 2019;105(3):581-590. doi:10.1016/j.ijrobp.2019.06.2549

4. Huang WYY, Lin CLYYL, Lin CLYYL, et al. Survival outcome of patients with nasopharyngeal carcinoma: a nationwide analysis of 13407 patients in Taiwan. *CLINICAL OTOLARYNGOLOGY*. 2015;40(4):327-334. doi:10.1111/coa.12371

5. Chan SK, Chau SC, Chan SY, et al. Incidence and Demographics of Nasopharyngeal Carcinoma in Cheung Chau Island of Hong Kong-A Distinct Geographical Area With Minimal Residential Mobility and Restricted Public Healthcare Referral Network. *Cancer Control*. 2021;28. doi:10.1177/10732748211047117

6. World Health Organization. Global Cancer Observatory. Accessed February 3, 2023. https://gco.iarc.fr/

7. Bray F, Colombet M, Mery L, et al. Cancer Incidence in Five Continents, Vol. XI. IARC Scientific Publication No. 166. Lyon: International Agency for Research on Cancer. Published 2021. Accessed October 21, 2021. https://publications.iarc.fr/597

8. Wei KR, Zheng RS, Zhang SW, Liang ZH, Li ZM, Chen WQ. Nasopharyngeal carcinoma incidence and mortality in China, 2013. *Chin J Cancer*. 2017;36(1):90. doi:10.1186/s40880-017-0257-9

9. Hsu C, Shen YC, Cheng CC, Hong RL, Chang CJ, Cheng AL. Difference in the Incidence Trend of Nasopharyngeal and Oropharyngeal Carcinomas in Taiwan: Implication from Age-Period-Cohort Analysis. *Cancer Epidemiology, Biomarkers & Prevention*. 2006;15(5):856-861. doi:10.1158/1055-9965.EPI-05-0821

10. Yu WM, Hussain SSM. 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. doi:10.1017/S0022215109005623

11. Mousavi SM, Sundquist J, Hemminki K. Nasopharyngeal and hypopharyngeal carcinoma risk among immigrants in Sweden. *Int J Cancer*. 2010;127(12):2888-2892. doi:10.1002/IJC.25287

12. Rottenberg Y, Levine H, Keinan-Boker L, Derazne E, Leiba A, Kark JD. Risk of nasopharyngeal carcinoma penetrates across immigrant generations: A migrant cohort study of 2.3 million Jewish Israeli adolescents. *Int J Cancer*. 2017;140(5):1060-1067. doi:10.1002/IJC.30525

13. Parkin DM, Iscovich J. RISK OF CANCER IN MIGRANTS AND THEIR DESCENDANTS IN ISRAEL: II. CARCINOMAS AND GERM-CELL TUMOURS. *J Cancer*. 1997;70:654-660. doi:10.1002/(SICI)1097-0215(19970317)70:6

14. 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. doi:10.1002/IJC.2910540403

15. Buell P. Race and place in the etiology of nasopharyngeal cancer: a study based on California death certificates. *Int J Cancer*. 1973;11(2):268-272. doi:10.1002/IJC.2910110204

16. Jia WH, Collins A, Zeng YX, et al. Complex segregation analysis of nasopharyngeal carcinoma in Guangdong, China: evidence for a multifactorial mode of inheritance (complex segregation analysis of NPC in China). *European Journal of Human Genetics 2005 13:2*. 2004;13(2):248-252. doi:10.1038/sj.ejhg.5201305

17. Bei JX, Li Y, Jia WH, et al. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat Genet*. 2010;42(7). doi:10.1038/ng.601

18. Tse KP, Su WH, Chang KP, et al. Genome-wide Association Study Reveals Multiple Nasopharyngeal Carcinoma-Associated Loci within the HLA Region at Chromosome 6p21.3. *The American Journal of Human Genetics*. 2009;85(2):194-203. doi:10.1016/J.AJHG.2009.07.007

19. Tang M, Lautenberger JA, Gao X, et al. The Principal Genetic Determinants for Nasopharyngeal Carcinoma in China Involve the HLA Class I Antigen Recognition Groove. *PLoS Genet*. 2012;8(11):e1003103. doi:10.1371/JOURNAL.PGEN.1003103

20. Chin YM, Mushiroda T, Takahashi A, et al. HLA-A SNPs and amino acid variants are associated with nasopharyngeal carcinoma in Malaysian Chinese. *Int J Cancer*. 2015;136(3):678-687. doi:10.1002/IJC.29035

21. Dai W, Zheng H, Cheung AKL, et al. Whole-exome sequencing identifies MST1R as a genetic susceptibility gene in nasopharyngeal carcinoma. *Proc Natl Acad Sci U S A*. 2016;113(12). doi:10.1073/pnas.1523436113

22. Ng CC, Yew PY, Puah SM, et al. A genome-wide association study identifies ITGA9 conferring risk of nasopharyngeal carcinoma. *Journal of Human Genetics 2009 54:7*. 2009;54(7):392-397. doi:10.1038/jhg.2009.49

23. Qin H De, Shugart YY, Bei JX, et al. Comprehensive Pathway-Based Association Study of DNA Repair Gene Variants and the Risk of Nasopharyngeal Carcinoma. *Cancer Res*. 2011;71(8):3000. doi:10.1158/0008-5472.CAN-10-0469

24. de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob Health*. 2020;8(2):e180-e190. doi:10.1016/S2214-109X(19)30488-7/ATTACHMENT/8970D333-63A3-477A-88C5-AD322F2D0DDA/MMC1.PDF

25. 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. doi:10.1038/S41588-019-0436-5

26. 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. doi:10.1093/AJE/KWS404

27. IARC. Volume 100E of IARC Monographs on the Evaluation of Carcinogenic risks to Humans: Personal habits and indoor combustions. *IARC Monogr Eval Carcinog Risks Hum*. 2012;100E.

28. 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. *Journal of Nutrition*. 2019;149(9):1596-1605.

29. 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. doi:10.1002/CNCR.30643

30. Chang ET, Liu Z, Hildesheim A, et al. Active and Passive Smoking and Risk of Nasopharyngeal Carcinoma: A Population-Based Case-Control Study in Southern China. *Am J Epidemiol*. 2017;185(12):1272-1280.

31. Huang T, Ploner A, Chang ET, et al. Dietary patterns and risk of nasopharyngeal carcinoma: a population-based case-control study in southern China. *American Journal of Clinical Nutrition*. 2021;114(2):462-471.

32. Lin C, Cao SM, Chang ET, et al. Chinese nonmedicinal herbal diet and risk of nasopharyngeal carcinoma: A population-based case-control study. *Cancer*. 2019;125(24):4462-4470.

33. Lyu YH, Lin CY, Xie SH, et al. Association Between Traditional Herbal Diet and Nasopharyngeal Carcinoma Risk: A Prospective Cohort Study in Southern China. *Front Oncol*. 2021;11. doi:10.3389/fonc.2021.715242

34. Feng R, Chang ET, Liu Q, et al. Intake of alcohol and tea and risk of nasopharyngeal Carcinoma: A population-based case–control study in Southern China. *Cancer Epidemiology Biomarkers and Prevention*. 2021;30(3):545-553. doi:10.1158/1055-9965.EPI-20-1244

35. WHO. *Pathology and Genetics of Head and Neck Tumours*. IARC Press; 2005.

36. Haddad RI, Hicks WL, Hitchcock YJ, et al. NCCN Guidelines Version 1.2023 Head and Neck Cancers Continue NCCN Guidelines Panel Disclosures. Published online 2022. Accessed December 30, 2022. https://www.nccn.org/home/member-

37. Bai R, Sun J, Xu Y, Sun Z, Zhao X. Incidence and mortality trends of nasopharynx cancer from 1990 to 2019 in China: an age-period-cohort analysis. *BMC Public Health*. 2022;22(1). doi:10.1186/S12889-022-13688-7

38. Carioli G, Negri E, Kawakita D, Garavello W, La Vecchia C, Malvezzi M. Global trends in nasopharyngeal cancer mortality since 1970 and predictions for 2020: Focus on low-risk areas. *Int J Cancer*. 2017;140(10). doi:10.1002/ijc.30660

39. Fitzmaurice C, Dicker D, Pain A, et al. The Global Burden of Cancer 2013. *JAMA Oncol*. 2015;1(4):505-527. doi:10.1001/jamaoncol.2015.0735

40. Tang LL, Chen WQ, Xue WQ, et al. Global trends in incidence and mortality of nasopharyngeal carcinoma. *Cancer Lett*. 2016;374(1):22-30. doi:10.1016/J.CANLET.2016.01.040

41. Nasopharyngeal Cancer Survival Rates | American Cancer Society. Accessed April 28, 2023. https://www.cancer.org/cancer/nasopharyngeal-cancer/detection-diagnosis-staging/survival-rates.html

42. Pan XX, Liu YJ, Yang W, Chen YF, Tang WB, Li CR. Histological subtype remains a prognostic factor for survival in nasopharyngeal carcinoma patients. *Laryngoscope*. 2020;130(3):E83-E88. doi:10.1002/LARY.28099

43. Lu X, Wang FL, Guo X, et al. Favorable prognosis of female patients with nasopharyngeal carcinoma. *Chin J Cancer*. 2013;32(5):283. doi:10.5732/CJC.012.10058

44. Su L, Lin Q, Li R, et al. Prognostic value of nutritional impairment on treatment-related toxicity and survival in patients with nasopharyngeal carcinoma taking normal nutrition before radiotherapy. *Head Neck*. 2020;42(12):3580-3589. doi:10.1002/HED.26426

45. Moon SH, Cho KH, Lee CG, et al. IMRT vs. 2D-Strahlentherapie oder konformaler 3D-Strahlentherapie beim Nasopharynxkarzinom: Folgen für das Überleben in einer koreanischen multizentrischen retrospektiven Studie (KROG 11-06). *Strahlentherapie und Onkologie*. 2016;192(6):377-385. doi:10.1007/S00066-016-0959-Y/FIGURES/3

46. Zeng H, Chen W, Zheng R, et al. Changing cancer survival in China during 2003–15: a pooled analysis of 17 population-based cancer registries. *Lancet Glob Health*. 2018;6(5):e555-e567. doi:10.1016/S2214-109X(18)30127-X

47. Gupta S, Kong W, Peng Y, Miao Q, Mackillop WJ. Temporal trends in the incidence and survival of cancers of the upper aerodigestive tract in Ontario and the United States. *Int J Cancer*. 2009;125(9):2159-2165. doi:10.1002/IJC.24533

48. ENCR | European Network of Cancer Registries. Accessed March 17, 2023. https://www.encr.eu/

49. 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. doi:10.1093/annonc/mdz231

50. Johnson CJ, Weir HK, Yin D, Niu X. The impact of patient follow-up on population-based survival rates. *J Registry Manag*. 2010;37(3):86-103. Accessed March 15, 2023. https://europepmc.org/article/med/21462880

51. Overview of the SEER Program. Accessed March 17, 2023. https://seer.cancer.gov/about/overview.html

52. Organization WH. Obesity and overweight. Published online 2017. Accessed May 2, 2023. http://www.who.int/mediacentre/factsheets/fs311/en/

53. Nishida C, Barba C, Cavalli-Sforza T, et al. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *The Lancet*. 2004;363(9403). doi:10.1016/S0140-6736(03)15268-3

54. Bentham J, Di Cesare M, Bilano V, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *The Lancet*. 2017;390(10113):2627-2642. doi:10.1016/S0140-6736(17)32129-3

55. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Absence of Excess Body Fatness. *New England Journal of Medicine*. 2016;375(8):794-798. doi:10.1056/NEJMSR1606602

56. Piqueras P, Ballester A, Durá-Gil J V., Martinez-Hervas S, Redón J, Real JT. Anthropometric Indicators as a Tool for Diagnosis of Obesity and Other Health Risk Factors: A Literature Review. *Front Psychol*. 2021;12:2618. doi:10.3389/FPSYG.2021.631179/BIBTEX

57. Park Y, Peterson LL, Colditz GA. The Plausibility of Obesity Paradox in Cancer. *Cancer Res*. 2018;78(8):1898. doi:10.1158/0008-5472.CAN-17-3043

58. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body Fatness and Cancer — Viewpoint of the IARC Working Group. *New England Journal of Medicine*. 2016;375(8):794-798. doi:10.1056/NEJMSR1606602/SUPPL_FILE/NEJMSR1606602_DISCLOSURES.PDF

59. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *The Lancet*. 2014;384(9945):755-765. doi:10.1016/S0140-6736(14)60892-8

60. Tu H, McQuade JL, Davies MA, et al. Body mass index and survival after cancer diagnosis: A pan-cancer cohort study of 114 430 patients with cancer. *The Innovation*. 2022;3(6):100344. doi:10.1016/J.XINN.2022.100344

61. Greenlee H, Unger JM, LeBlanc M, Ramsey S, Hershman DL. Association between body mass index (BMI) and cancer survival in a pooled analysis of 22 clinical trials. *Cancer Epidemiol Biomarkers Prev*. 2017;26(1):21. doi:10.1158/1055-9965.EPI-15-1336

62. Petrelli F, Cortellini A, Indini A, et al. Association of Obesity With Survival Outcomes in Patients With Cancer: A Systematic Review and Meta-analysis. *JAMA Netw Open*. 2021;4(3). doi:10.1001/JAMANETWORKOPEN.2021.3520

63. Gupta A, Majumder K, Arora N, et al. Premorbid body mass index and mortality in patients with lung cancer: A systematic review and meta-analysis. *Lung Cancer*. 2016;102:49-59. doi:10.1016/j.lungcan.2016.10.017

64. Pierce BL, Neuhouser ML, Wener MH, et al. Correlates of circulating C-reactive protein and serum amyloid A concentrations in breast cancer survivors. *Breast Cancer Res Treat*. 2009;114(1):155. doi:10.1007/S10549-008-9985-5

65. Sung H, Siegel RL, Torre LA, et al. Global patterns in excess body weight and the associated cancer burden. *CA Cancer J Clin*. 2019;69(2). doi:10.3322/CAAC.21499

66. Feng R, Chang ET, Liu Z, et al. Body mass index, body shape, and risk of nasopharyngeal carcinoma: A population-based case–control study in Southern China. *Cancer Med*. 2019;8(4):1835-1844.

67. Shen LJ, Chen C, Li BF, Gao J, Xia YF. High Weight Loss during Radiation Treatment Changes the Prognosis in Under-/Normal Weight Nasopharyngeal Carcinoma Patients for the Worse: A Retrospective Analysis of 2433 Cases. *PLoS One*. 2013;8(7):e68660. doi:10.1371/JOURNAL.PONE.0068660

68. Hu JY, Yl W, Xia YF, Gao J, Liu ZG, Tao YL. Impact of pretherapy body mass index on prognosis of nasopharyngeal carcinoma. *Chin J Cancer*. 2009;28(10). doi:10.5732/CJC.009.10415

69. Ouyang PY, Zhang LN, Tang J, et al. Evaluation of Body Mass Index and Survival of Nasopharyngeal Carcinoma by Propensity-Matched Analysis: An Observational Case-Control Study. *Medicine*. 2016;95(2). doi:10.1097/MD.0000000000002380

70. Shen GP, Xu FH, He F, et al. Pretreatment lifestyle behaviors as survival predictors for patients with nasopharyngeal carcinoma. *PLoS One*. 2012;7(5). doi:10.1371/JOURNAL.PONE.0036515

71. Huang PY, Wang CT, Cao KJ, et al. Pretreatment body mass index as an independent prognostic factor in patients with locoregionally advanced nasopharyngeal carcinoma treated with chemoradiotherapy: Findings from a randomised trial. *Eur J Cancer*. 2013;49(8):1923-1931. doi:10.1016/j.ejca.2013.01.027

72. Gruberg L, Weissman NJ, Waksman R, et al. The impact of obesity on the short-term and long-term outcomes after percutaneous coronary intervention: The obesity paradox? *J Am Coll Cardiol*. 2002;39(4):578-584. doi:10.1016/S0735-1097(01)01802-2

73. Flegal KM, Graubard BI, Williamson DF, Cooper RS. Reverse causation and illness-related weight loss in observational studies of body weight and mortality. *Am J Epidemiol*. 2011;173(1):1-9. doi:10.1093/AJE/KWQ341

74. Horesh A, Tsur AM, Bardugo A, Twig G. Adolescent and Childhood Obesity and Excess Morbidity and Mortality in Young Adulthood—a Systematic Review. *Current Obesity Reports 2021 10:3*. 2021;10(3):301-310. doi:10.1007/S13679-021-00439-9

75. 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. doi:10.18632/ONCOTARGET.19692

76. Sanosyan A, De Maudave AF, Bollore K, et al. The impact of targeting repetitive BamHI-W sequences on the sensitivity and precision of EBV DNA quantification. *PLoS One*. 2017;12(8):e0183856. doi:10.1371/JOURNAL.PONE.0183856

77. Ryan JL, Fan H, Glaser SL, Schichman SA, Raab-Traub N, Gulley ML. Epstein-Barr Virus Quantitation by Real-Time PCR Targeting Multiple Gene Segments : A Novel Approach to Screen for the Virus in Paraffin-Embedded Tissue and Plasma. *J Mol Diagn*. 2004;6(4):378. doi:10.1016/S1525-1578(10)60535-1

78. Shao JY, Zhang Y, Li YH, et al. Comparison of Epstein-Barr Virus DNA Level in Plasma, Peripheral Blood Cell and Tumor Tissue in Nasopharyngeal Carcinoma. *Anticancer Res*. 2004;24(6):4059-4066. Accessed April 12, 2023. https://ar.iiarjournals.org/content/24/6/4059

79. Fung SYH, Lam JWK, Chan KCA. Clinical utility of circulating Epstein-Barr virus DNA analysis for the management of nasopharyngeal carcinoma. *Chin Clin Oncol*. 2016;5(2):18-18. doi:10.21037/CCO.2016.03.07

80. Mo WN, Tang AZ, Zhou L, Huang GW, Wang Z, Zeng Y. Analysis of Epstein-Barr viral DNA load, EBV-LMP2 specific cytotoxic T-lymphocytes and levels of CD4+CD25+ T cells in patients with nasopharyngeal carcinomas positive for IgA antibody to EBV viral capsid antigen. *Chin Med J (Engl)*. 2009;122(10):1173-1178. doi:10.3760/CMA.J.ISSN.0366-6999.2009.10.011

81. Lo YMD, Chan LYS, Lo KW, et al. Quantitative analysis of cell-free Epstein-Barr virus DNA in plasma of patients with nasopharyngeal carcinoma. *Cancer Res*. 1999;59(6).

82. Dennis Lo YM. Quantitative analysis of Epstein-Barr virus DNA in plasma and serum: Applications to tumor detection and monitoring. *Ann N Y Acad Sci*. 2001;945:68-72. doi:10.1111/j.1749-6632.2001.tb03865.x

83. Alami IEL, Gihbid A, Charoute H, et al. Prognostic value of Epstein-Barr virus DNA load in nasopharyngeal carcinoma: a meta-analysis. *Pan Afr Med J*. 2022;41. doi:10.11604/PAMJ.2022.41.6.28946

84. Zhang W, Chen Y, Chen L, et al. The Clinical Utility of Plasma Epstein–Barr Virus DNA Assays in Nasopharyngeal Carcinoma: The Dawn of a New Era?: A Systematic Review and Meta-Analysis of 7836 Cases. *Medicine*. 2015;94(20). doi:10.1097/MD.0000000000000845

85. Prayongrat A, Chakkabat C, Kannarunimit D, Hansasuta P, Lertbutsayanukul C. Prevalence and significance of plasma Epstein-Barr Virus DNA level in nasopharyngeal carcinoma. *J Radiat Res*. 2017;58(4):509. doi:10.1093/JRR/RRW128

86. He SS, Wang Y, Peng H, et al. Pretreatment Alkaline Phosphatase and Epstein-Barr Virus DNA Predict Poor Prognosis and Response to Salvage Radiotherapy in Patients with Nasopharyngeal Carcinoma and Metachronous Bone-Only Metastasis. *J Cancer*. 2017;8(3):417-424. doi:10.7150/JCA.17310

87. Alfieri S, Iacovelli NA, Marceglia S, et al. Circulating pre-treatment Epstein-Barr virus DNA as prognostic factor in locally-advanced nasopharyngeal cancer in a non-endemic area. *Oncotarget*. 2017;8(29):47780. doi:10.18632/ONCOTARGET.17822

88. Leung S fai, Chan ATC, Zee B, et al. Pretherapy quantitative measurement of circulating Epstein–Barr virus DNA is predictive of posttherapy distant failure in patients with early-stage nasopharyngeal carcinoma of undifferentiated type. *Cancer*. 2003;98(2):288-291. doi:10.1002/CNCR.11496

89. Lo YMD, Chan ATC, Chan LYS, et al. Molecular prognostication of nasopharyngeal carcinoma by quantitative analysis of circulating Epstein-Barr virus DNA. *Cancer Res*. 2000;60(24):6878-6881.

90. Jin YN, Yao JJ, Zhang F, et al. Is pretreatment Epstein-Barr virus DNA still associated with 6-year survival outcomes in locoregionally advanced nasopharyngeal carcinoma? *J Cancer*. 2017;8(6):976. doi:10.7150/JCA.18124

91. David LA, Maurice CF, Carmody RN, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature 2013 505:7484*. 2013;505(7484):559-563. doi:10.1038/nature12820

92. Jernberg C, Löfmark S, Edlund C, Jansson JK. Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *The ISME Journal 2007 1:1*. 2007;1(1):56-66. doi:10.1038/ismej.2007.3

93. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature 2012 486:7402*. 2012;486(7402):222-227. doi:10.1038/nature11053

94. Liu Z, Chang ET, Liu Q, et al. Oral hygiene and risk of nasopharyngeal carcinoma - A population-based case-control study in China. *Cancer Epidemiology Biomarkers and Prevention*. 2016;25(8):1201-1207.

95. Liu QY, Liao Y, Wu YX, et al. The Oral Microbiome as Mediator between Oral Hygiene and Its Impact on Nasopharyngeal Carcinoma. *Microorganisms*. 2023;11(3):719. doi:10.3390/MICROORGANISMS11030719/S1

96. Li L, Deng X, Zou Y, Lv XP, Guo Y. Characterization of the nasopharynx microbiota in patients with nasopharyngeal carcinoma vs. healthy controls. *Brazilian Journal of Microbiology*. 2021;52(4):1873. doi:10.1007/S42770-021-00594-0

97. 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). doi:10.1128/MSYSTEMS.00065-20/SUPPL_FILE/REVIEWER-COMMENTS.PDF

98. Liao Y, Zhang JB, Lu LX, et al. Oral Microbiota Alteration and Roles in Epstein-Barr Virus Reactivation in Nasopharyngeal Carcinoma. *Microbiol Spectr*. 2023;11(1). doi:10.1128/SPECTRUM.03448-22

99. Zhu XX, Yang XJ, Chao YL, et al. The Potential Effect of Oral Microbiota in the Prediction of Mucositis During Radiotherapy for Nasopharyngeal Carcinoma. *EBioMedicine*. 2017;18:23-31. doi:10.1016/J.EBIOM.2017.02.002

100. Huang T, Debelius JW, Ploner A, et al. Radiation Therapy-Induced Changes of the Nasopharyngeal Commensal Microbiome in Nasopharyngeal Carcinoma Patients. *Int J Radiat Oncol Biol Phys*. 2021;109(1):145-150. doi:10.1016/J.IJROBP.2020.08.054

101. Xu Y, Teng F, Huang S, et al. Changes of saliva microbiota in nasopharyngeal carcinoma patients under chemoradiation therapy. *Arch Oral Biol*. 2014;59(2):176-186. doi:10.1016/J.ARCHORALBIO.2013.10.011

102. Qiao H, Tan XR, Li H, et al. Association of Intratumoral Microbiota With Prognosis in Patients With Nasopharyngeal Carcinoma From 2 Hospitals in China. *JAMA Oncol*. 2022;8(9):1301. doi:10.1001/JAMAONCOL.2022.2810

103. Stunkard A, Sørensen T, Schulsinger F. Use of the Danish Adoption Register for the study of obesity and thinness. *undefined*. Published online 1983.

104. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. 2016;4(10). doi:10.7717/PEERJ.2584

105. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology 2019 37:8*. 2019;37(8):852-857. doi:10.1038/s41587-019-0209-9

106. Bokulich NA, Subramanian S, Faith JJ, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods 2012 10:1*. 2013;10(1):57-59. doi:10.1038/nmeth.2276

107. Amir A, McDonald D, Navas-Molina JA, et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems*. 2017;2(2). doi:10.1128/MSYSTEMS.00191-16/SUPPL_FILE/SYS002172091SF3.PDF

108. Janssen S, McDonald D, Gonzalez A, et al. Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. *mSystems*. 2018;3(3). doi:10.1128/MSYSTEMS.00021-18

109. Antoine Lucas by, Tuszynski J, Bengtsson H, et al. Package "digest" Title Create Compact Hash Digests of R Objects. Published online 2022. Accessed April 28, 2023. http://dirk.eddelbuettel.com/code/digest.html

110. Shannon CE. A Mathematical Theory of Communication. *The Bell System Technical Journal*. 1948;27:623-656.

111. Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Conserv*. 1992;61(1):1-10. doi:10.1016/0006-3207(92)91201-3

112. Lozupone C, Knight R. UniFrac: A new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*. 2005;71(12):8228-8235. doi:10.1128/AEM.71.12.8228-8235.2005/ASSET/9AE40EDF-8FF9-4184-8DD7-367105439C23/ASSETS/GRAPHIC/ZAM0120562270003.JPEG

113. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and Qualitative β Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Appl Environ Microbiol*. 2007;73(5):1576. doi:10.1128/AEM.01996-06

114. Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol Monogr*. 1957;27(4):325-349. doi:10.2307/1942268

115. Whittaker RJ, Willis KJ, Field R. Scale and species richness: towards a general, hierarchical theory of species diversity. *J Biogeogr*. 2001;28(4):453-470. doi:10.1046/J.1365-2699.2001.00563.X

116. Tuomisto H. A diversity of beta diversities: Straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography*. 2010;33(1):2-22. doi:10.1111/J.1600-0587.2009.05880.X

117. Chinese Service Center. China's Education System. Published 2006. Accessed November 1, 2021. http://www.china.org.cn/english/LivinginChina/185292.htm

118. Timmermann C. 'Just give me the best quality of life questionnaire': the Karnofsky scale and the history of quality of life measurements in cancer trials. *Chronic Illn*. 2013;9(3):179. doi:10.1177/1742395312466903

119. Carolyn C. Compton, David R. Byrd, Julio Garcia-Aguilar, Scott H. Kurtzman, Alexander Olawaiye, Mary Kay Washington. AJCC Cancer Staging Atlas. *AJCC Cancer Staging Atlas*. Published online 2012. doi:10.1007/978-1-4614-2080-4

120. Rawlings JO, Pantula SG, Dickey D a. Applied Regression Analysis : A Research Tool , Second Edition Springer Texts in Statistics. Published online 1998:671.

121. Abeysekera WWM, Sooriyarachchi MR. Use of Schoenfeld's global test to test proportional hazards assumption in the Cox proportional hazards model: An application to clinical study. *J Natl Sci Found*. 2009;37(1):41-51. doi:10.4038/jnsfsr.v37i1.456

122. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1972;34(2):187-202. doi:10.1111/J.2517-6161.1972.TB00899.X

123. Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. *Stat Med*. 2010;29(9):1037-1057. doi:10.1002/SIM.3841

124. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. *Stata J*. 2009;9(2):265-290.

125. Royston P, Lambert PC, Royston P, Lambert PC. Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model. *Flexible parametric survival analysis using Stata: beyond the Cox model*. Published online 2011:25-59. Accessed April 19, 2023. https://EconPapers.repec.org/RePEc:tsj:spbook:fpsaus

126. Anderson MJ. Permutational Multivariate Analysis of Variance (PERMANOVA). *Wiley StatsRef: Statistics Reference Online*. Published online November 15, 2017:1-15. doi:10.1002/9781118445112.STAT07841

127. Knights D, Costello EK, Knight R. Supervised classification of human microbiota. *FEMS Microbiol Rev*. 2011;35(2):343-359. doi:10.1111/J.1574-6976.2010.00251.X

128. Martino C, Morton JT, Marotz CA, et al. A Novel Sparse Compositional Technique Reveals Microbial Perturbations. *mSystems*. 2019;4(1). doi:10.1128/MSYSTEMS.00016-19/SUPPL_FILE/MSYSTEMS.00016-19-ST002.XLSX

129. Du Y, Feng R, Chang ET, et al. Body mass index and body shape before treatment and nasopharyngeal carcinoma prognosis: a population-based patient cohort study in southern China. *Int J Cancer*. Published online March 25, 2023. doi:10.1002/IJC.34524

130. Du Y, Feng R, Chang ET, et al. Influence of Pre-treatment Saliva Microbial Diversity and Composition on Nasopharyngeal Carcinoma Prognosis. *Front Cell Infect Microbiol*. 2022;0:269. doi:10.3389/FCIMB.2022.831409

131. Wang L, Lu J, Lang J. SC09.04 Radiotherapy in China. *Journal of Thoracic Oncology*. 2017;12(1):S98-S100. doi:10.1016/j.jtho.2016.11.085

132. Q L, JO C, QH H, YH L. Trends in the survival of patients with nasopharyngeal carcinoma between 1976 and 2005 in Sihui, China: a population-based study. *Chin J Cancer*. 2013;32(6):325-333. doi:10.5732/CJC.012.10189

133. LY Sheng, P Zhang, SY Zhong. China statistical yearbook. Published 2010. Accessed October 31, 2022. http://www.stats.gov.cn/english/Statisticaldata/AnnualData/

134. Wang C, Wang F, Min X, et al. Toxicities of chemoradiotherapy and radiotherapy in nasopharyngeal carcinoma: an updated meta-analysis. *J Int Med Res*. 2019;47(7):2832. doi:10.1177/0300060519858031

135. Zhao F, Yang D, Li X. Effect of radiotherapy interruption on nasopharyngeal cancer. *Front Oncol*. 2023;13:1507. doi:10.3389/FONC.2023.1114652/BIBTEX

136. Yang XL, Zhou GQ, Lin L, et al. Prognostic value of radiation interruption in different periods for nasopharyngeal carcinoma patients in the intensity-modulated radiation therapy era. *Cancer Med*. 2021;10(1):143. doi:10.1002/CAM4.3580

137. Su ZY, Siak PY, Leong CO, Cheah SC. The role of Epstein–Barr virus in nasopharyngeal carcinoma. *Front Microbiol*. 2023;14. doi:10.3389/FMICB.2023.1116143

138. Zhu QY, Zhao GX, Li Y, et al. Advances in pathogenesis and precision medicine for nasopharyngeal carcinoma. *MedComm (Beijing)*. 2021;2(2):175-206. doi:10.1002/MCO2.32

139. He J, Tang XF, Chen QY, et al. Ex vivo expansion of tumor-infiltrating lymphocytes from nasopharyngeal carcinoma patients for adoptive immunotherapy. *Chin J Cancer*. 2012;31(6):287. doi:10.5732/CJC.011.10376

140. Louis CU, Straathof K, Bollard CM, et al. Adoptive transfer of EBV-specific T cells results in sustained clinical responses in patients with locoregional nasopharyngeal carcinoma. *J Immunother*. 2010;33(9):983. doi:10.1097/CJI.0B013E3181F3CBF4

141. Lambert PC, Thompson JR, Weston CL, Dickman PW. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*. 2007;8(3):576-594. doi:10.1093/BIOSTATISTICS/KXL030

142. Tomkovich S, Jobin C. Microbiota and host immune responses: a love–hate relationship. *Immunology*. 2016;147(1):1. doi:10.1111/IMM.12538

143. Hou J, Zheng HM, Li P, Liu HY, Zhou HW, Yang XJ. Distinct shifts in the oral microbiota are associated with the progression and aggravation of mucositis during radiotherapy. *Radiotherapy and Oncology*. 2018;129(1):44-51. doi:10.1016/j.radonc.2018.04.023

144. Fessler J, Matson V, Gajewski TF. Exploring the emerging role of the microbiome in cancer immunotherapy. *J Immunother Cancer*. 2019;7(1):108. doi:10.1186/S40425-019-0574-4

145. Bae J, Park K, Kim YM. Commensal Microbiota and Cancer Immunotherapy: Harnessing Commensal Bacteria for Cancer Therapy. *Immune Netw*. 2022;22(1). doi:10.4110/IN.2022.22.E3

146. Routy B, Le Chatelier E, Derosa L, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science (1979)*. 2018;359(6371):91-97. doi:10.1126/SCIENCE.AAN3706/SUPPL_FILE/AAN3706_ROUTY_SM.PDF

147. Rothman KJ. Epidemiology in Clinical Settings. *Epidemiology An Introduction*. 2012;72:235-253.

148. Ahlbom A. Modern Epidemiology, 4th edition. TL Lash, TJ VanderWeele, S Haneuse, KJ Rothman. Wolters Kluwer, 2021. *Eur J Epidemiol*. 2021;36(8):767-768. doi:10.1007/S10654-021-00778-W

149. Zaura E, Brandt BW, de Mattos MJT, et al. Same Exposure but two radically different responses to antibiotics: Resilience of the salivary microbiome versus long-term microbial shifts in feces. *mBio*. 2015;6(6). doi:10.1128/MBIO.01693-15/SUPPL_FILE/MBO005152533S1.PDF

150. Kim KY, Le QT, Yom SS, et al. Current State of PCR-Based Epstein-Barr Virus DNA Testing for Nasopharyngeal Cancer. *JNCI Journal of the National Cancer Institute*. 2017;109(4):7. doi:10.1093/JNCI/DJX007

151. Santonocito S, Giudice A, Polizzi A, et al. A Cross-Talk between Diet and the Oral Microbiome: Balance of Nutrition on Inflammation and Immune System's Response during Periodontitis. *Nutrients*. 2022;14(12). doi:10.3390/NU14122426

152. Ogbanga N, Nelson A, Ghignone S, et al. The Oral Microbiome for Geographic Origin: An Italian Study. *Forensic Sci Int Genet*. 2023;64:102841. doi:10.1016/J.FSIGEN.2023.102841

153. Lam WKJ, King AD, Miller JA, et al. Recommendations for Epstein-Barr virus–based screening for nasopharyngeal cancer in high- and intermediate-risk regions. *JNCI: Journal of the National Cancer Institute*. 2023;2023:1-10. doi:10.1093/JNCI/DJAD012

154. Zhou X, Cao SM, Cai YL, et al. A comprehensive risk score for effective risk stratification and screening of nasopharyngeal carcinoma. *Nat Commun*. 2021;12(1). doi:10.1038/S41467-021-25402-Z

155. Miller JA, Huang C, Sahoo MK, Wang H, Le QT, Pinsky B. Comparison of Multiplex Real-Time PCR and Droplet Digital PCR for Detection of Plasma Epstein-Barr Virus DNA in Nasopharyngeal Carcinoma. *International Journal of Radiation Oncology*Biology*Physics*. 2021;111(3):e386-e387. doi:10.1016/J.IJROBP.2021.07.1130

156. Miller JA, Huang C, Yamamoto F, et al. Comparison of Real-Time PCR and Digital PCR for Detection of Plasma Epstein-Barr Virus DNA in Nasopharyngeal Carcinoma. *The Journal of Molecular Diagnostics*. 2023;0(0). doi:10.1016/J.JMOLDX.2023.03.007

157. Katzke VA, Le Cornet C, Mahfouz R, et al. Are Circulating Immune Cells a Determinant of Pancreatic Cancer Risk? A Prospective Study Using Epigenetic Cell Count Measures. *Cancer Epidemiology Biomarkers and Prevention*. 2021;30(12):2179-2187. doi:10.1158/1055-9965.EPI-21-0169/671136/AM/ARE-CIRCULATING-IMMUNE-CELLS-A-DETERMINANT-OF

158. Sadrekarimi H, Gardanova ZR, Bakhshesh M, et al. Emerging role of human microbiome in cancer development and response to therapy: special focus on intestinal microflora. *Journal of Translational Medicine 2022 20:1*. 2022;20(1):1-20. doi:10.1186/S12967-022-03492-7

159. Huang J, Liu W, Kang W, et al. Effects of microbiota on anticancer drugs: Current knowledge and potential applications. *EBioMedicine*. 2022;83:31900056-32000096. doi:10.1016/j.ebiom.2022.104197

160. Li HL, Lu L, Wang XS, et al. Alteration of gut microbiota and inflammatory cytokine/chemokine profiles in 5-fluorouracil induced intestinal mucositis. *Front Cell Infect Microbiol*. 2017;7(OCT):455. doi:10.3389/FCIMB.2017.00455/FULL

161. Shen S, Lim G, You Z, et al. Gut microbiota is critical for the induction of chemotherapy-induced pain. *Nature Neuroscience 2017 20:9*. 2017;20(9):1213-1216. doi:10.1038/nn.4606