Charles University

Faculty of Science

Dept. of Social Geography and Regional Development

Study programme: Epidemiology
Branch of study: Social epidemiology

Mgr. Alžběta Sekavová

Rizikové faktory HPV infekce a nádorů hlavy a krku

Risk factors of HPV infection and head and neck tumours

Diploma thesis

Prague 2017

Supervisor: RNDr. Ruth Tachezy, Ph.D.

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

Praha podpis autorky

Abstract

Epidemiology of head and neck cancers is currently intensively studied topic. Recent shift in the age of incidence towards younger population is generally attributed to growing proportion of head and neck cancers caused by human papillomavirus - sexually transmitted virus which causes asymptomatic, but sometimes persisting infection that can lead to malignant transformation of the infected tissue. Significance of the topic lies mainly in the prognostic advantage of patients with virally induced head and cancers and preventability of infections with human papillomavirus. First aim of this thesis is to demonstrate epidemiological trends of head and neck cancers in the Czech Republic, with the focus on change of age-specific incidence and mortality in the last three decades. Second aim of this thesis is to identify risk factors of oral infections with human papillomaviruses and head and neck cancer in a case-control and cross-sectional study of a hospital-based cohort.

Keywords: epidemiology, head and neck, HPV, tumour, cancer, risk factors

Abstrakt

Epidemiologie nádorů hlavy a krku je v současnosti intenzivně studovaným tématem. Nedávný posun věku incidence do mladších věkových kategorií je připisován vzůrstajícímu podílu nádorů hlavy a krku, které jsou způsobené lidským papilomavirem – sexuálně přenosným virem, který navozuje asymptomatické a někdy persistující infekce, které pak mohou vést k nádorové transformaci infikované tkáně. Lepší prognóza pacientů s nádory hlavy a krku indukovanými virovou infekcí a také preventabilita infekce lidským papilomavirem jsou hlavní důvody pro intenzivní studium této problematiky. Prvním cílem této práce je ukázat epidemiologické trendy nádorů hlavy a krku v České republice, s důrazem na změny věkově-specifické indicence a úmrtnosti během posledních tří dekád. Druhá část práce si klade za cíl identifikovat rizikové faktory infekce lidským papillomavirem v dutině ústní a také rizikové faktory nádorů hlavy a krku. K této studii případů a kontrol a průřezové studii je využita nemocniční kohorta.

Klíčová slova: epidemiologie, hlava a krk, HPV, nádor, rakovina, rizikové faktory

Ráda bych na tomto místě poděkovala svým rodičům a celé rodině za podporu při studiu a za motivaci při výzvách, které studium přineslo. Děkuji své školitelce doktorce Ruth Tachezy za její cenné rady a čas, který mé práci obětovala, a za příležitost být součástí její výzkumné skupiny a učit se pod jejím vedením. Fantastické čtyřce děkuji za skvělé zážitky během celého studia a zejména pak za molární podporu při sepisování našich diplomových prací.

Díky patří i doktorce Lustigové za připomínky k formálním náležitostem práce, Aničce Borovcové za rady k demografickým výpočtům a hlavně Lukášovi Hejtmánkovi za pomoc se skriptováním v R a statistickými analýzami.

V neposlední řadě děkuji Národní referenční laboratoři pro papillomaviry a polyomaviry a jejím členům za poskytnutí dat a vyšetřování vzorků, bez čehož by tato práce nemohla vzniknout.

Contents

List of Figures

List of Tables

List of Abbreviations

1 Introduction **8**

2 Literature review **9**
2.1 Head and neck cancer in international classification of diseases . . 9
2.2 Epidemiology and etiology of head and neck cancer 10
2.3 Two distinct categories of head and neck cancers 11
2.4 Human papillomavirus, its tumorigenesis and epidemiology 13
2.5 Risk factors of HPV infection 15
2.6 Risk factors of head and neck cancer 15
2.7 Prevention of head and neck cancer and HPV infection 16
2.8 Diagnostics and therapy of head and neck cancer 17
2.9 Significance of the research topic 18

3 Hypothesis and aims **19**

4 Data and methods **20**
4.1 Data 20
4.1.1 Data on HNSCC mortality and population 20
4.1.2 Data on HNSCC mortality in regions of the Czech Republic 20
4.1.3 Data on HNSCC incidence 21
4.1.4 Clinical data for risk factor analyses 21
4.2 Methods 23
4.2.1 Demographic analysis 23
4.2.2 Clinical data preprocessing 24
4.2.3 Statistical analysis 24

5 Results **25**
5.1 Epidemiological trends of head and neck cancer in the Czech Republic 27
5.1.1 Mortality from head and neck cancer in Czech population 27
5.1.2 Incidence of head and neck cancer in the Czech Republic . 33
5.1.3 Head and neck cancer incidence and mortality in regions of the Czech Republic 36
5.2 Risk factors of head and neck cancer and oral HPV infection . . . 40
5.2.1 Characteristics of participants 41
5.2.2 Risk factors of oral HPV infection 45
5.2.3 Risk factors of head and neck cancer 53

6 Discussion **66**

7 Conclusion **74**

Bibliography **75**

A Appendix **82**

List of Figures

2.1 Schematic depiction of head and neck cancer body regions. 9
2.2 Head and neck cancer death rate, Europe, men and women, 2012 10
2.3 Simplified schema of head and neck cancer etiology 12
2.4 Structure of human papillomavirus 13
2.5 Differences in HPV^+ and HPV^- HNSCC and their risk factors . . 15

5.1 Standardised death rates for ICD-10: C00–C14 and ICD-10: C00–C06; C09–C10, Czech Republic, men and women, 1994–2013 . . . 26
5.2 Standardised death rates from HNSCC, Czech Republic, men and women, years 1986–2013 27
5.3 Age-specific death rates from HNSCC, Czech Republic, men and women, 1986 and 2013 31
5.4 Age-specific death rates from HNSCC in two age categories (65–69 and 75–79), Czech Republic, men and women, 1986–2013 32
5.5 Age-specific incidence of HNSCC, Czech Republic, men and women, 2014 33
5.6 Age-specific incidence of HNSCC, Czech Republic, men and women, 1986 and 2014 35
5.7 Incidence of HNSCC in regions of the Czech Republic, men and women, 2014 37
5.8 Mortality from HNSCC in regions of the Czech Republic, men and women, 2014 39
5.9 Schema of study design 40
5.10 Distribution of continuous variables in the dataset 43
5.11 Frequencies of categorical factors in the dataset 44
5.12 Gender of $oHPV^+$ and $oHPV^-$ participants 45
5.13 Age of $oHPV^+$ and $oHPV^-$ participants 46
5.14 Completed years of completed education of $oHPV^+$ and $oHPV^-$ participants 47
5.15 Drinking habits of $oHPV^+$ and $oHPV^-$ participants 48
5.16 Smoking habits of $oHPV^+$ and $oHPV^-$ participants 48
5.17 Sexual behaviours of $oHPV^+$ and $oHPV^-$ participants 49
5.18 Gender of patients and controls 53
5.19 HPV and HR HPV presence in oral washings of patients and controls 54
5.20 Age of patients and controls 55
5.21 Completed years of completed education of patients and controls . 56
5.22 Drinking habits of patient and control groups 57
5.23 Smoking habits of patient and control groups 58
5.24 Sexual behaviours of patients and controls 59

A.1 Regions of the Czech Republic 82

List of Tables

2.1 Incidence and death rate from HNSCC, Czech Republic, men and women, 2014 . . . 11
2.2 High risk and low risk HPV types . . . 14

5.1 Standardised death rates from HNSCC, Czech Republic, men and women, 1986–2013 . . . 28
5.2 Age-specific death rates from HNSCC, Czech Republic, men and women, 1986 and 2013 . . . 30
5.3 Age-specific incidence of HNSCC, Czech Republic, men and women, 1986 and 2014 . . . 34
5.4 Incidence of HNSCC in regions of the Czech Republic, men and women, 2014 . . . 36
5.5 Death rate from HNSCC in regions of the Czech Republic, men and women, 2014 . . . 38
5.6 Characteristics of participants. . . . 42
5.7 Association of oHPV status with participants' demographics and risk behaviours . . . 50
5.8 Risk factors of oHPV infection . . . 51
5.9 Presence of HR HPVs in oral washing of controls and patients . . . 54
5.10 Association of HNSCC diagnosis with participants' demographics and risk behaviours . . . 60
5.11 Risk factors of head and neck cancer . . . 62
5.12 Strenght of prediction models . . . 63
5.13 Risk factors of HR HPV positive head and cancers . . . 64
5.14 Risk factors of HR HPV negative head and cancers . . . 65

List of Abbreviations

BTL	Basic Tabulation List
CoDQL	Cause of Death Query online
CI	confidence interval
DNA	deoxyribonucleic acid
HNSCC	head and neck squamous cell carcinoma
HPV	human papillomavirus
HPV^- HNSCC	human papillomavirus negative HNSCC
HPV^+ HNSCC	human papillomavirus positive HNSCC
HR HPV	high-risk human papillomavirus
ICD	International Classification of Diseases
LR HPV	low-risk human papillomavirus
NA	not available
NRL	National Reference Laboratory
oHPV	oral HPV
$oHPV^+$	oral HPV-positive (in oral washing)
$oHPV^-$	oral HPV-negative (in oral washing)
OR	odds ratio
ORL	otorhinolaryngology
SD	standard deviation
WHO	World Health Organization

1. Introduction

Incidence of head and neck cancers has been increasing in recent years. This is attributed to increasing number of cases associated with oral infection with human papillomavirus (HPV). Not only is the infection with HPV preventable by immunisation, but prognosis of patients with head and neck cancer associated with HPV is better and treatment plans for such patients could be accordingly modified. In the Czech Republic, no study yet described the epidemiological trends of head and neck cancers observed worldwide. Additionally, currently only one study in Czech population investigated the risk factors of head and neck cancers and none addressed the risk factors of oral HPV infection itself. In this thesis we aim to contribute to the state of knowledge about these issues in the Czech Republic.

Firstly, we demonstrate epidemiological trends of head and neck cancers in the Czech Republic. We present results of demographic analyses of mortality and incidence of head and neck cancers.

Secondly, we aim to analyse risk factors of head and neck cancers and oral HPV infection with data provided by National Reference Laboratory for Papillomaviruses and Polyomaviruses. We present results of cross-sectional study of risk factors of oral HPV infections (oHPV) as well as results of case-control study investigating the risk factors of head and neck cancers.

Identifying potential risk factors of oHPV infections and HPV-associated head and neck cancers should help to target primary and secondary prevention of both HPV infection and the associated cancers.

2. Literature review

Head and neck squamous cell carcinoma (HNSCC) is an overarching diagnosis (ICD-10: C00–C14) which includes carcinomas at several anatomical sites. These are namely: lip, base of tongue, gum, floor of mouth, palate, parotid gland, tonsil, oropharynx, nasopharynx, piriform sinus, hypopharynx and other locations such as major salivary glands, parts of tongue, oral cavity and pharynx (Figure 2.1). HNSCC is one of the most prevalent cancers worldwide. HNSCC is currently in the spotlight of research because of its interestingly changing epidemiological patterns and unavailable preventive screening program.

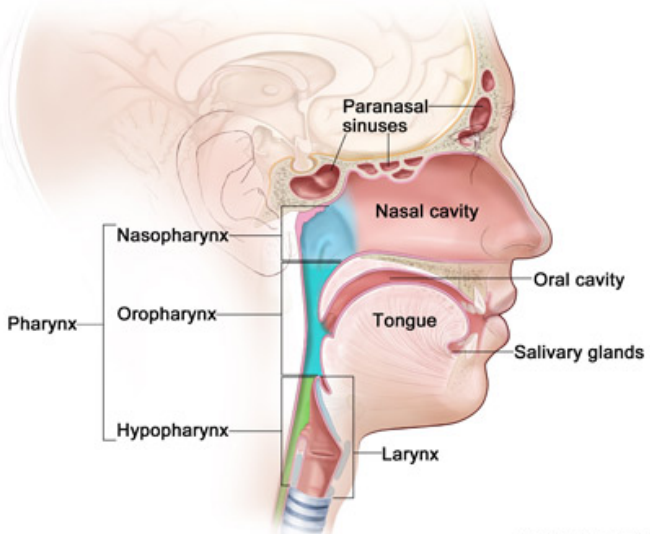

Figure 2.1: Schematic depiction of head and neck cancer body regions
Retrieved from NCI (2017)

2.1 Head and neck cancer in international classification of diseases

While in ICD-10 the term HNSCC includes diagnoses C00–C14, in older revision ICD-9 (used in years 1979–1993), the included diagnoses are 140–149 (i.e. B08 in Basic Tabulation List, BTL). As has been pointed out in review by Hashibe & Sturgis (2013), changes in classification might compromise epidemiological stud-

ies which use data from public databases. Nevertheless, with the knowledge of these limitations and careful interpretation, these data still provide important information about recent changes in HNSCC epidemiology.

2.2 Epidemiology and etiology of head and neck cancer

Head and neck cancer is the fifth most prevalent cancer in the world. Among European countries in 2012, Hungary had the highest incidence and mortality from cancers of lip, oral cavity and pharynx (ICD-10: C00–C14). The Czech Republic had the seventeenth highest mortality from HNSCC among men and the sixth highest mortality among women in Europe (Figure 2.2). Similarly to other countries, both incidence and mortality in the Czech Republic are higher for men than women (Table 2.1). Also, HNSCC is the seventh most frequent cause of death by cancer in Czech men, whereas sixteenth most frequent for Czech women (Ferlay et al. 2013).

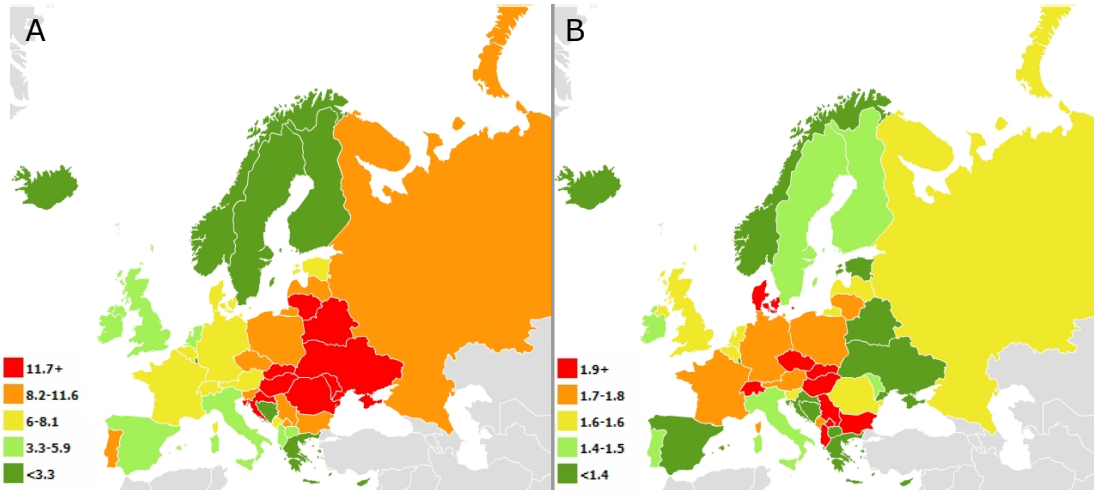

Figure 2.2: Head and neck cancer (ICD-10: C00–C14) death rate per 100, 000 population, Europe, men (A) and women (B), 2012

Retrieved from International Agency for Research on Cancer (2012)

Table 2.1: Incidence and death rate from HNSCC, Czech Republic, men and women, 2014

Source of data: IHIS CR (2014)

	Incidence per 100, 000 population	Death rate per 100, 000 population
Males	21.4	11.0
Females	8.4	3.0

Numerous studies and reviews focus on epidemiology of HNSCC in USA and Western Europe (reviewed in Gooi et al. (2016)), but there is only limited number of studies on this topic in the Czech Republic, even though Czech Republic has excellent (and publicly available) sources of data on incidence and mortality (namely National Oncology Registry, web portal SVOD.cz, Czech Demographic Handbook of the Czech Statistical Office).

2.3 Two distinct categories of head and neck cancers

Although HNSCC was historically perceived as one disease, recent discoveries identified two distinct populations of these tumours. HNSCC can be either associated with tobacco and alcohol consumption or can be induced by persistent human papillomavirus (HPV) infection of oral tissue. Traditionally, smoking and alcohol consumption were considered to be the major risk factors for HNSCC (Hashibe et al. 2007, Gandini et al. 2008). Even though the prevalence of smoking in the world decreases (Agaku et al. 2014), the incidence and mortality from HNSCC do not decline as is expected. Lately, subpopulation of these tumours was shown to be associated with persistent infection with HPV, which can be detected in tissue of these HPV-positive tumours (HPV^+ HNSCC) (Figure 2.3).

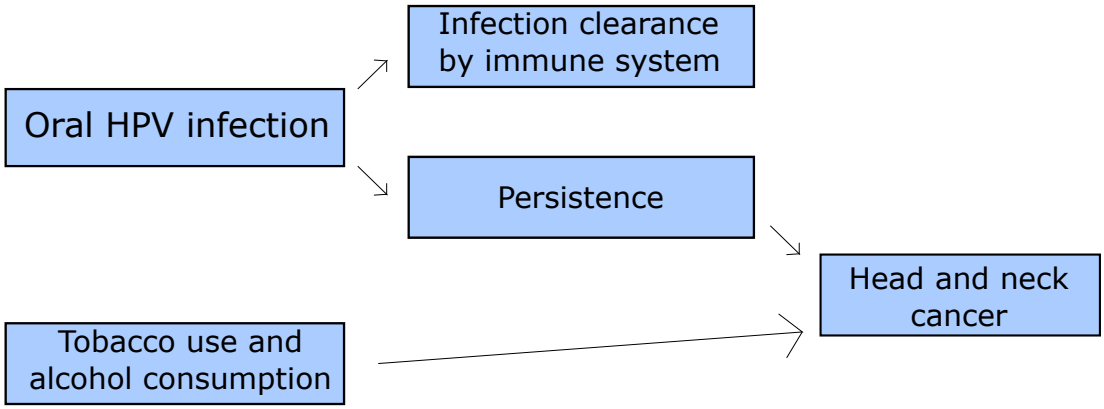

Figure 2.3: Simplified schema of head and neck cancer etiology

Author's own work

While incidence of HPV-negative (HPV^-) HNSCC which are induced by traditional risk factors declines, distinctive proportion of HPV^+ HNSCC is emerging (Gooi et al. 2016). The two subgroups have also distinct risk factors as well as population at risk (reviewed in Gillison et al. (2008)). Patients with HPV^+ HNSCC are typically younger (<60 years of age) and are likely to smoke and drink alcohol less than patients with HPV^- HNSCC. Another difference is the anatomical site of these tumours: HPV^+ HNSCC are usually found at the base of tongue and in tonsils, whereas HPV^- HNSCC are mainly located in oral cavity, larynx and hypopharynx (reviewed in Lewis et al. (2015)). Up to 70% of HNSCC are currently associated with HPV infection (reviewed in Chaturvedi et al. (2011)).

In USA, significant increase of HPV^+ HNSCC incidence since 1984 was shown to be associated with better prognosis and survival of patients (Chaturvedi et al. 2011). The importance of distinction between the two etiologies lies in potentially different treatment plans for the patients (Mirghani et al. 2015, Urban et al. 2014)(see section 2.8).

2.4 Human papillomavirus, its tumorigenesis and epidemiology

Human papillomavirus (HPV) was first described in 1950s, but came into the spotlight of research when it was found to be associated with human malignancies in patients with epidermodysplasia verruciformis in 1978 (Orth et al. 1978). HPV was also found in cervical cancer tissue by Harald zur Hausen and his team in 1984 (Boshart et al. 1984). Since then, its causative relationship with cervical cancer has been established. HPV is also etiological agent of benign warts as well as several malignancies including anal, vulvar and penile cancers (zur Hausen 2009) and head and neck cancers.

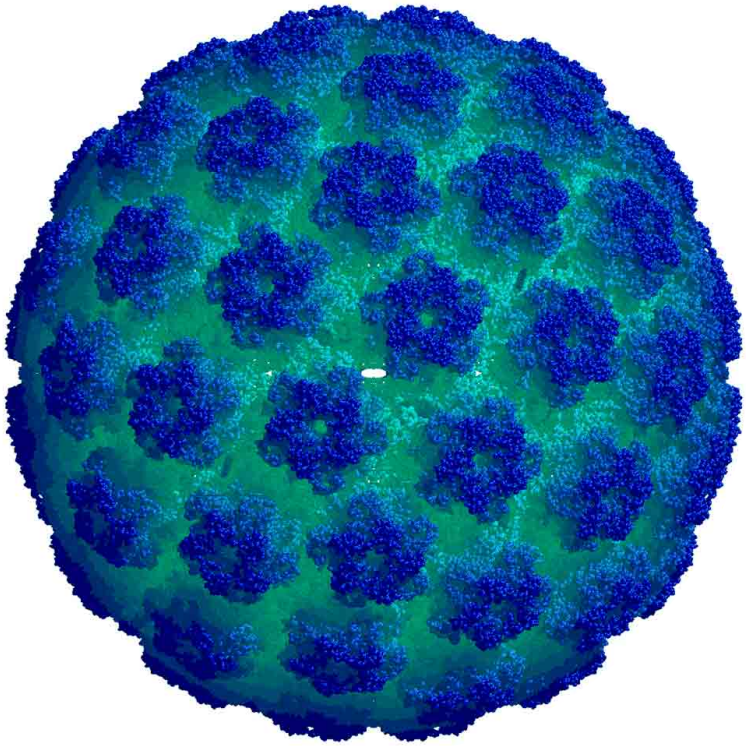

Figure 2.4: Structure of human papillomavirus

Adapted from Modis et al. (2002)

HPV is a small non-enveloped virus with double stranded circular DNA genome enclosed inside the 50 nm diameter capsid (Figure 2.4). There are more than 200 subtypes of HPV which differ in sequence of gene coding major capsid

protein L1. HPV subtypes are divided based on their tissue tropism - infecting either mucosal tissues or skin (or both in case of some HPV types) (reviewed in Knipe et al. (2002)). HPV from genus alfa are also commonly classified as low risk (LR HPV) or high risk (HR HPV) depending on their ability to cause malignancies (Table 2.2). LR HPV types (e.g. HPV-6 and HPV-11) are causing benign warts and HR HPV (most notably HPV-16 and HPV-18) are responsible for malignant tumours. Most cases (up to 90%) of HPV^+ HNSCC are caused by HPV-16 (Gillison et al. 2000).

Table 2.2: High risk (HR) and low risk (LR) HPV types

Adapted from Hamšíková (2008)

	HPV types
HR HPV	16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, 82
LR HPV	6, 11, 40, 42, 43, 44, 54, 61, 70, 72, 81, cand89

HPV is a sexually transmitted virus and it is considered to be the most common sexually transmitted viral infection. The estimated prevalence is up to 39% in young women (Howell-Jones et al. 2012), but up to 90% of adult population has antibodies against HPV, i.e. they encountered the virus during their life course. However, not all infections will lead to virus persistence and not every persistent HPV infection necessarily leads to cancer transformation. The infection is usually naturally resolved by the immune system within two years after the transmission. Small proportion of infections becomes persistent and can cause precancerous lesions and subsequently cancer. For HNSCC, especially HPV infection of oral tissues (oHPV) is thought to be the most relevant. Oral HPV infection is estimated to be less prevalent than genital HPV, ranging from 0.7% to 11.4% depending on the study and population in focus (reviewed in Chung et al. (2014)).

2.5 Risk factors of HPV infection

Since HPV is sexually transmitted, main risk factors for the infection are associated with unsafe sexual behaviour. Risk factors for contracting HPV in general include having more than 6 sexual partners, higher frequency of sexual intercourse and earlier age of sexual debut (age 16 or younger) (Blomberg et al. 2011).

Oral HPV infection which can develop into HNSCC is more frequent in individuals with poor oral hygiene, smokers and the risk is also greater for people engaging in sexual activities involving oral-genital or oral-anal contact (Shi et al. 2014).

2.6 Risk factors of head and neck cancer

Both HPV^+ and HPV^- HNSCC are more frequent in men, but risk factors for HPV^+ and HPV^- HNSCC differ (summarised in Figure 2.5). However, not all studies differentiate between these two when assessing risk factors. Risk factors of HPV^+ HNSCC are of course closely associated with oHPV risk factors as described above.

Characteristics	HPV-Negative	HPV-Positive
Site	All sites	Tonsil, BOT
Risk factors	Tobacco/alcohol	Sexual behavior
Cofactors	Poor oral hygiene	Marijuana use
Age	Older cohorts	Younger cohorts
Sex	3:1 men	3:1 men
Incidence	Decreasing	Increasing
Stage	Variable	Early T stage, advanced N stage
Histology	Keratinized	Basaloid/poorly differentiated
p53	Mutated	Wild type
p16	Decreased expression	Increased expression

BOT = base of tongue; HPV = human papillomavirus.

Figure 2.5: Differences in HPV^+ and HPV^- HNSCC and their risk factors

Adapted from Lewis et al. (2015)

Systematic review by Chancellor et al. (2016) compiles risk factors found in 21 studies. Found risk factors for oropharyngeal cancer and oral cavity cancers include: higher number of sexual partners and experience with performing oral-genital sex. In some reviewed studies, homosexual relationships were identified as a risk factor. Further risk factors, which were however not as consistent among the reviewed studies, were young age of sexual debut, number of oral sexual partners, oral-anal contact (including the practice of oral-anal sex and the number of oral-anal sexual partners) and also experience with anal sex. As authors discuss, some risk factors such as absence of condom use and casual sexual relationships are unfortunately very rarely followed. In only one of the reviewed studies (D'Souza et al. 2007) association between oral cancers and these two factors was tested and were, indeed, found significant.

The number of studies in the Czech population is very limited. One case control study focusing on cohort of 86 patients with HNSCC and 117 controls by Tachezy et al. (2009) focused on demographic factors and the role of oral HPV infection. The study found association between HNSCC and presence of HR HPV, experience with oral-anal contact and tobacco and alcohol use.

2.7 Prevention of head and neck cancer and HPV infection

Prevention of HPV^- HNSCC, which is associated with lifestyle risk factors, should be focused on implementing anti-smoking policies and further educating the public about the risks associated with smoking and drinking. HPV^+ HNSCC is preventable by vaccination against HR HPV. For HPV infections and other STDs, primary prevention oriented on good sex education programs is also very relevant.

HPV infection can be prevented by vaccination by one of the three available vaccines. Bivalent vaccine (Cervarix) protects against infection with HPV-16 and HPV-18. This vaccine is recommended mainly for girls before their first sexual experience and recently, HPV vaccines were also recommended to young boys as a means of prevention of ano-genital carcinomas and lesions. As of July 2017, the

cost of the vaccine is covered by public health insurance in the Czech Republic when administered to girls or boys between ages 13–14 (Czech Vaccination Society ČLS JEP 2017).

Other vaccines are also available and actually used preferentially in some countries. For example, United Kingdom also administers quadrivalent and nonavalent vaccines besides Cervarix. Quadrivalent vaccine (Silgard) protects against infection with HPV-16 and HPV-18 (HR HPV types associated with majority of cancers), as well as against low risk types HPV-6 and HPV-11 (which cause genital warts). Nonavalent vaccine (Gardasil9) protecting against nine of the types is also available since 2015.

Even though HPV vaccines are not yet indicated as a means of prevention of head and neck cancers, they should protect against ano-genital as well as oral infections when administered to naïve individuals who have not yet encountered the viral infection (reviewed in D'Souza & Dempsey (2011)). Since the introduction of the vaccine, gradual decline of HPV associated cancers is expected. However, this will not be apparent for several years to come, because the development of cancers is a long process (Herrero et al. 2013).

2.8 Diagnostics and therapy of head and neck cancer

Vaccines prevent infection with the virus but cannot prevent the cancer progression. In cases when the infection leads to precancerous changes of the tissue and cancerous transformation, the available prevention of cervical cancer is cervical cancer screening which can detect cytological changes before the cancer progresses and therefore enable early (and more successful) treatment. While for cervical cancer, the screening program is implemented in most countries, screening for head and neck cancers meets several challenges. The precancerous stages are not as clearly defined and such screening has not yet been recommended.

Detection of the virus in the tumour tissue is the most reliable way to establish association between the infection and the tumour. Serological tests (i.e. tests

detecting antibodies against the virus) can only determine whether the individual encountered the virus at some point of his life, but do not specify when it was and in which anatomical location the infection occured. On the other hand, testing the presence of viral DNA in healthy tissue might lead to misleading results - mere virus presence does not necessarily lead to cancer and the infection can be resolved by the immune system.

Problems specific to HNSCC include inaccessibility of tonsillar crypts for routine smear tests (as is routinely done for cervical cancer screening), absence of early signs of the disease or only microscopic pathological changes in the tissue (reviewed in Lewis et al. (2015)). Therefore, further research and identification of early markers of precancerous stages is highly needed. Studies from 2013 and 2017 investigated the use of serological markers in prediction of subsequent HNSCC diagnosis (Kreimer et al. 2013, 2017). According to their results, HPV antibody levels (specifically anti-HPV-16) were elevated at least ten years prior to HNSCC diagnosis and could serve as a prognostic marker in screening.

Good reason for HPV testing in HNSCC patients is that HPV^+ HNSCC patients have better prognosis and survival. HPV status of the tumour itself is strong positive prognostic marker and it might be beneficial to revise management guidelines and implement less aggressive treatment options for those with HPV^+ tumours (reviewed in Kimple & Harari (2015)).

2.9 Significance of the research topic

Changes in HNSCC epidemiology and massively increasing incidence of HPV^+ HNSCC are evident in the last three decades. Much is known about potential risk factors for both HPV^+ and HPV^- groups, but the population at risk is still not routinely screened. Even though this issue is a currently extensively studied topic, much still remains to be uncovered. In the Czech Republic, only limited number of studies was conducted. This thesis should contribute to the current state of knowledge about HNSCC specifics in the Czech Republic.

3. Hypothesis and aims

Currently, most prevention programs are aimed at genital HPV infections and associated cancers (cervical cancer), but further knowledge for implementing such course of action for oHPV has to be investigated. Main goal of this thesis is to identify behavioural risk factors of oHPV infection and of head and neck cancer.

First part of the thesis aims to demonstrate the epidemiological trends of HNSCC in the Czech Republic. By demographic analyses of publicly available data, we aim to illustrate that the population at risk changes as has been documented in other countries. Our goal here is to answer the following questions:

1. How does mortality from HNSCC in Czech population change overtime?
2. Do age-specific death rates from HNSCC change overtime in Czech population as we would expect based on epidemiology trends described in other populations?
3. How is HNSCC incidence and mortality distributed in the regions of the Czech republic?

In the second part of the thesis, we conduct cross-sectional and case-control study with 485 participants. From this cohort, we obtained questionnaires and clinical data which will be analysed to answer the following:

1. What risk factors are associated with oHPV infections?
   oHPV infections are transmitted through oral sexual contacts. Therefore, we hypothesise that the risk factors identified in this sections will be consistent with known mechanisms of infection. We expect that oral sexual contacts and higher number of sexual partners will be associated with the infection.
2. What risk factors are associated with HNSCC?
   We hypothesise that HR HPV and previously described risk behaviours (e.g. smoking, risk sexual behaviours and alcohol abuse) will be associated with HNSCC diagnosis.

4. Data and methods

4.1 Data

4.1.1 Data on HNSCC mortality and population

Publicly available data on mortality from HNSCC accessible in WHO Mortality Database (WHO 2016) were used for the first part of this thesis. CoDQL (Cause of Death Query online) web interface was used to gather all necessary data. Data used contain information about number of deaths by diagnosis ICD-10: C00–C14 for both sexes and all ages in the Czech Republic in the years 1986–2013.

Data for year 1993 and prior were only available coded in the 9th revision of ICD in Basic Tabulation List (BTL). In order to convert diagnoses ICD-10: C00–C14 to ICD-9 BTL, online conversion tool between the ICD-9 and ICD-10 revisions (Health Fusion, Inc 2017) and online conversion table of detailed ICD-9 (diagnoses ICD-9:140–149) codes from BTL diagnoses (B08 - malignant neoplasm of lip, oral cavity and pharynx) were used (Wolfbane Cybernetic Ltd. 2001).

Population statistics for both sexes and all ages between years 1986–2013 were also extracted from CoDQL (WHO 2016).

For comparison of mortality when using different definition of HNSCC, data for diagnoses ICD-10: C00–C06; C09–C10 were filtered out of the previously accessed data.

In addition to the original 5-year age groups in the accessed data, two new age categories for ages 1–4 and 85+ were created.

4.1.2 Data on HNSCC mortality in regions of the Czech Republic

Data on mortality from HNSCC (ICD-10: C00–C14) in Czech regions (2014) were extracted from Czech Statistical Office (2016). Data were used for creating the map of HNSCC mortality of men and women in the regions of the Czech Republic.

4.1.3 Data on HNSCC incidence

Data on HNSCC incidence in 2014 (ICD-10: C00–C14) were extracted from Czech National Cancer Registry (IHIS CR 2014). Data were used for comparison of incidence between males and females in the Czech Republic and for creating the map of HNSCC incidence in the regions of the Czech Republic. Data on the age-specific incidence in 1986 were extracted from Dušek et al. (2005) and used for comparison of the age-specific incidence in the years 1986 and 2014.

4.1.4 Clinical data for risk factor analyses

Data used for analyses of risk factors for HPV infections and HNSCC were gathered by National Reference Laboratory (NRL) for Papillomaviruses and Polyomaviruses in collaboration with Dept. of Otorhinolaryngology (ORL) and Head and Neck Surgery, Faculty Hospital Motol in years 2000–2010. Participants (n=485) were recruited either as patients treated for HNSCC (ICD-10: C00–C06; C09–C10) or controls that were not treated for HNSCC but were also recruited in the ORL clinic. Controls were chosen to match the patient cohort in frequency of age and sex. Enroled participants were asked to fill out personal questionnaire on risk behaviours. Biological samples for serological exam (i.e. serum for the detection of antibodies against HPV) and PCR test of oral washing (i.e. HPV DNA presence in oral cavity) were recovered by hospital personnel and tested by members of NRL for Papillomaviruses and Polyomaviruses. All participants signed informed consent and confirmed that they participate voluntarily, that they can withdraw their consent and enrolment in the study at any time without changes to their treatment plan and are not obliged to give their reason(s) for withdrawal.

The questionnaire contained questions on age, gender, completed years of education and three sections of questions on risk behaviours: alcohol consumption, tobacco use and sexual activities. Sex and gender were used as interchangeable terms throughout this thesis.

Alcohol consumer was defined as person drinking alcoholic beverages at least

once a week on average for at least a year[1]. Participants were also asked how many beers (0.5l), wine (0.2l) and liquor (0.05l) drinks they drink per week.

Smoker in our questionnaire was defined as person "smoking regularly at least half a pack of cigarettes per week for at least one year"[2]. Participants were also asked whether they currently smoke, at what age they stopped smoking (if so) and at what age they started. The same set of questions was also asked for other means of tobacco use: cigar and pipe smoking and tobacco chewing and sniffing.

Participants were asked about the number of their sexual partners, whether they have experience with oral-oral (kissing), oral-genital and oral-anal contact and if so, at what age they first experienced oral-genital and oral-anal contact. At the beginning of this set of question, participants were again ensured about anonymity and confidentiality of the collected data. At this point, it was also explained why these questions are relevant to the study (i.e. sexually transmitted virus is an etiological agent of head and neck cancers and other diseases).

[1]Original question in Czech: "Pil(a) jste někdy v průměru více než jednou týdně alkoholické nápoje po dobu alespoň jednoho roku a déle?"

[2]Original question in Czech: "Kouřil(a) jste někdy pravidelně cigarety, alespoň půl balíčku za týden po dobu jednoho roku nebo déle?"

4.2 Methods

4.2.1 Demographic analysis

Age-standardised death rate was chosen to demonstrate changes in the epidemiological trends of HNSCC mortality. Age-standardised death rate was calculated for each year between the years 1986–2013 as follows:

$$\mathrm{SDR} = \frac{\sum_0^{\omega-1} m_x \times P_x^{st}}{\sum_0^{\omega-1} P_x^{st}} \quad \text{, where}$$

SDR = age-standardised death rate for each sex and 1 year in the Czech Republic, normalised to 100, 000;

ω is the first age that ano one in population reaches;

m_x = age-specific death rate for HNSCC;

P_x^{st} = number of living persons in specific age in standard population

Age-specific death rate (m_x) was calculated as follows:

$$m_x = \frac{D_x}{P_x} \quad \text{, where}$$

D_x = number of deaths caused by HNSCC in specific age;

P_x = number of living persons in specific age

Direct standardisation using New European Standard (2013) was performed to calculate standardised death rates.

All graphs of demographic data were generated in Microsoft® Excel® 2016 MSO (16.0.7726.1042) 32-bit. Because of the higher incidence and death rates in men, the y axis scale in graphs differs between the two sexes. All maps were generated in ArcMap 10.4 (10.4.1). Regions of the Czech Republic are labeled as in Figure A.1 (see Appendix A).

4.2.2 Clinical data preprocessing

Participants whose oral washing status (positive/negative) was unknown were excluded (n=4) because this information was essential to our analyses. Unknown values for other variables were marked as "NA" (not available), but participants with NA values were kept in dataset. This may cause discrepancies for frequencies in some variables.

Positivity for any HR HPV in oral washing was computed as new variable: individuals were labeled as positive if any HR HPV type (see Table 2.2) was present in their oral washing.

For the risk factor analyses, all participants were grouped pooled and then categorised either according to the patient/control status or according to their oHPV staus - $oHPV^+$ or $oHPV^-$ (based on the result of examination of HPV presence in oral washings).

New variable 'pack-years' was computed. One pack-year equals to 1 year of smoking 1 pack of cigarettes (20 cigarettes) per day. Number of drinks per week was calculated from reported number of beers (one drink equals 0.5l), wine (one drink equals 0.2l) and liquor (one drink equals 0.05l) drank on average by participant per one week.

4.2.3 Statistical analysis

All statistical analyses were carried out in R version 3.3 R Core Team (2017).

Basic statistical tests (t-test, χ^2 test or Pearson correlation coefficient) were computed using psych package (Revelle 2017). Logistic regression models were fitted with generalized linear model (R base glm function). Despite common practice of using stepwise model selection in demographic studies, statistical literature argues against this approach because it yields oversimplified and biased coefficients (Harrell 2010). Therefore all relevant variables were entered into the model without stepwise selection of predictors. Regression model's pseudo R-squared were obtained using pscl package using Nagelkerke-Cragg & Uhler's measure of goodness of fit (Cragg & Uhler 1970, Nagelkerke 1991). Plots were generated using ggplot2 package (Wickham 2009).

5. Results

In this thesis, we used two sets of data: In the first part, we used publicly available data on mortality and incidence for diagnoses ICD-10: C00–C14 and equivalent diagnoses in ICD-9. In second part of the thesis, the clinical data used concern only patients with diagnoses ICD-10: C00–C06; C09–C10 (cancer of the oral cavity and oropharynx).

This discrepancy is caused firstly by complications with ICD coding conversion when accessing older data coded in ICD-9 and combining them with data for specific diagnoses coded in ICD-10. We decided to use diagnoses ICD-10: C00–C14 which were routinely used in other studies and are easily convertible to their ICD-9 BTL equivalents (see methods section) in the demographic analyses. Secondly, clinical data that we used for the analysis of risk factors originated from study performed by the National Reference Laboratory for Papillomaviruses and Polyomaviruses.

To demonstrate that the two sets of diagnoses follow similar epidemiological trends and that the difference is marginal, we show standardised death rates for men and women in the Czech Republic for both sets of diagnoses (Figure 5.1).

While the trend is similar for both variants of the diagnosis, diagnosis ICD-10: C00–C14 has overall higher standardised death rate, which is not surprising considering it encompasses the ICD-10: C00–C06; C09–C10 diagnoses plus additional diagnoses C07–C08 and C11–C12.

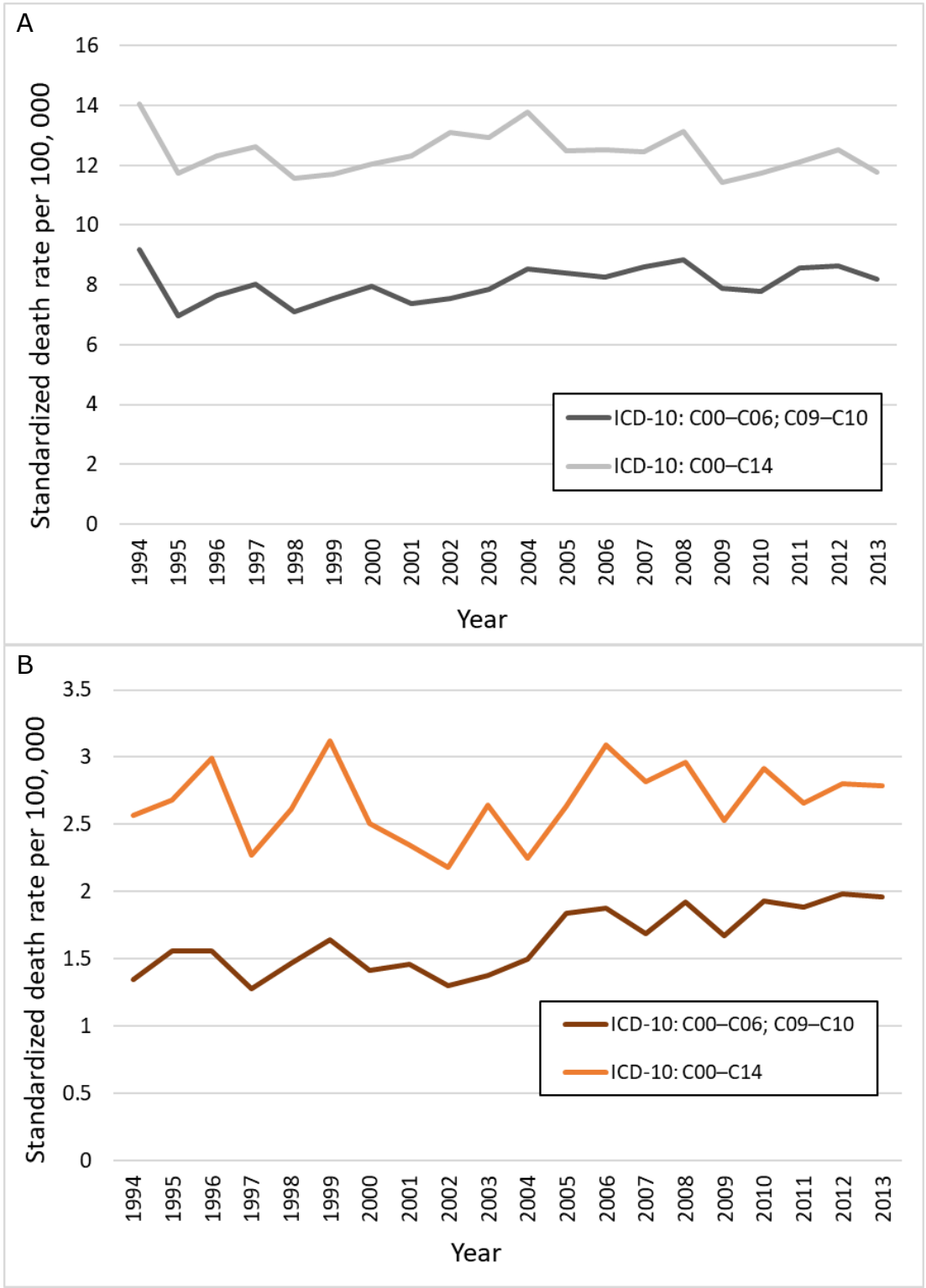

Figure 5.1: Comparison of standardised death rates per 100, 000 population for ICD-10: C00–C14 and ICD-10: C00–C06; C09–C10 in the Czech Republic, men (A) and women (B), years 1994–2013

Author's own work based on data from WHO mortality database

5.1 Epidemiological trends of head and neck cancer in the Czech Republic

5.1.1 Mortality from head and neck cancer in Czech population

According to available literature, head and neck cancer mortality does not decrease over the years. To confirm this trend in the Czech Republic, we calculated standardised death rates for Czech men and women for the years 1986–2013 (Table 5.1). We indeed found, that the data confirm the trend described in literature - overall death rate does not decline. The standardised death rate for women is lower than for men (Figure 5.2).

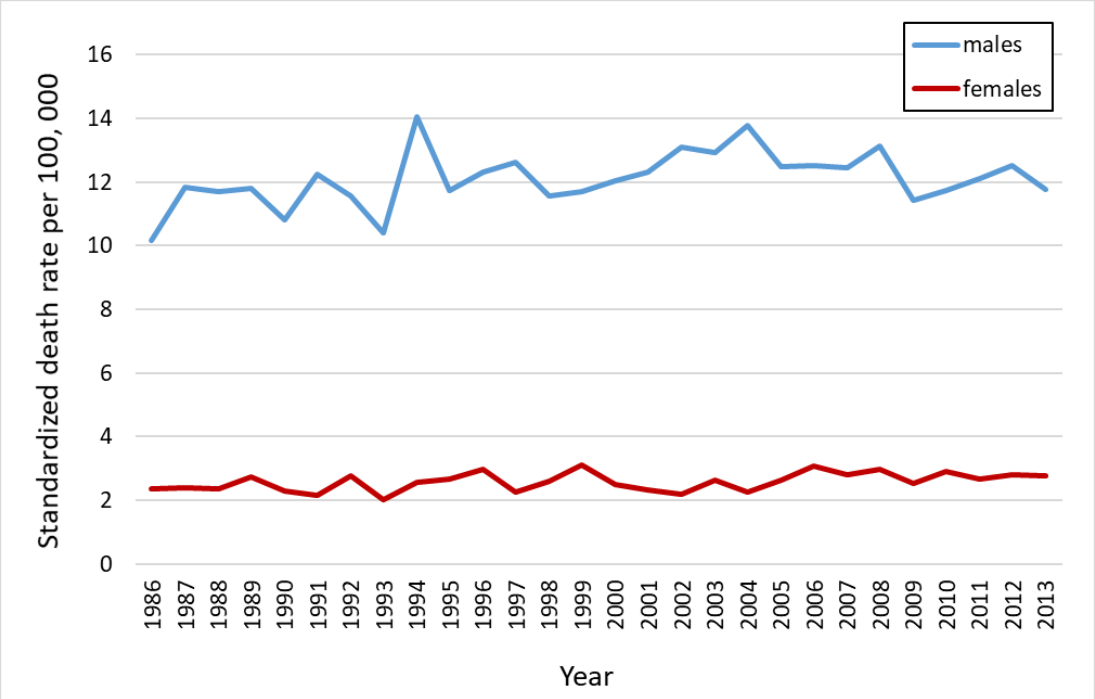

Figure 5.2: Standardised death rates from HNSCC per 100, 000 population in the Czech Republic, men and women, years 1986–2013

Author's own work based on data from WHO mortality database

Table 5.1: Standardised death rates from HNSCC per 100, 000 population in the Czech Republic, men and women, years 1986–2013

Data source: WHO mortality database

Year	Men	Women
1986	10.17	2.35
1987	11.83	2.41
1988	11.71	2.35
1989	11.81	2.75
1990	10.81	2.29
1991	12.26	2.17
1992	11.58	2.76
1993	10.40	2.03
1994	14.06	2.56
1995	11.75	2.68
1996	12.33	2.99
1997	12.64	2.27
1998	11.57	2.62
1999	11.69	3.12
2000	12.05	2.51
2001	12.32	2.35
2002	13.10	2.18
2003	12.92	2.64
2004	13.78	2.25
2005	12.49	2.64
2006	12.53	3.09
2007	12.47	2.82
2008	13.14	2.97
2009	11.42	2.53
2010	11.72	2.91
2011	12.11	2.66
2012	12.50	2.80
2013	11.79	2.79

The currently accepted hypothesis why the overall death rate does not decline is that now significant proportion of head and neck cancer cases is caused by infectious agent (HPV) rather than traditional risk factors (tobacco and alcohol). This results in shift of the incidence to younger population since the 1980's. We calculated head and neck cancer age-specific death rates (Table 5.2) for men and women in the Czech Republic in years 1986 and 2013. The goal was to investigate the changes in mortality distribution among age groups.

We saw changes in the age-specific death rates in both men and women. While in 1986, the majority of deaths by head and neck cancers occured in people 75 years old and older, it was predominant in younger population in 2013. This is especially apparent in men (Figure 5.3A), but similar pattern can be seen in women as well (Figure 5.3B), particularly in category 85+. Age-specific death rates in 2013 are higher in younger age categories (55–70) for both sexes.

We chose two age groups to better illustrate the trend: category 65–69 represents the younger age risk groups and category 75–79 represents the older group. For both men (Figure 5.4A) and women (Figure 5.4B) we see increase of death rate in the younger group while the death rate in the older group slightly declines. We also looked at data for younger age groups (e.g. 50–55). Contrary to our expectations and trends observed in other populations, the data did not show the same trend as in age group 65–59 (not shown).

Table 5.2: Age-specific death rates from HNSCC per 100, 000 population, Czech Republic, men and women, years 1986 and 2013

Data source: WHO mortality database

Age group	Men		Women	
	1986	2013	1986	2013
0	0	0	0	0
1 – 4	0	0	0	0
5 – 9	0	0	0.3	0
10 – 14	0	0	0	0
15 – 19	0	0	0	0
20 – 24	0	0	0	0
25 – 29	0	0	0	0
30 – 34	0.5	0	0.3	0
35 – 39	2.1	1.0	0.5	0
40 – 44	5.9	2.0	0.5	1.1
45 – 49	11.9	11.7	0.7	1.7
50 – 54	16.3	13.4	1.1	2.8
55 – 59	20.0	35.1	3.6	5.7
60 – 64	20.6	35.2	1.9	6.3
65 – 69	18.5	34.8	3.3	8.8
70 – 74	19.8	31.1	5.8	6.0
75 – 79	25.9	21.7	8.9	4.4
80 – 84	35.3	25.0	8.3	11.7
85+	45.7	18.5	29.5	15.5

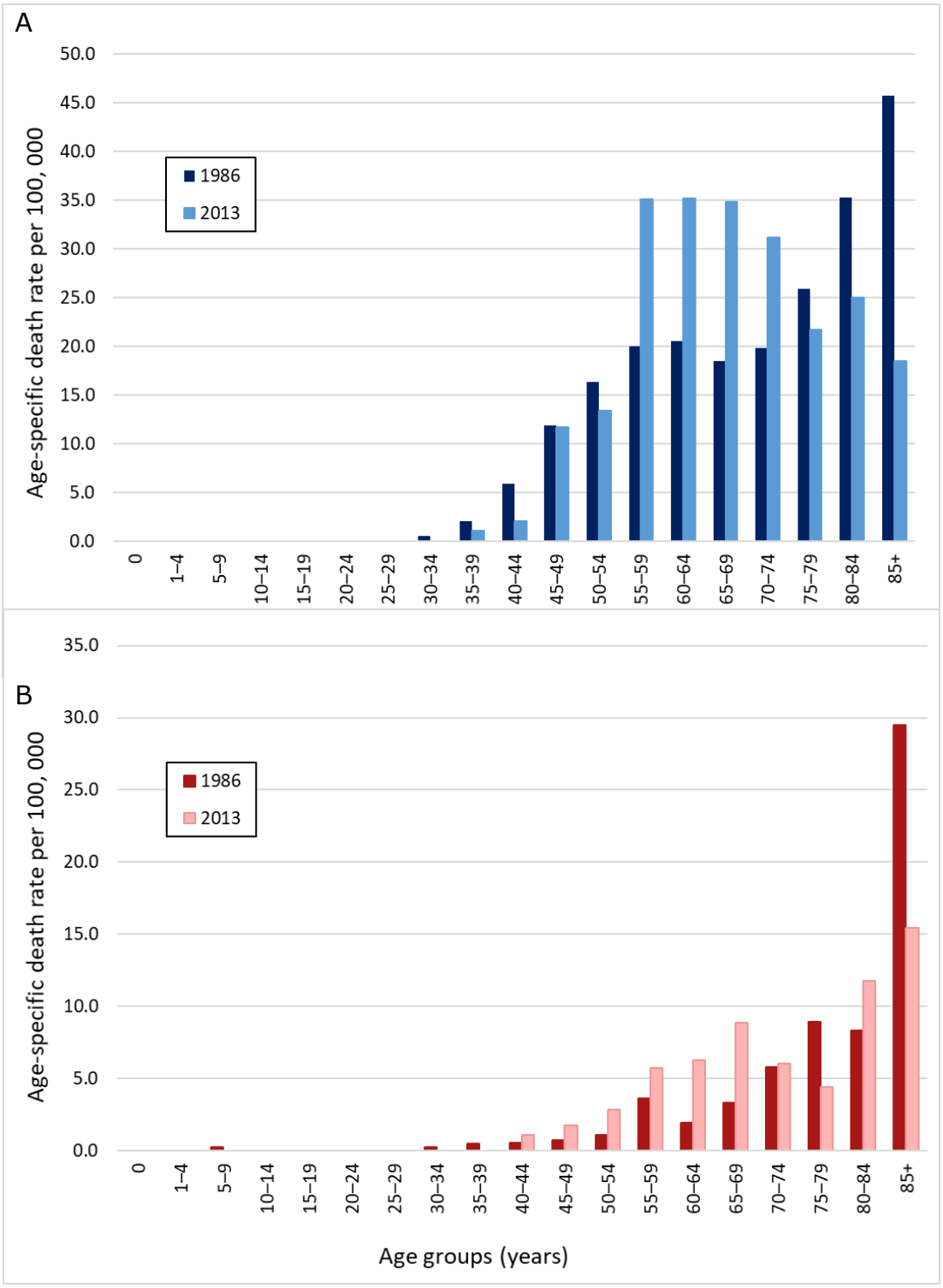

Figure 5.3: Age-specific death rates from HNSCC per 100, 000 population, Czech Republic, men (A) and women (B), years 1986 and 2013

Author's own work based on data from WHO mortality database

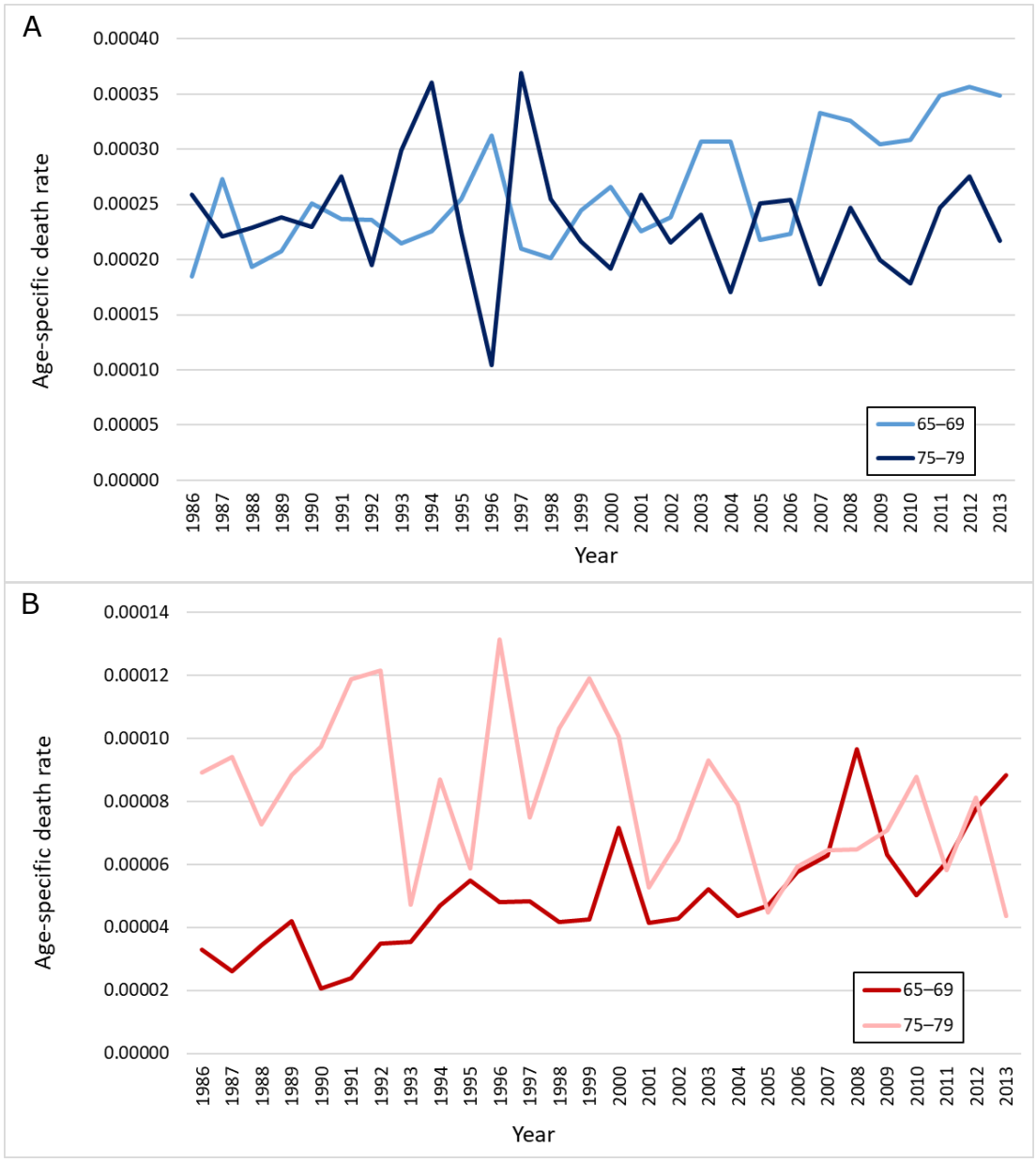

Figure 5.4: Age-specific death rates from HNSCC in two age categories (65–69 and 75–79), Czech Republic, men (A) and women (B), years 1986–2013

Author's own work based on data from WHO mortality database

5.1.2 Incidence of head and neck cancer in the Czech Republic

Differences in death rates do not have to necessarily reflect the incidence of a disease. To complete the picture we also show data on age-specific incidence of HNSCC in the Czech Republic (Table 5.3 and Figure 5.5). It is also higher in men and peaks in age group 60–69 for both sexes.

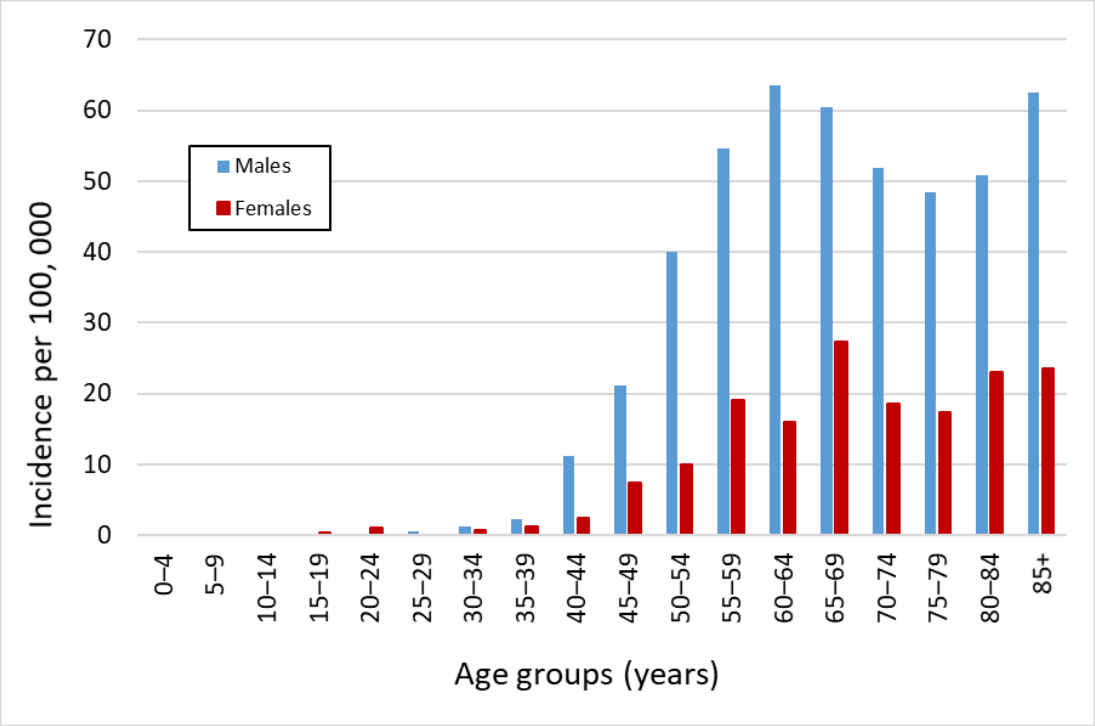

Figure 5.5: Age-specific incidence of HNSCC per 100, 000 in age groups, Czech Republic, men and women, 2014

Author's own work based on data from WHO mortality database

When we compared age-specific incidence in years 1986 and 2014 (Table 5.3), we see similar differences as we saw in death rates (Figure 5.3). In men, we see that incidence is higher in age groups 50–75 in the year 2014 than it was in the year 1986 (Figure 5.6A). In women, the overall incidence is higher in the year 2014 but the increase is again more apparent in age groups 50–75 (Figure 5.6B).

Table 5.3: Age-specific incidence of HNSCC per 100, 000 population, Czech Republic, 1986 and 2014

Data source: Czech National Cancer Registry

Age group	Men		Women	
	1986	2014	1986	2014
0–4	0	0	0	0
5–9	0	0	0	0
10–14	0	0	0	0
15–19	0.56	0	0.58	0.40
20–24	0.28	0	0.29	1.00
25–29	0.58	0.60	0.31	0
30–34	1.20	1.30	0.50	0.80
35–39	6.24	2.30	0.72	1.30
40–44	14.20	11.20	3.24	2.50
45–49	24.09	21.10	4.58	7.40
50–54	35.02	40.10	4.77	9.90
55–59	34.84	54.70	7.24	19.00
60–64	35.20	63.50	7.57	16.00
65–69	38.20	60.40	6.11	27.30
70–74	41.80	51.80	7.26	18.50
75–79	44.33	48.50	14.69	17.40
80–84	47.69	50.90	12.01	23.00
85+	57.02	62.50	14.72	23.60

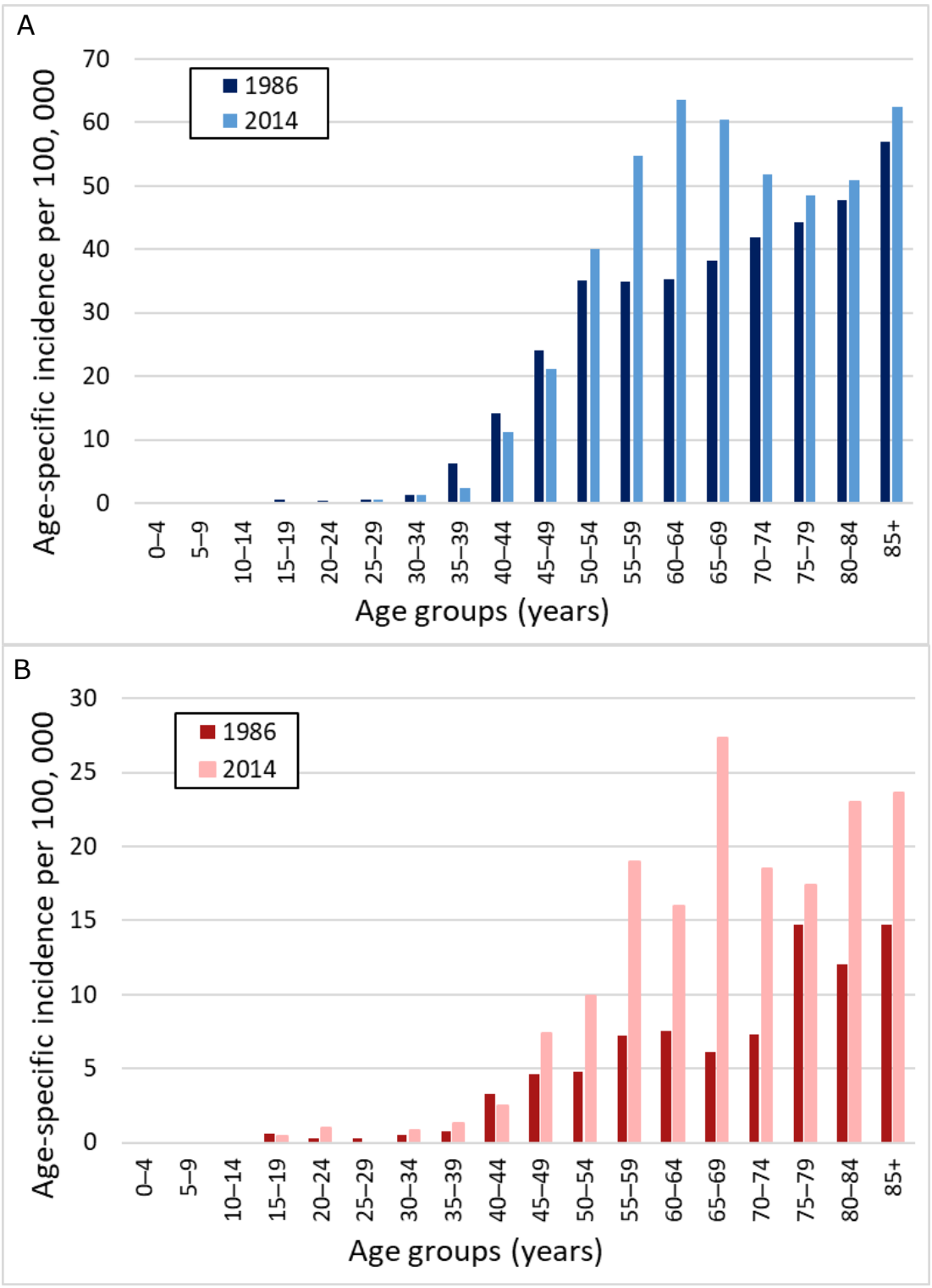

Figure 5.6: Age-specific incidence of HNSCC, Czech Republic, men (A) and women (B), 1986 and 2014

Author's own work based on data from WHO mortality database

5.1.3 Head and neck cancer incidence and mortality in regions of the Czech Republic

We also looked at the distribution of HNSCC incidence in regions of the Czech Republic (Figure 5.7). We see that the region with highest incidence in men are Zlín Region and Karlovy Vary Region. In women, Zlín Region is actually one of the regions with the lowest incidence whereas the highest incidence is in Hradec Králové Region and Prague. However, it is important to note that the differences of incidence between the regions are very small - ranging from 5 cases per 100, 000 in women to 18.8 cases per 100,000 in men (Table 5.4).

Table 5.4: Incidence of HNSCC per 100, 000 population in regions of the Czech Republic, men and women, 2014

Data source: Czech National Cancer Registry

Region	Cases of HNSCC (per 100, 000 men)	Cases of HNSCC (per 100, 000 women)
Prague	20.0	10.5
South Bohemian Region	16.2	7.7
South Moravian Region	19.9	9.5
Karlovy Vary Region	27.0	9.2
Hradec Králové Region	22.9	10.3
Liberec Region	22.3	7.2
Moravian-Silesian Region	25.8	8.5
Olomouc Region	15.4	6.5
Pardubice Region	23.9	9.2
Plzeň Region	21.8	8.3
Central Bohemian Region	18.3	7.4
Ústí nad Labem Region	19.8	9.4
Vysočina Region	18.2	5.5
Zlín Region	34.2	6.3

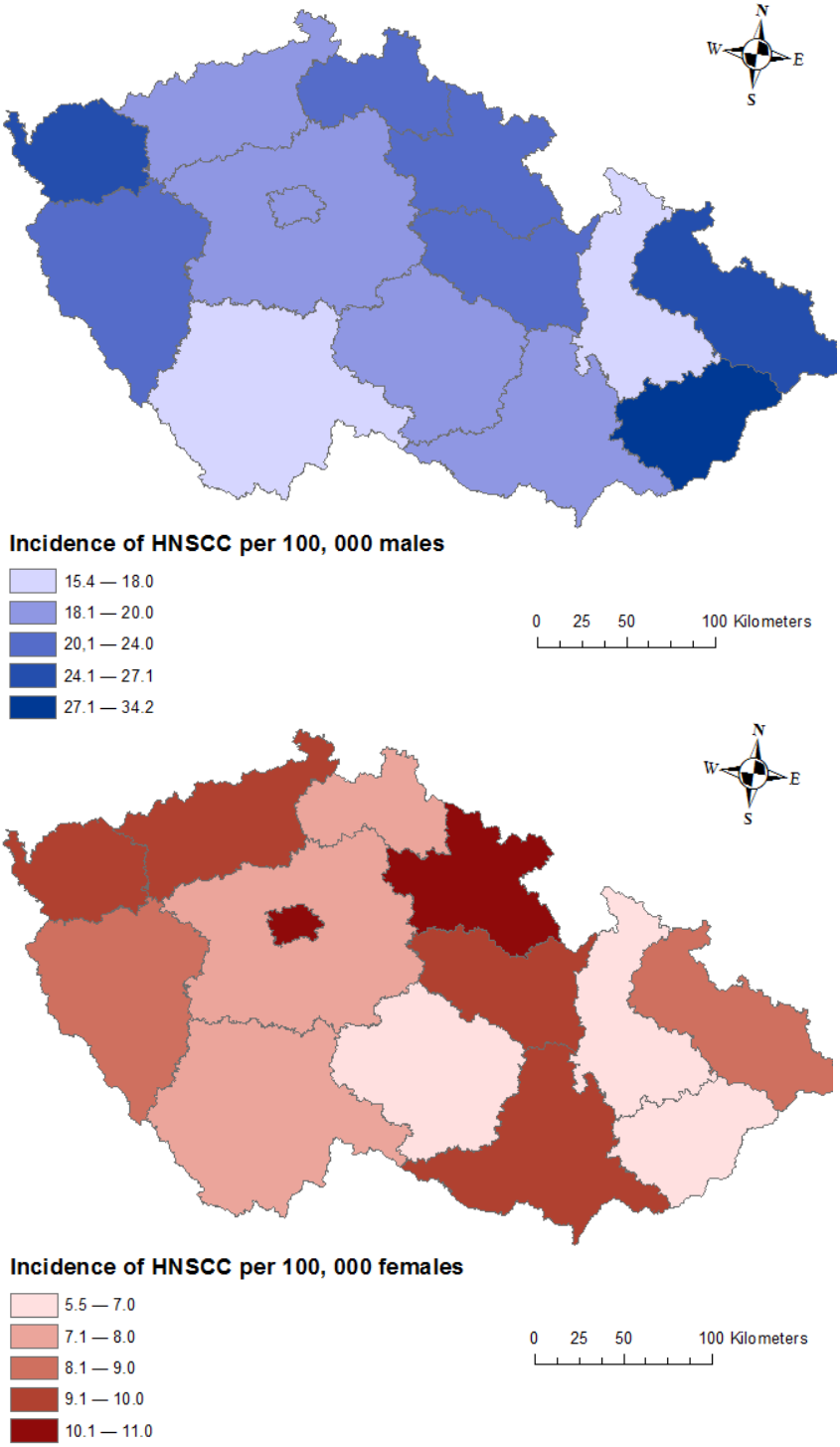

Figure 5.7: Incidence of HNSCC per 100, 000 population in regions of the Czech Republic, men (top map) and women (bottom map), 2014

Author's own work based on data from Czech National Cancer Registry, generated in ArcMap

When we look at the map depicting death rates in regions of the Czech Republic (Figure 5.8, data shown in Table 5.5), it resembles the pattern seen on the map of incidence. Similarly, we also saw random differences in the distribution of death rate for the two sexes among European countries as was shown in Figure 2.2.

Table 5.5: Death rate from HNSCC per 100, 000 population in regions of the Czech Republic, men and women, 2014

Data source: Czech National Cancer Registry

Region	Deaths (per 100, 000 men)	Deaths (per 100, 000 women)
Prague	8.90	3.41
South Bohemian Region	8.28	2.79
South Moravian Region	12.74	2.68
Karlovy Vary Region	10.13	3.95
Hradec Králové Region	8.48	2.49
Liberec Region	6.04	4.03
Moravian-Silesian Region	12.89	3.86
Olomouc Region	10.29	3.38
Pardubice Region	6.28	2.68
Plzeň Region	10.91	3.10
Central Bohemian Region	11.45	1.96
Ústí nad Labem Region	12.00	3.6
Vysočina Region	15.01	1.95
Zlín Region	16.75	1.34

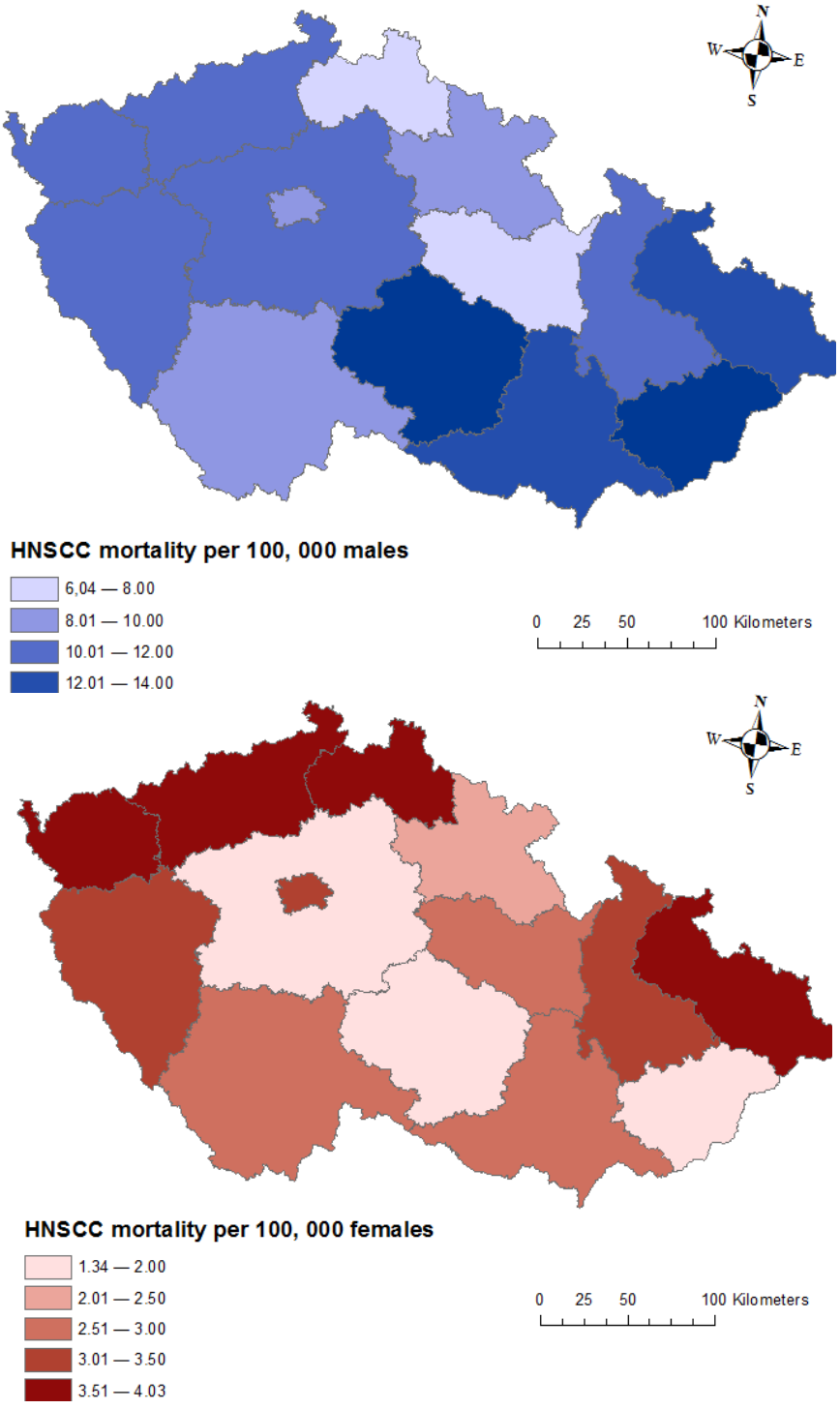

Figure 5.8: Mortality from HNSCC per 100, 000 population in regions of the Czech Republic, men (top map) and women (bottom map), 2014

Author's own work based on data from Czech National Cancer Registry, generated in ArcMap

5.2 Risk factors of head and neck cancer and oral HPV infection

Assessing the HPV etiology of significant subgroup of head and neck cancers is important for prevention as well as in therapy. In future, the therapeutical approach for HPV-associated and non-HPV-associated cancers might change due to different prognosis in each subgroup. Also, other ways of prevention of head and neck cancers are available. HPV infection can be prevented by vaccination and HPV presence can be tested and even routinely screened for if suitable target population is identified.

For these reasons, we conducted cross-sectional study to assess risk factors of oral HPV infection as well as case-control study to analyse the risk factors of head and neck cancers with data collected by NRL for Papillomaviruses and Polyomaviruses. The schema of the study design is shown in Figure 5.9.

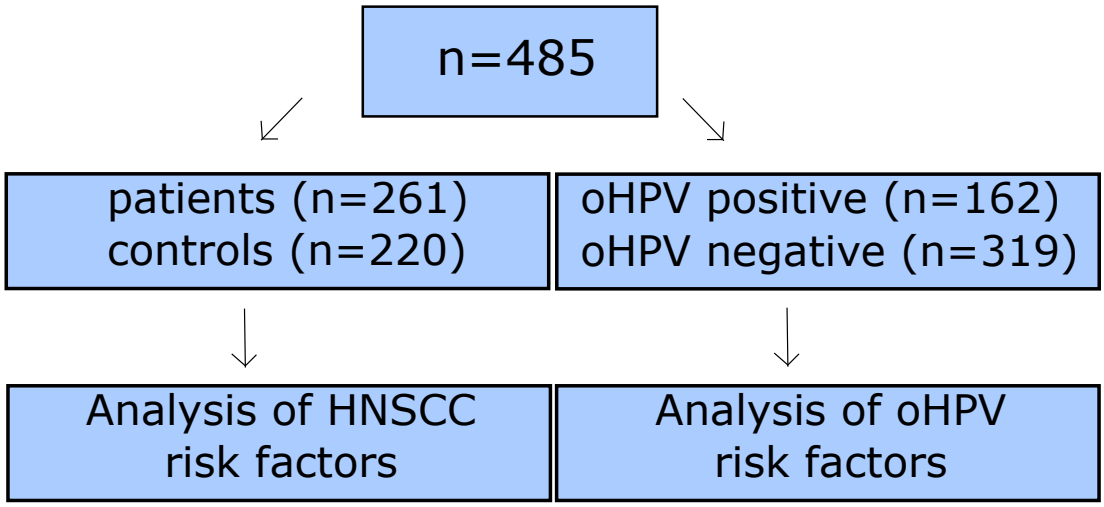

Figure 5.9: Schema of study design

Author's own work

5.2.1 Characteristics of participants

Data from 481 participants were used for the analyses of risk factors of oHPV infection and HNSCC. 261 participants (54.26%) were recruited as patients treated for HNSCC and 220 (45.74%) as controls (without HNSCC diagnosis). The mean age of participants is 56.54 years (SD = 10.05), they completed on average 12.97 years of education (SD = 2.97) and most of them were males (75.68%). Majority of participants were also smokers (72.07%) and alcohol consumers (72.36%). Most of them did not have HPV DNA present in their oral washing (i.e. are $oHPV^-$) (66.32%). Out of 162 participants with positive result of HPV DNA test ($oHPV^+$), 85.19% (n = 138) were positive for any of the HR HPV types (defined in section 2.4).

Average number of sexual partners of participants was 9.64 (SD = 14.08), but 34 participants did not disclose this information. Most of the participants had experience with oral-oral (88.57%) contact (i.e. kissing) and majority also with oral-genital contact (67.98%). Only few of the participants had experience with oral-anal contact (8.32%) and 28 participants did not disclose this information.

Main characteristics of the data set are summarized in Table 5.6. Distribution of continuous variables in the dataset is depicted in Figure 5.10 and frequencies of selected categorical factors in Figure 5.11.

Table 5.6: Characteristics of participants

Data source: NRL for Papillomaviruses and Polyomaviruses

	n=481* (% or SD)
Mean age	56.54 (SD = 10.05)
Mean education	12.97 (SD = 2.97)
Gender	
males	364 (75.68%)
females	116 (24.12%)
HPV presence in oral wash	
positive (oHPV$^+$)	162 (33.68%)
negative (oHPV$^-$)	319 (66.32%)
Patient/control status	
patient	261 (54.26%)
control	220 (45.74%)
Smoking	
smokers	338 (70.27%)
non-smokers	131 (27.23%)
Mean pack-years	27.84 (SD = 20.45)
Ever consumed alcohol	
alcohol consumers	330 (68.61%)
life long abstinence	139 (28.90%)
Number of sexual partners	9.64 (SD = 14.08)
Experience with oral-oral contact	
yes	426 (88.57%)
no	32 (6.65%)
Experience with oral-genital contact	
yes	327 (67.98%)
no	129 (26.82%)
Experience with oral-anal contact	
yes	40 (8.32%)
no	413 (85.86%)

* Where percentages do not add up to 100%, some values were missing due to non-response.

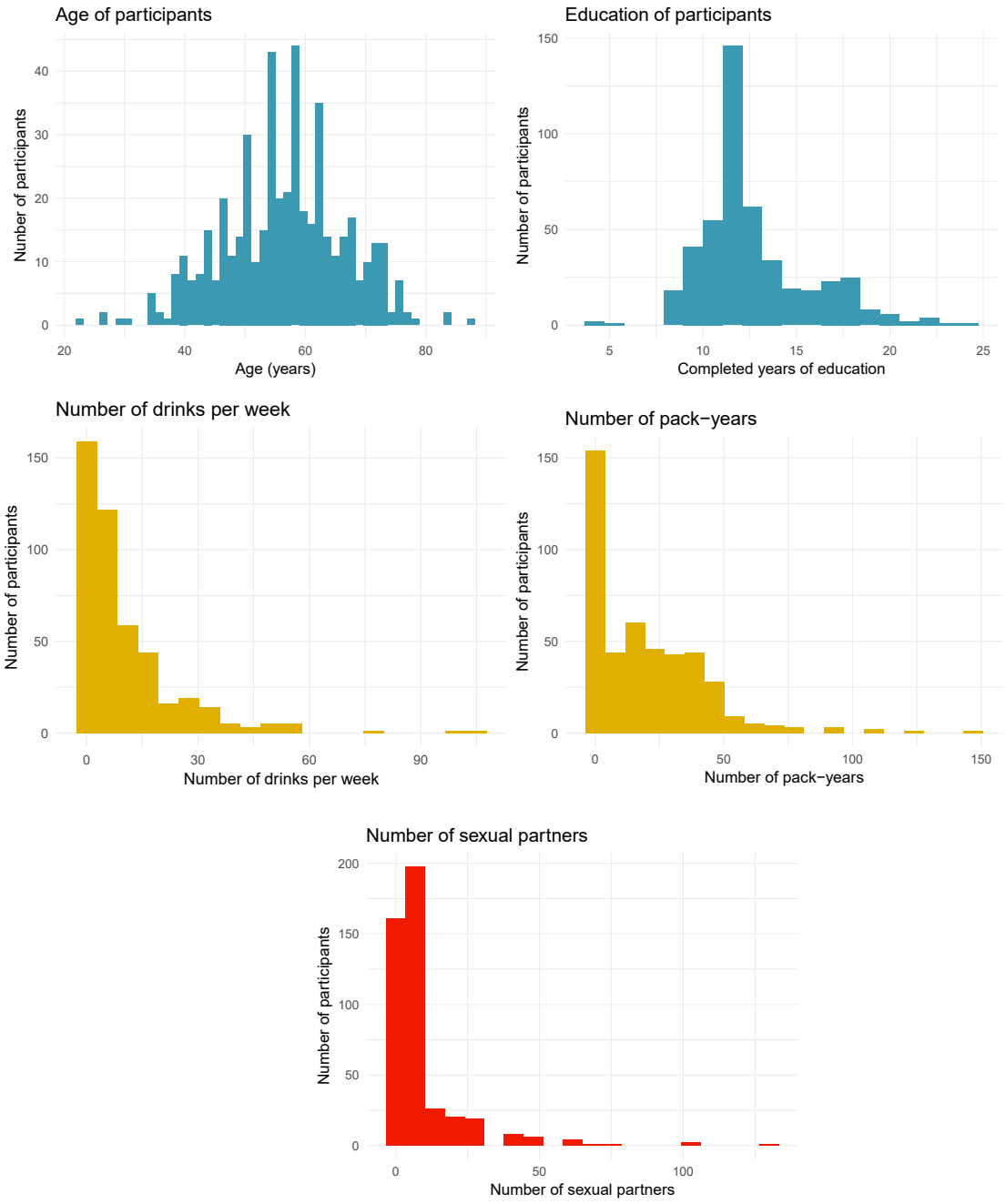

Figure 5.10: Distribution of continuous variables in the dataset

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

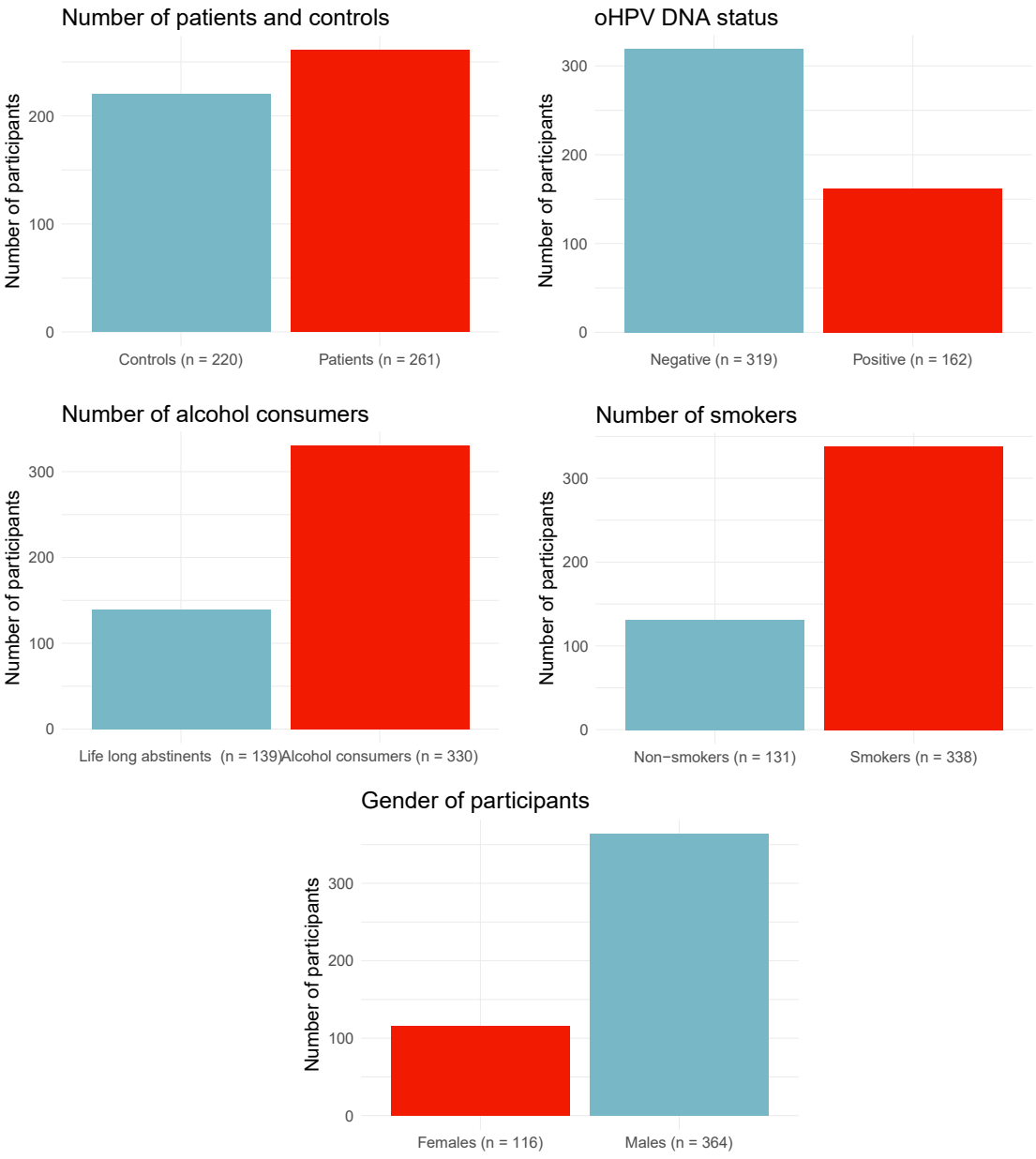

Figure 5.11: Frequencies of categorical factors in the dataset

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

5.2.2 Risk factors of oral HPV infection

For the analyses of risk factors of HPV infection, we divided participants (n=481) into two groups: those who tested negative (n=319) for HPV DNA in their oral washings (oHPV$^-$) and those who tested positive (n=162) (oHPV$^+$). Before the final analysis of risk factors with logistic regression, we inspected individual associations between oHPV status and demographic and behavioural characteristics of the participants.

Association between gender and oHPV infection

Since the majority of participants overall were men, we saw similar distribution in both oHPV$^+$ and oHPV$^-$ categories Figure 5.12. Gender did not affect oHPV infection ($\chi^2(1) = 0.36$, p = 0.55).

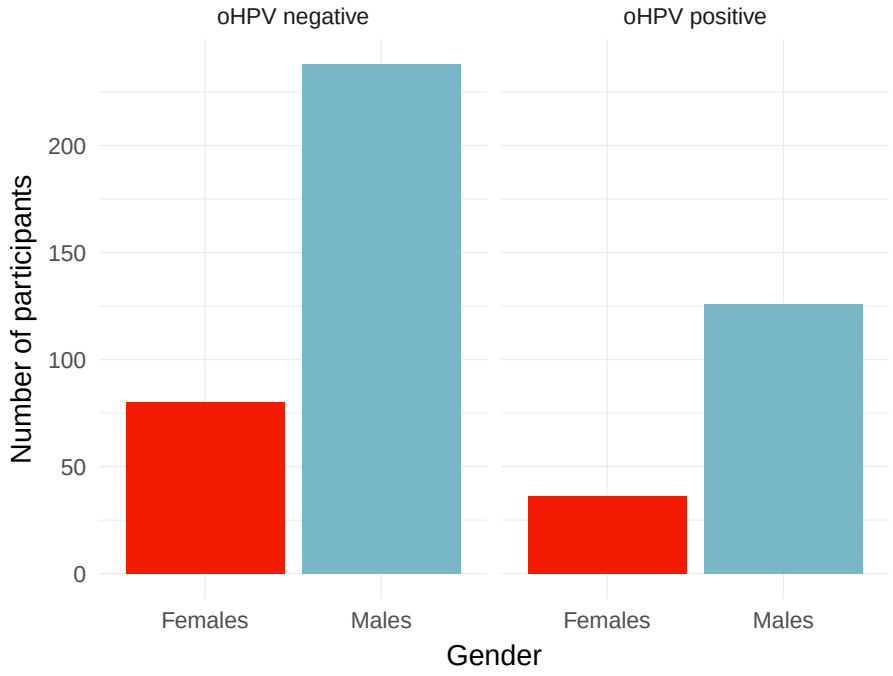

Figure 5.12: Gender of oHPV$^+$ and oHPV$^-$ participants

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

Association between age and oHPV infection

Mean age of oHPV$^+$ group was 55.86 years (SD = 10.35), while mean age of the oHPV$^-$ group was 57.86 years (SD = 9.33) (Figure 5.13), which is a significant difference (t(355.46) = -2.14, p = 0.03).

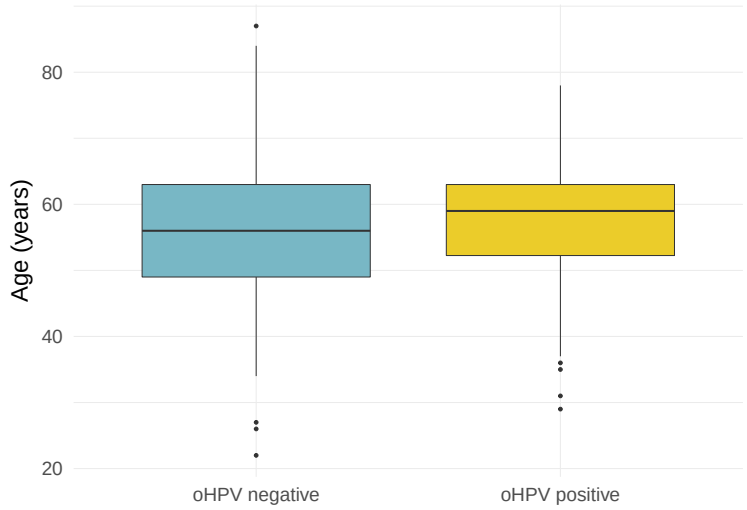

Figure 5.13: Age of oHPV$^+$ and oHPV$^-$ participants

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

Association between education and oHPV infection

Mean length of completed education of oHPV$^+$ group was 13.18 years (SD = 2.84), while length of completed education of the oHPV$^-$ group was 12.86 years (SD = 3.04). Distribution of completed years of education in both groups is depicted in Figure 5.14 We have not found significant effect of education on oHPV infection (t(342.62) = -1.13, p = 0.26).

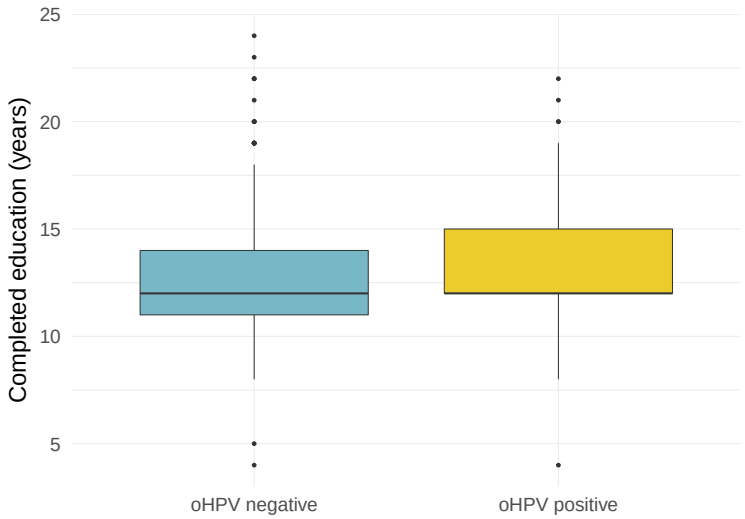

Figure 5.14: Completed years of completed education of oHPV$^+$ and oHPV$^-$ participants

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

Association between alcohol consumption and oHPV infection

Majority of participants in our sample responded that they do consume alcohol. Alcohol does not affect oHPV infection ($\chi^2(1) = 0.14$, 0.70). We also tested for the effect of consumption of beer (t(279.15) = 0.64, p = 0.5231), wine (t(186.84) = -0.33, p = 0.7395) or liquor (t(329.33) = 1.06, p = 0.2889) per week and did not found any association with any specific beverage consumption even when we added them up (t(283.46) = 0.89, p = 0.38) (Figure 5.15).

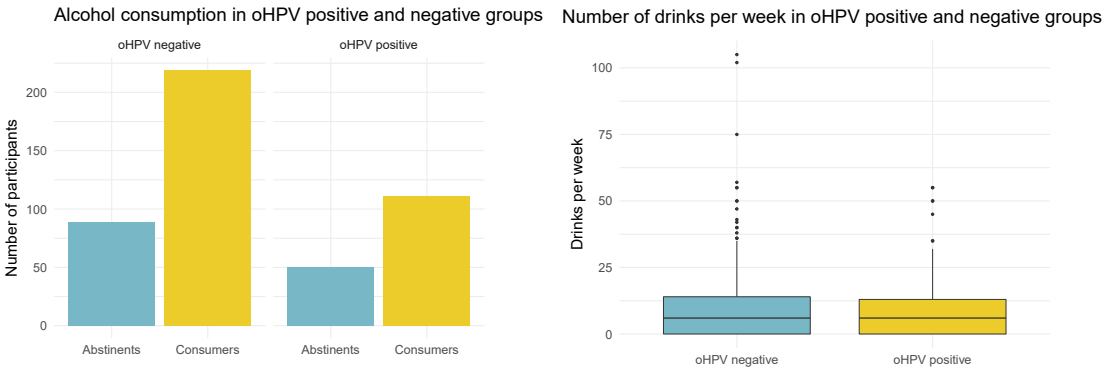

Figure 5.15: Drinking habits of oHPV$^+$ and oHPV$^-$ participants

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

Association between tobacco use and oHPV infection

Majority of our participants identified as smokers, irrespective of oHPV infection. Smoking does not affect oHPV infection in our sample ($\chi^2(1) = 0.10$, p = 0.75). For other types of smoking (cigars, pipe) and tobacco use (chewing, sniffing), the numbers of users among our respondents were very low - smoking cigars (n=9), pipes (n=10), sniff (n=3), chew tobacco (n=0) and therefore we did not assess the effect of these behaviours. When we evaluated effect of pack-years, we found significant difference (t(247.16) = 2.12, p = 0.04). In the oHPV$^+$ group the mean of pack-years is actually lower (M = 24.67, SD = 19.33) than in the oHPV$^-$ group (M = 29.61, SD = 20.9) (Figure 5.16).

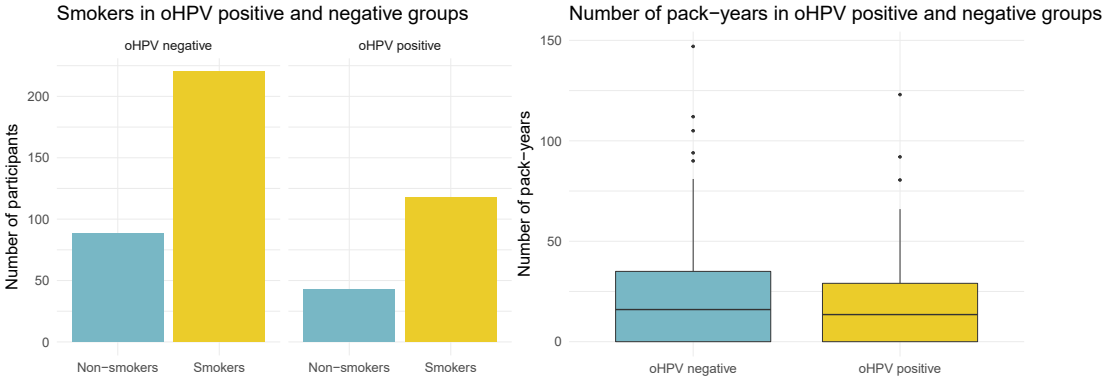

Figure 5.16: Smoking habits of oHPV$^+$ and oHPV$^-$ participants

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

Association between sexual behaviour and oHPV infection

It is known that HPV is sexually transmitted virus. When assessing oHPV infection, we focused on sexual activities involving oral contacts (Figure 5.17). We analysed the relationship with number of sexual partners as well. In oHPV$^-$ group, the number of sexual partners was lower (M = 8.44, SD = 12.03) than in the oHPV$^+$ group (M = 11.97, SD = 17.18) and the association is statistically significant (t(229.46) = -2.26, p = 0.02).

Even though frequencies of oral-oral ($\chi^2(1) = 0.32$, p = 0.57), oral-genital ($\chi^2(1) = 0.37$, p = 0.54) and oral-anal ($\chi^2(1) = 2.60$, p = 0.10) contact did not differ, the mean age of first oral-genital contact was lower in oHPV$^+$ group (M = 19.51, SD = 4.09) than in oHPV$^-$ group (M = 21.02, SD = 6.17) and is a significant predictor of oHPV infection (t(286.61) = 2.59, p = 0.01).

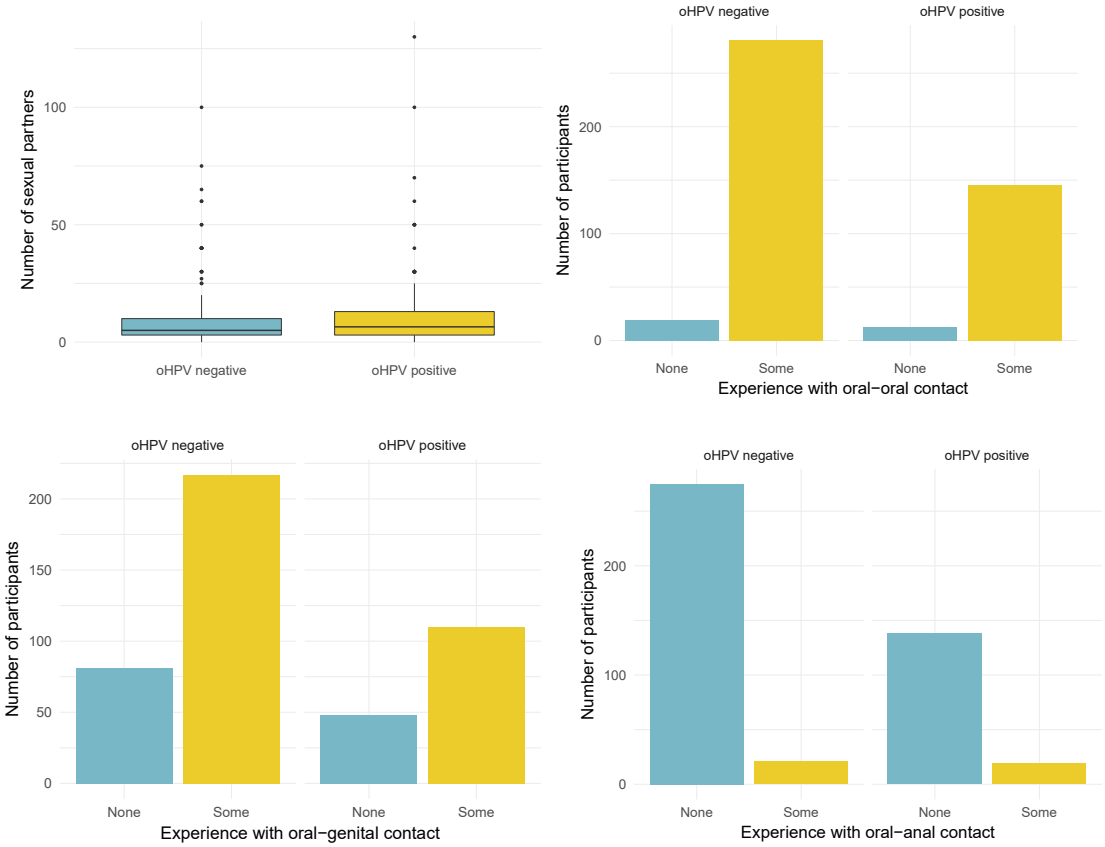

Figure 5.17: Sexual behaviours of oHPV$^+$ and oHPV$^-$ participants

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

Descriptives of the tested variables in oHPV$^+$ and oHPV$^-$ participants and associations found in preliminary tests are summarized in Table 5.7.

Table 5.7: Association of oHPV status with participants' demographics and risk behaviours

Data source: NRL for Papillomaviruses and Polyomaviruses

	HPV$^+$ (n=162*) N (% or SD)	HPV$^-$ (n=319*) N (% or SD)	p-value
Gender			0.55
males	126 (77.78%)	238 (74.61%)	
females	36 (22.22%)	80 25.08%)	
Mean age (years)	57.86 (SD = 9.33)	55.86 (SD = 10.35)	0.03
Completed years of education	13.18 (SD = 2.84)	12.86 (SD = 3.04)	0.26
Alcohol consumption			0.70
alcohol consumers	111 (68.52'%)	219 (71.10%)	
life long abstinence	50 (30.86%)	89 (28.90%)	
Drinks per week	9.21 (SD = 11.54)	10.3 (SD = 14.47)	0.39
Smoking			0.75
smokers	118 (72.84%)	220 (68.97%)	
non-smokers	43 (26.54%)	88 (27.59%)	
Pack-years	24.67 (SD = 19.33)	29.61 (SD = 20.9)	0.04
Oral-oral contact (kissing)			0.57
never	13 (8.02%)	19 (5.96%)	
have experience	145 (89.51%)	281 (88.09%)	
Oral-genital contact			0.54
never	48 (29.63%)	81 (25.39%)	
have experience	110 (67.90%)	217 (68.03%)	
Mean age of first oral-genital contact	19.51 (SD = 4.09)	21.02 (SD = 6.17)	0.01
Oral-anal contact			0.10
never	138 (85.19%)	275 (86.21%)	
have experience	19 (11.73%)	21 (6.58%)	

* Where percentages do not add up to 100%, some values were missing due to non-response.

Generalized linear model of oHPV risk factors

After considering all possible interactions and assumptions based on the literature, we decided on fitting binomial generalized linear model predicting HPV infection with the following variables: gender, years of completed education, number of pack-years, number of drinks per week, number of sexual partners and experience with oral-oral, oral-genital and oral-anal contact.

Odds ratios and p-values that these predictors yield can be found in Table 5.8. The model has overall pseudo R-squared 0.29.

Table 5.8: Risk factors of oHPV infection

Data source: NRL for Papillomaviruses and Polyomaviruses

	OR	CI (2.5% - 97.5%)		p-value
Gender				
female	1.00			
male	1.01	0.61	1.69	0.98
Age	1.02	1.00	1.04	0.07
Number of sexual partners	1.02	1.00	1.04	0.04
Years of education	1.04	0.97	1.12	0.28
Number of pack-years	1.00	0.98	1.01	0.41
Drinks per week	1.00	0.98	1.01	0.62
Experience with				
oral-oral contact				
no	1.00			
yes	0.81	0.33	2.02	0.64
oral-genital contact				
no	1.00			
yes	0.89	0.52	1.53	0.67
oral-anal contact				
no	1.00			
yes	1.44	0.67	3.00	0.34

As can be seen, only the number of sexual partners seems to be a significant predictor of oHPV infection. Our results did not show any other statistically significant ($p = 0.4$) association of oHPV status with tested behaviours, with OR = 1.02. Possible reasons and further research options will be further discussed.

5.2.3 Risk factors of head and neck cancer

In the analyses of risk factors of HNSCC, we first explored associations of control/patient status with participants' demographics and reported risk behaviours before we proceeded to analysis of mixed effects in the final logistic regression.

Association between gender and HNSCC

Number of women and men in our sample was not equal, with the majority being men. This is not surprising, as HNSCC incidence is higher in men (subsection 5.1.2). The observed difference (Figure 5.18) was even greater and statistically significant in the patient group ($\chi^2(1) = 9.73$, $p < 0.01$), which is caused by sampling.

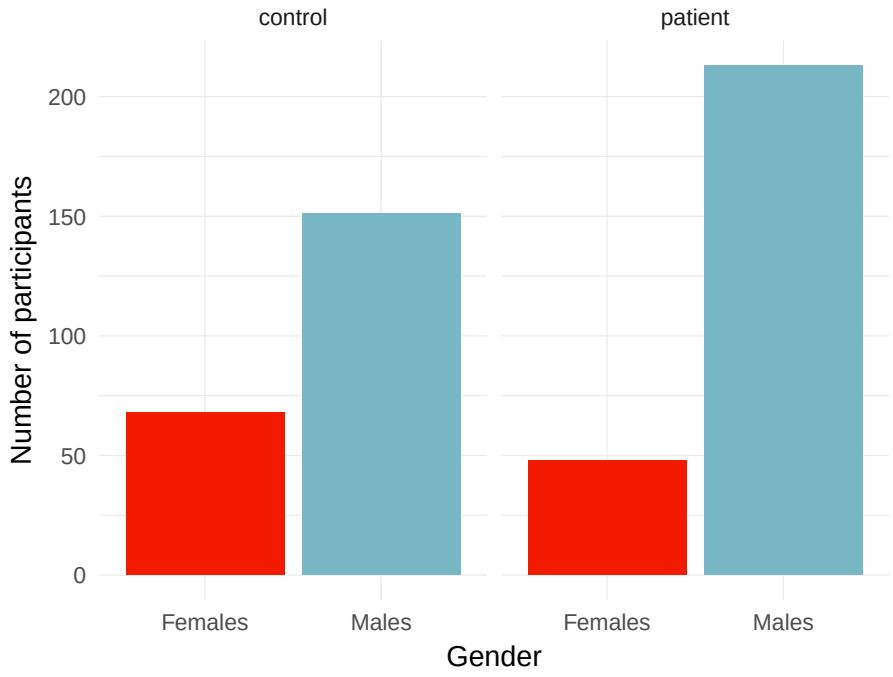

Figure 5.18: Gender of patients and controls

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

Association between oHPV and HNSCC

Infection with HR HPV is a known risk factor for HNSCC since HR HPV are the causative agent. Indeed, we found that the majority of HPV non-infected belong to the control group ($\chi^2(1) = 74.59$, $p < 0.01$).

When we focused specifically on the effect of HR HPV presence in oral washing (Table 5.9 and Figure 5.19), we saw that in comparison to patients only very small proportion of controls is HR HPV positive (n=14) and the association was strong ($\chi^2(1) = 96.78$, $p < 0.01$) for HNSCC.

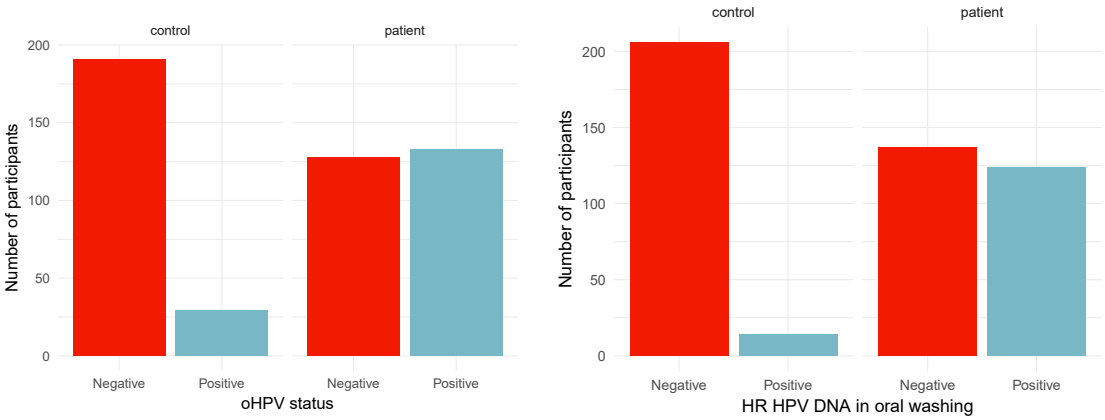

Figure 5.19: HPV and HR HPV presence in oral washings of patients and controls

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

Table 5.9: Presence of HR HPVs in oral washing of controls and patients

Data source: NRL for Papillomaviruses and Polyomaviruses

	controls	patients
HR HPV positive	14	124
HR HPV negative	206	137

Association between age and HNSCC

Mean age was almost two years higher (M = 57.3, SD = 9.48) in the patient group than in the control group (M= 55.6, SD = 10.64), but this difference was not statistically significant (t(441.06) = -1.79, p = 0.07). The distribution of age in both groups is shown in Figure 5.20.

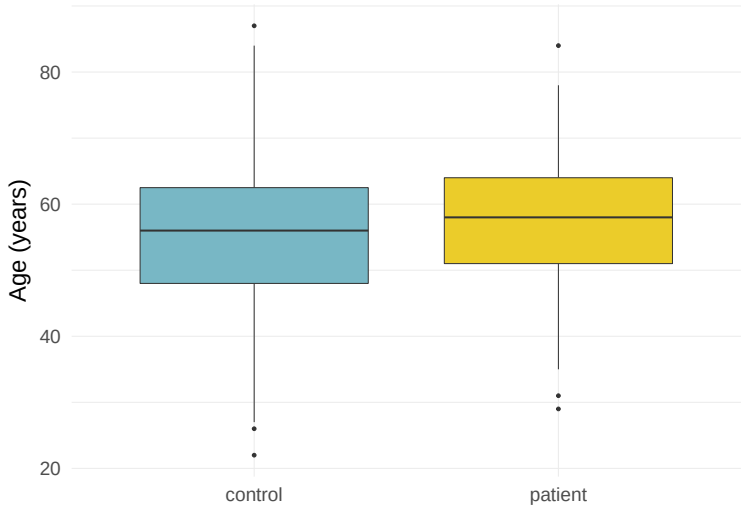

Figure 5.20: Age of patients and controls

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

Association between education and HNSCC

Mean number of years of completed education was higher in the control group ($t(417.55) = 3.51$, $p < 0.01$). Distribution of this variable is shown in Figure 5.21.

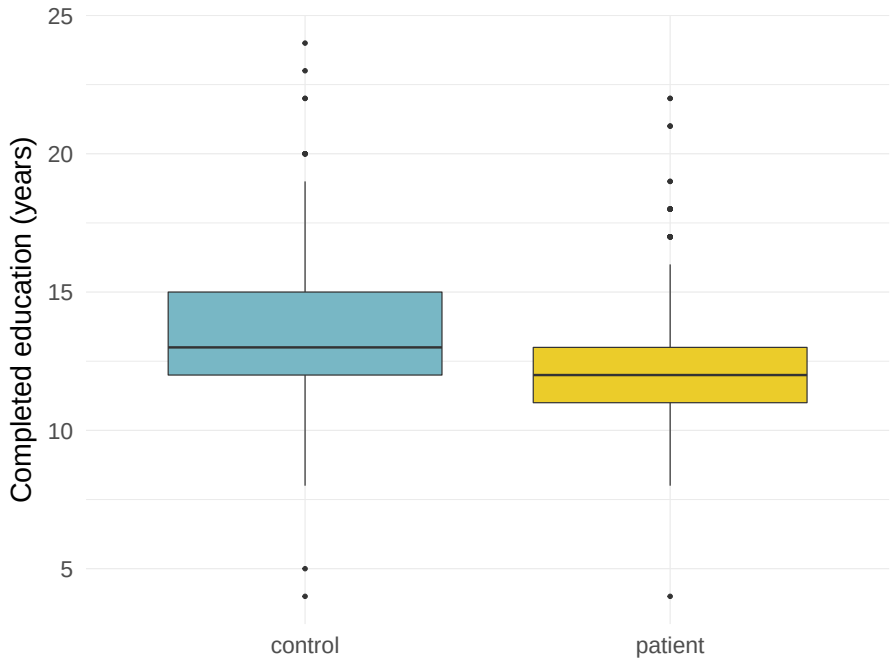

Figure 5.21: Completed years of completed education of patients and controls

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

Association between alcohol consumption and HNSCC

Alcohol consumption is known to be strong risk factor for HNSCC. In our dataset, the number of alcohol consumers was higher in the patient group (Figure 5.22) and the association is statistically significant ($\chi^2(1) = 4.79$, $p = 0.03$).

We quantified the alcohol consumption as number of drinks per week reported by the participants (Figure 5.22). We found strong correlation with beer consumption ($t(317.39) = -5.3032$, $p < 0.01$), but none with wine ($t(306.07) = -0.081$, $p = 0.94$) and liquor ($t(317.43) = -1.35$, $p = 0.18$) consumption. With all types of drinks added up, we found significant difference in alcohol consumption

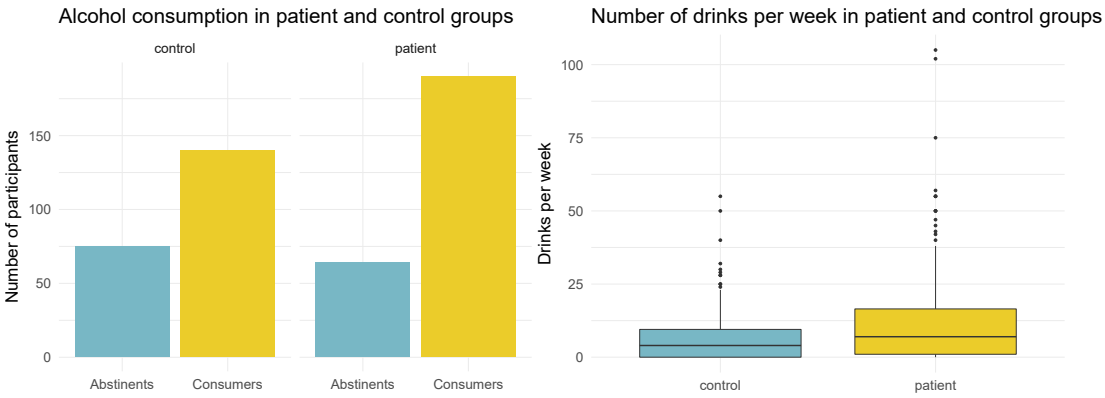

Figure 5.22: Drinking habits of patient and control groups

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

between patients (M = 16.38, SD = 16.51) and controls (M = 9.59, SD = 8.86) (t(305.9) = -4.8306, $p < 0.01$), with patients consuming more drinks.

Association between tobacco use and HNSCC

Our dataset contained a large proportion of smokers. In patient group, smokers represented the majority (78.93%) and association of patient/control status with smoking was found significant ($\chi^2(1) = 21.50$, $p < 0.01$). We could not evaluate other means of tobacco use because of the small proportion of participants who reported smoking pipe (n = 10), cigars (n = 9), or using tobacco intranasally (sniff; n = 3) or by chewing (n = 0).

We also assessed smoking as continuous variable through pack-years. We see significant association (t(286.18) = -2.90, $p < 0.01$), as the mean of pack-years in patients was much higher (M = 30.29, SD = 21.62) than in controls (M = 23.79, SD = 17.73)(Figure 5.23).

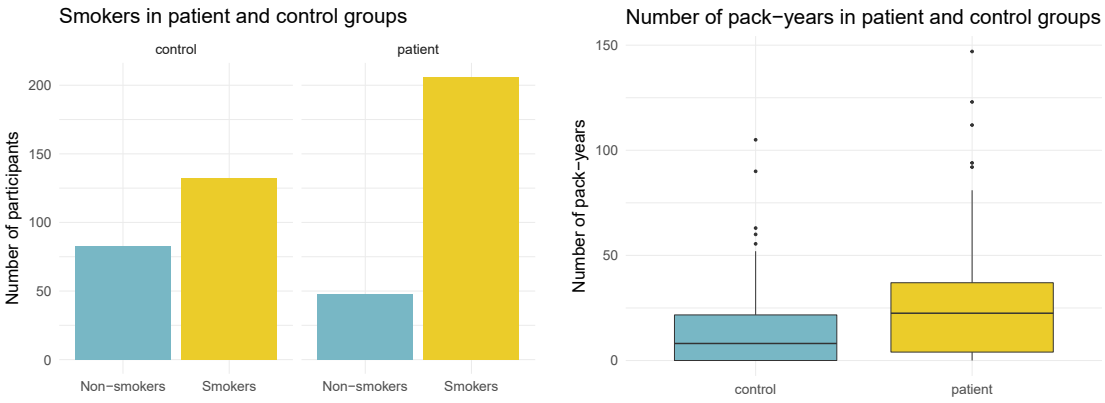

Figure 5.23: Smoking habits of patient and control groups

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

Association between sexual behaviour and HNSCC

Since HPV (an etiological agent) is a sexually transmitted virus, some sexual behaviours (e.g. early age of sexual debut, oral-oral, oral-genital and oral-anal contact and higher number of sexual partners) were found to be risk factors for cervical cancer. We assessed number of sexual partners and oral-genital and oral-anal contact. Mean number of sexual partners was significantly ($t(416.14) = -3.17$, $p < 0.01$) higher in patients ($M = 11.55$, $SD = 16.2$) than in controls ($M = 7.46$, $SD = 10.8$).

When we analysed association of HNSCC with other sexual risk behaviours (Figure 5.24), we have not found significant association with oral-oral contact ($\chi^2(1) = 0.05$, $p = 0.82$), oral-genital contact ($\chi^2(1) = 1.98$, $p = 0.16$) or even oral-anal contact ($\chi^2(1) = 0.19$, $p = 0.66$). However, we found that the mean age of first oral-genital contact was lower in patients ($M = 19.91$, $SD = 5.49$) than in controls ($M = 21.16$, $SD = 5.67$) and the association with early age of the first oral-genital contact is significant ($t(309.2) = 1.97$, $p = 0.0497$).

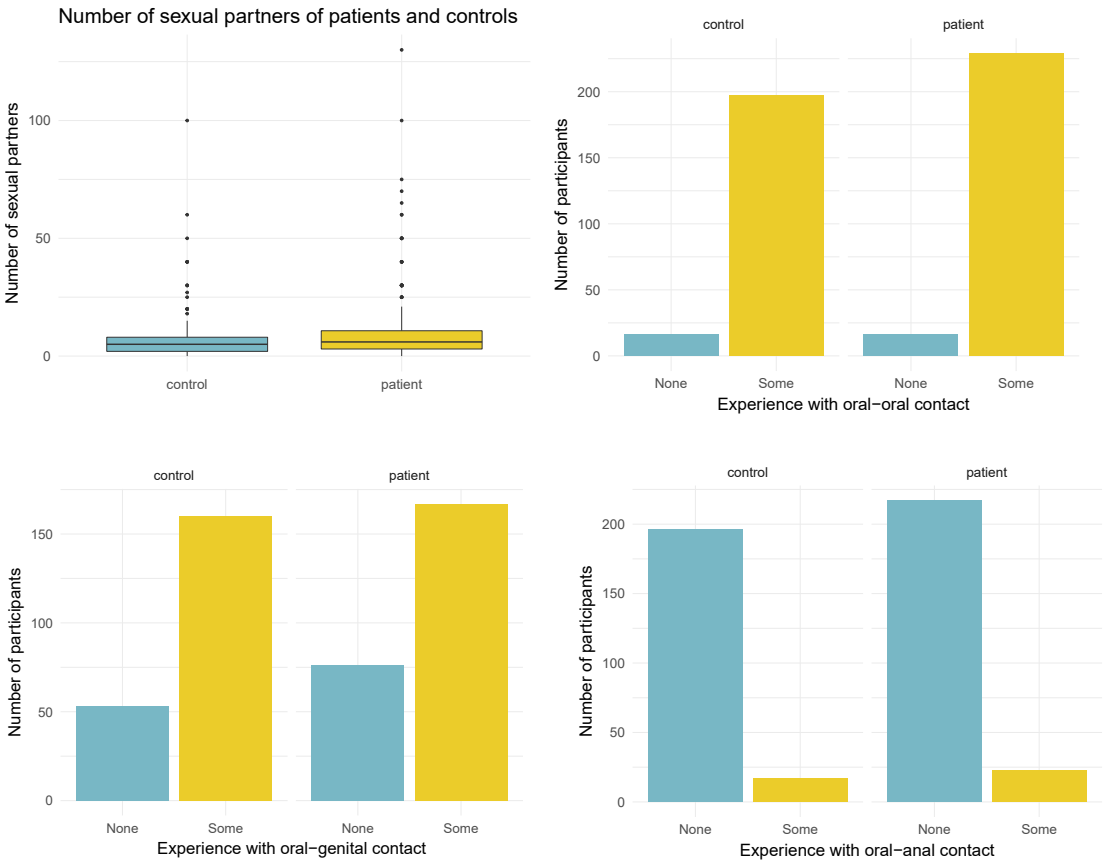

Figure 5.24: Sexual behaviours of patients and controls

Authors own work based on data from NRL for Papillomaviruses and Polyomaviruses, generated in R

Descriptives of the tested variables in patient and control group and associations found in preliminary tests are summarized in Table 5.10.

Table 5.10: Association of HNSCC diagnosis with participants' demographics and risk behaviours

Data source: NRL for Papillomaviruses and Polyomaviruses

	Patients (n=261*) N (% or SD)	Controls (n=220*) N (% or SD)	p-value
Gender			<0.01
males	213 (81.61%)	151 (68.64%)	
females	48 (18.39%)	68 (30.91%)	
HPV presence in oral wash			<0.01
positive (oHPV$^+$)	133 (50.96%)	29 (13.18%)	
negative (oHPV$^-$)	128 (49.04%)	191 (86.82%)	
Mean age	57.3 (SD = 9.48)	55.63 (SD = 10.64)	0.07
Completed years of education	12.53 (SD = 2.7)	13.5 (SD = 3.19)	<0.01
Alcohol			0.03
alcohol consumers	190 (72.80%)	140 (63.64%)	
life long abstinence	64 (24.52%)	75 (34.09%)	
Drinks per week	12.73 (SD = 16.07)	6.58 (SD = 8.58)	<0.01
Smoking			<0.01
smokers	206 (78.93%)	132 (60%)	
non-smokers	48 (18.39%)	83 (37.73%)	
Pack-years	30.29 (SD = 21.62)	23.79 (SD = 17.73)	<0.01
Number of sexual partners	11.55 (SD = 16.2)	7.46 (SD = 7.46)	<0.01
Oral-oral contact (kissing)			0.82
never	16 (6.13%)	16 (7.27%)	
have experience	229 (87.74%)	197 (89.55%)	
Oral-genital contact			0.16
never	76 (29.12%)	53 (24.09%)	
have experience	167 (63.98%)	160 (72.72%)	
Mean age of first oral-genital contact	19.91 (SD = 5.49)	21.16 (SD = 5.67)	0.05
Oral-anal contact			0.66
never	217 (83.14%)	196 (89.09%)	
have experience	23(8.81%)	17 (7.73%)	

* Where percentages do not add up to 100%, some values were missing due to non-response.

Generalized linear model of HNSCC risk factors

There are many potential risk factors of HNSCC. Some are known for a long time (alcohol consumption, tobacco exposure), while some arose in recent years (HPV). We decided to include the following variables as predictors of HNSCC: HR HPV in oral washing, age, gender, years of completed education, number of pack-years, number of drinks per week, number of sexual partners and experience with oral-oral, oral-genital and oral-anal contact.

Odds ratios, confidence intervals and p-values of the fitted model can be found in Table 5.11. The model has overall pseudo R-squared of 0.6.

We identified several significant predictors of HNSCC: we confirmed previously published risk factors - alcohol and cigarettes consumption, as well as protective factor of years of completed education. Unsurprisingly, we found the infection with HR HPV to be significant predictor as well. We confirmed that HR HPV is a stronger predictor than pack-years, number of sexual partners and drinks per week using Wald Test ($p < 0.001$). As a surprising discovery, we observed that engaging in oral-genital contact to be significant protective factor and engaging in oral-anal contact is not associated with the disease.

Table 5.11: Risk factors of head and neck cancer

Data source: NRL for Papillomaviruses and Polyomaviruses

	OR	CI (2.5% - 97.5%)		p-value
HR HPV in oral cavity				
HR HPV negative	1.00			
HR HPV positive	22.23	11.37	47.24	< 0.01
Age	1.00	0.98	1.03	0.88
Gender				
female	1.00			
male	1.41	0.76	2.65	0.27
Drinks per week	1.03	1.01	1.06	0.01
Years of education	0.90	0.82	0.98	0.02
Number of pack-years	1.03	1.01	1.04	< 0.01
Number of sexual partners	1.03	1.01	1.05	0.01
Experience with				
oral-oral contact				
no	1.00			
yes	0.95	0.34	2.76	0.92
oral-genital contact				
no	1.00			
yes	0.53	0.28	0.99	0.05
oral-anal contact				
no	1.00			
yes	1.57	0.61	4.05	0.35

To test the robustness of the prediction model, we implemented a randomizing script that recreated the general linearized model using data from two thirds of participants and then tried to predict the diagnoses of the remainder. All participants with higher than 0.5 probability of a positive diagnosis were marked to have HNSCC. We ran the randomizing procedure 100 times and then calculated a median proportion of correct model decisions, as well as false positives and false negatives (see Table 5.12).

The results clearly show, that the logistic model is quite capable in predicting HNSCC. Both false positives and negatives are kept low, with correct diagnosis of HNSCC patient reaching above 75 percent. The model still predicts diagnosis quite well even after removing all non significant variables (i.e. experience with oral-oral contact, age etc.). However, after removing data on HR HPV presence in oral washing, the accuracy drops.

Table 5.12: Strenght of prediction models

	Complete model	Model without non-significant variables	Model without HR-HPV status
False positive	0.18	0.18	0.40
False negative	0.24	0.28	0.35
Correct negative	0.82	0.82	0.60
Correct positive	0.76	0.72	0.65

To better investigate the two distinct groups that suffer from HNSCC, we created two separate prediction models for each of them: HR HPV positive and HR HPV negative. Because there were only 7 patients with no experience with oral-oral contact, we were forced to remove this predictor from the separated models. Results of these two models are presented in Table 5.13 and Table 5.14. The model of HR HPV positive HNSCC risk factors has overall pseudo R-squared 0.42 and model of HR HPV negative HNSCC risk factors has overall pseudo R-squared 0.50.

There are clear differences between these two groups. While in the HPV negative group, drinking and pack-years are a strong predictor of HNSCC, it is not the case in the HPV positive group. Also, age is a predictor in the

HPV positive group, but not in the HPV negative. This can be explained by progressive nature of HPV induced cancer. While HR HPV positive patients are older than HR positive controls. In HR HPV negative group, the age predictor is probably not significant because of its inter-correlation with pack-years. Years of completed education are a protective factor in both groups, although it is significant only in the HPV positive group.

Table 5.13: Risk factors of HR positive head and neck cancer

Data source: NRL for Papillomaviruses and Polyomaviruses

	OR	CI (2.5% - 97.5%)		p-value
Age	1.09	1.02	1.18	0.02
Gender				
female	1.00			
male	3.12	0.69	14.85	0.14
Drinks per week	0.97	0.90	1.04	0.32
Years of education	0.70	0.52	0.91	0.01
Number of pack-years	1.00	0.95	1.05	0.87
Number of sexual partners	1.08	0.99	1.25	0.18
Experience with				
oral-genital contact				
no	1.00			
yes	0.31	0.04	1.66	0.21
oral-anal contact				
no	1.00			
yes	2.84	0.28	76.25	0.44

Table 5.14: Risk factors of HR negative head and neck cancer

Data source: NRL for Papillomaviruses and Polyomaviruses

	OR	CI (2.5% - 97.5%)		p-value
Age	0.99	0.96	1.02	0.46
Gender				
female	1.00			
male	1.24	0.62	2.57	0.55
Drinks per week	1.04	1.02	1.08	< 0.01
Years of education	0.92	0.83	1.02	0.12
Number of pack-years	1.03	1.01	1.04	< 0.01
Number of sexual partners	1.03	1	1.05	0.02
Experience with				
oral-genital contact				
no	1.00			
yes	0.49	0.25	0.94	0.03
oral-anal contact				
no	1.00			
yes	1.57	0.54	4.36	0.4

6. Discussion

Epidemiology of head and neck cancers is a currently trending topic. Growing proportion of the tumors is caused by HPV infection which can be easily prevented. However, no study of epidemiological trends of this disease focused on the Czech Republic. In this thesis, we analysed mortality and incidence of head and neck cancer in the Czech Republic to show that the observed trends correspond to those of other populations described in the literature. We also used data from NRL for Papillomaviruses and Polyomaviruses to analyse risk factors of head and neck cancers and oHPV infection. Risk factors of oHPV infection have never been analysed in Czech population and our study of HNSCC risk factors is the biggest one performed in the Czech Republic up to now.

We demonstrated that the trends in HNSCC epidemiology in the Czech Republic follow the same pattern as in other countries. We identified higher number of sexual partners to be a risk factor of oHPV infection and presence of HR HPV in oral washing to be major risk factor of HNSCC in Czech patients. According to our results, higher number of sexual partners, smoking and alcohol consumption are also risk behaviours associated with HNSCC diagnosis.

When studying HNSCC epidemiology, it is important to clearly define the diagnoses in question. In lot of studies, the definition is vague and the studies often compare data from different time points, therefore probably using different ICD revisions. Even though this phenomenon is quite common, it is rarely discussed in these studies (Hashibe & Sturgis 2013). For example, Danish study of HNSCC incidence by Blomberg et al. (2011) used data on incidence in the years 1978–2007 with ICD-10 categories C00–C14 and C30–C32. Categories C30–C32 cover several codes of ICD-9 coding but are, however, not represented in the ICD-9 BTL classification. We could not use categories C30–C32 without also somehow converting mortality and incidence data from period before 1993 to ICD-9 BTL (death rates in those years in the Czech Republic are only accessible in CoDQL using BTL coding). In the mentioned study, the used data were extracted from the Danish Cancer Registry database, but authors do not explain which codes

of older ICD revision they used. For the demographic analyses in this thesis, we used ICD-10: C00–C14 and its corresponding ICD-9: 140–149 codes (BTL: B08) as is usual in HNSCC epidemiological studies. In the second part about HNSCC and HPV risk factors, we used ICD-10: C00–C06; C09–C10 which was due to the patient enrolment criteria in the original study. We used available online tools for the ICD-10 to ICD-9 conversion and for its translation to Basic Tabulation List codes. However, our results cannot be easily compared to studies where authors do not specify their ICD conversion approach.

In demographic analyses, we can only use aggregated data. With such data, it is impossible to follow association between incidence or mortality with specific factors, such as etiology of the disease. However, it enables us to follow population trends. Here we showed how HNSCC age-specific mortality changes over time in the Czech Republic. We observed that in 1986 mortality was higher in the older age groups (>70 years) than in the year 2013 where the incidence increased in younger age groups. This is generally attributed to increasing incidence of HPV-associated HNSCC, which affects younger individuals. In Denmark population, the identified age group at risk for HPV-associated is >60 and 60+ for non-HPV-associated HNSCC (Blomberg et al. 2011). In the USA, the increase of incidence has been also observed in patients younger than 60 (reviewed in D'Souza & Dempsey (2011)). In data on incidence and death rates between the years 1986 and 2013 in the Czech Republic, we observed the dividing age to be 70 years rather than 60. This difference might be specific to the Czech population, but it could also be an effect of disease rareness. The greatest differences in incidence within one age group were around 15 cases per 100, 000 in males and around 5 cases per 100, 000 in women. Overall differences in death rates over the years were similarly small: 27.2 deaths per 100, 000 in men (in 85+ category) and 14.5 deaths per 100, 000 in women (in 85+ category). Future studies could explore these differences in more depth - it would be interesting to learn about the changes in incidence or death rates of individual diagnosis (specific location of the tumour) and to compare epidemiology of head and neck cancer in the Czech Republic with epidemiology in other countries, perhaps in respect to their anti-smoking laws and the time of implementing of such measures.

When we look at the incidence in regions of the Czech Republic in 2014, we see differences among the regions for both men and women. In men, the incidence is highest in Zlín Region, while in women it is highest in Prague and Hradec Králové Region. It is important to note that the biggest differences are only 5 cases per 100, 000 women (Vysočina Region and Prague) and 18.8 cases per 100, 000 men (Olomouc and Zlín Region). We saw the same patterns for region-specific death rates. In future studies, it would be interesting to also look how region-specific death rate changes overtime. Correlating the incidence and death rate with alcohol and tobacco consumption in those regions could show whether these risk behaviours explain observed differences.

Throughout our demographic analyses, we saw higher HNSCC death rates and incidence in men, which is in agreement with worldwide trends. Even though from biological point of view there is no evidence of different mechanisms of HNSCC origin in men, it is interesting to consider potential differences in risk factors, for example possibly different route of transmission or genetic factors. Because of the lower incidence in women, there were also fewer women in our clinical sample and we could not analyse gender-specific risk factors. One of potential aims of future studies with even bigger sample could be investigation of gender-specific aspects of HNSCC.

Out of the variables fitted into the generalized linear model, the only predictor of oHPV infection was the number of sex partners. Surprisingly, we did not found association with any of the sexual practices. This may be due to the fact, that the number of participants who report that they have some experience with oral-oral contact was too great and for oral-anal contact it was too low. Nevertheless, we show that oHPV is not associated with smoking and drinking alcohol and that there are no gender-specific differences like we see for HNSCC. In future studies focused on oHPV infection, we could also investigate association with frequency of practices such as oral-genital contact, with number of sex partners for these practices and perhaps even the use of barrier methods of birth control as we further discuss below.

In analysis of risk factors of HNSCC, we confirmed previously found risk factors, such as smoking, alcohol consumption, higher number of sexual partners and most notably presence of HR HPV in the oral cavity. One of the most interesting results is the protective effect of oral-genital contact in HNSCC ($p = 0.05$). We could argue that by potentially more frequent encounter with the virus individuals engaging in oral-genital contact could develop stronger immunity against it, since preexisting antibodies can prevent reinfection with the virus and it has been previously shown, that HPV seropositivity is associated with higher number of sexual partners (Rahman et al. 2016, Rettig et al. 2016). In the analyses of oHPV risk factors, we did not see the same protective effect of oral-genital contact. Even so, the protective effect of oral-genital contact could potentially protect against developing persistent oHPV infection (and subsequently developing cancer) but not against the mere HPV presence at the point of the study. On the contrary - if this were true, we should actually see increased incidence of oHPV in individuals who engage more frequently in oral-genital contact (and thus boosting their immune system). Moreover, up to now no study has suggested such effect of "immunity boost by risk behaviours", even though several studies had result similar to ours: either no significant association or protective affect of oral-genital contact in head and neck cancers (reviewed in Chancellor et al. (2016)). Our study unfortunately did not collect information about the frequency of participants' sexual practices or the number of partners for such practices, but it would be yet another interesting topic for future studies. While the number of sexual partners for oral sex is sometimes included in the studies, but not always found to be significant risk factor in HNSCC (reviewed in Chancellor et al. (2016)), frequency of oral-genital sex is not a commonly investigated risk factor.

In both of our risk factor analyses, we used set of clinical data from ORL clinic. Our participants were enlisted as either patients (with HNSCC) or controls. While sampling is a critical point of any clinical study design, hospital-based sampling certainly has its advantages. We automatically filter out role of the en-

vironment because patients and controls live roughly in the same area. Above all, it is highly practical for conducting the study. That said, it is important to also be mindful that hospital-based population does have its specifics, be it higher prevalence of smoking or occurrence of other risk factors which might have brought them into the hospital in the first place. Although our questionnaire covered potential comorbidities, we decided not to use this data as it should not have major impact on the presented results. If used carefully and with awareness of potential biases which might occur when using hospital-based cohort, our gender and age matched cohort is quite suitable for analysis of such rare outcome as head and neck cancer.

The same cohort was also used in analysis of risk factors of oHPV. Unfortunately, it suffered from unequal representation of oHPV positive and negative individuals. However, case-control study for oHPV infection would be more complicated to conduct than for HNSCC because the infection is inapparent and participants would have to be tested in order to be designated as 'cases' or 'controls'. Making use of the data gathered in case-control study for HNSCC is a suitable way to inspect risk factors for oHPV infection.

Second important study limitation is the design of the questionnaire, which originated almost 17 years ago. In concordance with current literature, it would be helpful to also include question on age of sexual debut. Our cohort was sampled before the first HPV vaccine was available in the Czech Republic. Incidental vaccination with the first available vaccines in 2006 would not influence the results of HNSCC risk factor analysis, because the vaccine could not have reversed any ongoing pathological changes in the patient group. Recently, the question of HPV vaccine status became more relevant and therefore should be included in future studies. Our questionnaire also did not take into consideration the use of barrier methods of birth control. Since HPV is sexually transmitted, the use of condoms (in this case during oral sex) is an important way to prevent the infection and with vaccination could be an important target of prevention programs.

Comparison of our results to those of other studies is compromised by the fact that each study defines certain variables differently. For example in study by

Gillison et al. (2008), alcohol drinker is defined as “an individual who had ever consumed one or more drinks of alcohol per month for more than 1 year” and smoker as “individual who had ever smoked one or more cigarettes per day for 1 year or longer”. Some studies categorise age as “55 and younger” and “older than 55” (Tachezy et al. 2009), whereas some other studies create even more categories (Chaturvedi et al. 2011, Anantharaman et al. 2016). Given that published studies do not typically agree on the dividing age category and our demographic results suggest that the ”younger age” might go as high as to 70 years, we decided to control for age using it as continuous variable.

In final linear regression models, we included age of participants as well as pack-years. These two variables were intercorrelated (not shown) and could confound the results. However, when we created the model with smoking status as a categorical variable (have the participant or have not ever smoked), the results remained unchanged (not shown). Therefore, we decided to keep both age and pack-years in the final models. Unfortunately, due to larger number of NA values due to non-response, we could not include age of the first oral-genital contact in the model which we have found significantly associated with oHPV infection and HNSCC diagnosis in the preliminary tests. We were also able to confirm that there are two distinct groups suffering from HNSCC, each with different predictors and risk behaviours. The group infected with HR HPV did not show any increased risk of HNSCC in relation to alcohol or smoking that was observed in the HR HPV negative group of HNSCC.

Using seropositivity as a marker for HPV positivity is debatable, because antibodies persist even when the infection is cleared by the immune system and is no longer relevant. On the other hand, it has been shown that the antibody level in serum is elevated prior to HNSCC diagnosis and can be used as predictor of HNSCC (Kreimer et al. 2013, 2017). In this thesis, we did not use results of serology tests as a marker of HPV infection. Because we were specifically interested in oHPV infections which cannot be serologically distinguished from HPV infections in other locations, including serological markers in our analysis of oHPV risk factors would in this case not contribute to the study.

According to the published literature, the oHPV prevalence is usually less than 10% (reviewed in Chung et al. (2014)). In our data set, 33.68% of participants were positive for HPV DNA in their oral washing. This is not surprising given that the high oHPV prevalence is to be expected in HNSCC patients. In healthy (without HNSCC) controls, the prevalence was 13.18% which is slightly higher than previously published prevalences, but could be explained by higher age of the control group in our age matched sample as the oHPV prevalence is associated with age (reviewed in Chung et al. (2014)). Indeed, it would be interesting to conduct population-based study, to confirm that the oHPV incidence in the Czech Republic is not higher than it has been published for other populations. Coincidentally, such study is currently ongoing in the NRL for Papillomaviruses and Polyomaviruses. Preliminary data from university student population show that the oHPV prevalence is indeed very low and agrees with previously published studies (unpublished results of NRL, personal communication).

Upon identifying risk population for contracting oHPV and subsequently developing HNSCC, discussion about prospective prevention programmes should take place. Primary prevention in the form of already available HPV vaccines could be undeniably the first step in control of HNSCC. Here, it is important to again emphasise that HPV can cause not only cervical cancer but also other benign and malignant diseases also in men such as penile, perianal, anal and HNSCC cancer as well as genital warts and laryngeal papillomatosis. While HNSCC threatens especially men, the recommendation to also vaccinate boys came out only recently (Czech Vaccination Society ČLS JEP 2017).

Precancerous stages detectable by examination of oral or histological tests in individuals at risk (similarly to cervical carcinoma screening) have not yet been established as a suitable method for HNSCC screening. The data of the two recent studies suggest that serological markers of HPV infection could be an early marker of HNSCC, but currently, there is no intervention that can be offered to individuals found to be HPV positive (reviewed in Kreimer (2014)). Some researchers propose tonsillectomy (i.e. surgical removal of tonsils) as preventive measure for such individuals at risk (Fakhry et al. 2015). This procedure however

is not without its risks and further evaluation of its advantages and disadvantages is needed.

Some potential research targets were already mentioned throughout this discussion. Additionally, future studies could further explore for example risk factors associated with specific anatomic locations of the tumour, more biological markers as a tool for screening and also evaluate patient survival associated with these factors and potential individual treatment plans.

In this thesis, we illustrated epidemiological trends of HNSCC in the Czech Republic. We analysed risk factors of HNSCC in case-control study conducted by NRL for Papillomaviruses and Polyomaviruses. Furthermore, we explored risk factors for oHPV infection, which has previously never been done in Czech population.

7. Conclusion

Head and neck cancers are increasing in incidence worldwide and Czech Republic is no exception. The need of effective screening and means of prevention is apparent and largely debated in published literature. In this thesis, we contributed to knowledge of Czech population-specific trends of HNSCC.

In the first part, we have set out to show epidemiological trends of HNSCC in Czech republic with the use of publicly available data in large databases, such as WHO Mortality Database and Czech National Oncology Registry. We show that in agreement with trends observed in other countries, mortality in the Czech Republic does not decline and since 1980s, the HNSCC incidence and death rates shift to younger age groups. We also show that incidence and death rates in regions of the Czech Republic are differently distributed in men and women and argue that these differences resemble the overall trend among European countries and is probably attributable to relative rareness of the disease.

In the second part of the thesis, we conduct case-control and cross-sectional study to analyse risk factors of head and neck cancers and oHPV infections. We identified number of sexual partners to be risk factor of oHPV infection a and alcohol consumption, smoking and higher number of sexual partners as risk factors for head and neck cancers. Most importantly, we show that HR HPV presence in oral cavity is indeed strong predictor of head and neck cancer regardless of other risk factors and that two distinct groups of patients with different risk factors can be identified based on the presence of HR HPV in oral cavity.

Bibliography

Agaku, I. T., King, B. A. & Dube, S. R. (2014), 'Current cigarette smoking among adults—united states, 2005–2012', *MMWR Morb. Mortal. Wkly. Rep.* **63**(2), 29–34.

Anantharaman, D., Muller, D. C., Lagiou, P., Ahrens, W., Holcátová, I., Merletti, F., Kjærheim, K., Polesel, J., Simonato, L., Canova, C., Castellsague, X., Macfarlane, T. V., Znaor, A., Thomson, P., Robinson, M., Conway, D. I., Healy, C. M., Tjønneland, A., Westin, U., Ekström, J., Chang-Claude, J., Kaaks, R., Overvad, K., Drogan, D., Hallmans, G., Laurell, G., Bueno-de Mesquita, H. B., Peeters, P. H., Agudo, A., Larrañaga, N., Travis, R. C., Palli, D., Barricarte, A., Trichopoulou, A., George, S., Trichopoulos, D., Quirós, J. R., Grioni, S., Sacerdote, C., Navarro, C., Sánchez, M.-J., Tumino, R., Severi, G., Boutron-Ruault, M.-C., Clavel-Chapelon, F., Panico, S., Weiderpass, E., Lund, E., Gram, I. T., Riboli, E., Pawlita, M., Waterboer, T., Kreimer, A. R., Johansson, M. & Brennan, P. (2016), 'Combined effects of smoking and HPV16 in oropharyngeal cancer', *Int. J. Epidemiol.* pp. dyw069–.

Blomberg, M., Nielsen, A., Munk, C. & Kjaer, S. K. (2011), 'Trends in head and neck cancer incidence in denmark, 1978–2007: focus on human papillomavirus associated sites', *International journal of Cancer* **129**(3), 733–741.

Boshart, M., Gissmann, L., Ikenberg, H., Kleinheinz, A., Scheurlen, W. & zur Hausen, H. (1984), 'A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer', *EMBO J.* **3**(5), 1151–1157.

Chancellor, J. A., Ioannides, S. J. & Elwood, J. M. (2016), 'Oral and oropharyngeal cancer and the role of sexual behaviour: a systematic review', *Community Dent. Oral Epidemiol.* **45**(1), 20–34.

Chaturvedi, A. K., Engels, E. A., Pfeiffer, R. M., Hernandez, B. Y., Xiao, W., Kim, E., Jiang, B., Goodman, M. T., Sibug-Saber, M., Cozen, W., Liu, L., Lynch, C. F., Wentzensen, N., Jordan, R. C., Altekruse, S., Anderson,

W. F., Rosenberg, P. S. & Gillison, M. L. (2011), 'Human papillomavirus and rising oropharyngeal cancer incidence in the united states', *J. Clin. Oncol.* **29**(32), 4294–4301.

Chung, C. H., Bagheri, A. & D'Souza, G. (2014), 'Epidemiology of oral human papillomavirus infection', *Oral Oncol.* **50**(5), 364–369.

Cragg, J. G. & Uhler, R. S. (1970), 'The demand for automobiles', *Can. J. Econ.* **3**(3), 386–406.

Czech Statistical Office (2016), 'Zemřelí podle seznamu příčin smrti, pohlaví a věku v čr, krajích a okresech - 2006 až 2015 — čsú', `https://www.czso.cz/csu/czso/zemreli-podle-seznamu-pricin-smrti-pohlavi-a-veku-v-cr-krajich-a-okresech-20` Accessed: 2017-6-28.

Czech Vaccination Society ČLS JEP (2017), 'Czech vaccination society', `http://www.vakcinace.eu/novinky`. Accessed: 2017-7-4.

D'Souza, G. & Dempsey, A. (2011), 'The role of HPV in head and neck cancer and review of the HPV vaccine', *Prev. Med.* **53 Suppl 1**, S5–S11.

D'Souza, G., Kreimer, A. R., Viscidi, R., Pawlita, M., Fakhry, C., Koch, W. M., Westra, W. H. & Gillison, M. L. (2007), 'Case–control study of human papillomavirus and oropharyngeal cancer', *N. Engl. J. Med.* **356**(19), 1944–1956.

Dušek, L., Mužík, J., Kubásek, M., Koptíková, J., Žaloudík, J. & Vyzula, R. (2005), 'Epidemiologie zhoubných nádorů v české republice [online]', *Masarykova univerzita* .

EU2009.cz (2011), 'EU2009: Regions of the czech republic', `http://www.eu2009.cz/en/czech-republic/regions/regions-of-the-czech-republic-329/`. Accessed: 2017-6-30.

Fakhry, C., Andersen, K. K., Christensen, J., Agrawal, N. & Eisele, D. W. (2015), 'The impact of tonsillectomy upon the risk of oropharyngeal carcinoma diagno-

sis and prognosis in the danish cancer registry', *Cancer Prev. Res.* **8**(7), 583–589.

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J. W. W., Comber, H., Forman, D. & Bray, F. (2013), 'Cancer incidence and mortality patterns in europe: estimates for 40 countries in 2012', *Eur. J. Cancer* **49**(6), 1374–1403.

Gandini, S., Botteri, E., Iodice, S., Boniol, M., Lowenfels, A. B., Maisonneuve, P. & Boyle, P. (2008), 'Tobacco smoking and cancer: a meta-analysis', *Int. J. Cancer* **122**(1), 155–164.

Gillison, M. L., D'Souza, G., Westra, W., Sugar, E., Xiao, W., Begum, S. & Viscidi, R. (2008), 'Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers', *JNCI: Journal of the National Cancer Institute* **100**(6), 407–420.

Gillison, M. L., Koch, W. M., Capone, R. B., Spafford, M., Westra, W. H., Wu, L., Zahurak, M. L., Daniel, R. W., Viglione, M., Symer, D. E., Shah, K. V. & Sidransky, D. (2000), 'Evidence for a causal association between human papillomavirus and a subset of head and neck cancers', *J. Natl. Cancer Inst.* **92**(9), 709–720.

Gooi, Z., Chan, J. Y. K. & Fakhry, C. (2016), 'The epidemiology of the human papillomavirus related to oropharyngeal head and neck cancer', *Laryngoscope* **126**(4), 894–900.

Hamšíková, E. (2008), 'Očkování proti HPV z pohledu vi-rologa', `http://zdravi.euro.cz/clanek/postgradualni-medicina/ockovani-proti-hpv-z-pohledu-virologa-338167`. Accessed: 2017-6-30.

Harrell, F. (2010), *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*, Springer New York.

Hashibe, M., Brennan, P., Benhamou, S., Castellsague, X., Chen, C., Curado, M. P., Maso, L. D., Daudt, A. W., Fabianova, E., Wünsch-Filho, V. & Others

(2007), 'Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the international head and neck cancer epidemiology consortium', *J. Natl. Cancer Inst.* **99**(10), 777–789.

Hashibe, M. & Sturgis, E. M. (2013), 'Epidemiology of oral-cavity and oropharyngeal carcinomas: controlling a tobacco epidemic while a human papillomavirus epidemic emerges', *Otolaryngol. Clin. North Am.* **46**(4), 507–520.

Health Fusion, Inc (2017), 'ICD-9 to ICD-10 code search — ICD-10 code lookup & crosswalk', `http://www.icd10codesearch.com/`. Accessed: 2017-6-30.

Herrero, R., Quint, W., Hildesheim, A., Gonzalez, P., Struijk, L., Katki, H. A., Porras, C., Schiffman, M., Rodriguez, A. C., Solomon, D., Jimenez, S., Schiller, J. T., Lowy, D. R., van Doorn, L.-J., Wacholder, S., Kreimer, A. R. & CVT Vaccine Group (2013), 'Reduced prevalence of oral human papillomavirus (HPV) 4 years after bivalent HPV vaccination in a randomized clinical trial in costa rica', *PLoS One* **8**(7), e68329.

Howell-Jones, R., de Silva, N., Akpan, M., Oakeshott, P., Carder, C., Coupland, L., Sillis, M., Mallinson, H., Ellis, V., Frodsham, D., Robinson, T. I., Gill, O. N., Beddows, S. & Soldan, K. (2012), 'Prevalence of human papillomavirus (HPV) infections in sexually active adolescents and young women in england, prior to widespread HPV immunisation', *Vaccine* **30**(26), 3867–3875.

IHIS CR (2014), Cancer incidence in the czech republic, 2014, Technical report, Ústav zdravotnických informací a statistiky ČR.

International Agency for Research on Cancer (2012), 'European cancer observatory', `http://eco.iarc.fr/EUCAN/Cancer.aspx?Cancer=1`. Accessed: 2016-12-28.

Kimple, R. J. & Harari, P. M. (2015), 'The prognostic value of HPV in head and neck cancer patients undergoing postoperative chemoradiotherapy', *Ann Transl Med* **3**(Suppl 1), S14.

Knipe, D. M., Howley, P. M., Griffin, D. E., Lamb, R. A., Martin, M. A., Roizman, B., Straus, S. E. & Rubin, D. H. (2002), *Fields Virology, 4th Edition, Volumes I and II: Fundamental Virology*.

Kreimer, A. R. (2014), 'Prospects for prevention of HPV-driven oropharynx cancer', *Oral Oncol.* **50**(6), 555–559.

Kreimer, A. R., Johansson, M., Waterboer, T., Kaaks, R., Chang-Claude, J., Drogen, D., Tjønneland, A., Overvad, K., Quirós, J. R., González, C. A., Sánchez, M. J., Larrañaga, N., Navarro, C., Barricarte, A., Travis, R. C., Khaw, K.-T., Wareham, N., Trichopoulou, A., Lagiou, P., Trichopoulos, D., Peeters, P. H. M., Panico, S., Masala, G., Grioni, S., Tumino, R., Vineis, P., Bueno-de Mesquita, H. B., Laurell, G., Hallmans, G., Manjer, J., Ekström, J., Skeie, G., Lund, E., Weiderpass, E., Ferrari, P., Byrnes, G., Romieu, I., Riboli, E., Hildesheim, A., Boeing, H., Pawlita, M. & Brennan, P. (2013), 'Evaluation of human papillomavirus antibodies and risk of subsequent head and neck cancer', *J. Clin. Oncol.* **31**(21), 2708–2715.

Kreimer, A. R., Johansson, M., Yanik, E. L., Katki, H. A., Check, D. P., Lang Kuhs, K. A., Willhauck-Fleckenstein, M., Holzinger, D., Hildesheim, A., Pfeiffer, R., Williams, C., Freedman, N. D., Huang, W.-Y., Purdue, M. P., Michel, A., Pawlita, M., Brennan, P. & Waterboer, T. (2017), 'Kinetics of the human papillomavirus type 16 E6 antibody response prior to oropharyngeal cancer', *J. Natl. Cancer Inst.* **109**(8).

Lewis, A., Kang, R., Levine, A. & Maghami, E. (2015), 'The new face of head and neck cancer: The HPV epidemic', *Oncology* **29**(9), 616–626.

Mirghani, H., Amen, F., Blanchard, P., Moreau, F., Guigay, J., Hartl, D. M. & Lacau St Guily, J. (2015), 'Treatment de-escalation in HPV-positive oropharyngeal carcinoma: Ongoing trials, critical issues and perspectives', *International journal of cancer* **136**(7), 1494–1503.

Modis, Y., Trus, B. L. & Harrison, S. C. (2002), 'Atomic model of the papillomavirus capsid', *EMBO J.* **21**(18), 4754–4762.

Nagelkerke, N. J. D. (1991), 'A note on a general definition of the coefficient of determination', *Biometrika* **78**(3), 691–692.

NCI (2017), 'NIH: Head and neck cancers', `https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet`. Accessed: 2017-6-30.

Orth, G., Jablonska, S., Favre, M., Croissant, O., Jarzabek-Chorzelska, M. & Rzesa, G. (1978), 'Characterization of two types of human papillomaviruses in lesions of epidermodysplasia verruciformis', *Proc. Natl. Acad. Sci. U. S. A.* **75**(3), 1537–1541.

R Core Team (2017), 'R: A language and environment for statistical computing'.

Rahman, S., Pierce Campbell, C. M., Rollison, D. E., Wang, W., Waterboer, T., Michel, A., Pawlita, M., Villa, L. L., Lazcano Ponce, E., Borenstein, A. R. & Giuliano, A. R. (2016), 'Seroprevalence and associated factors of 9-valent human papillomavirus (HPV) types among men in the multinational HIM study', *PLoS One* **11**(11), e0167173.

Rettig, E. M., Fakhry, C., Rositch, A. F., Burke, A. E., Chang, K., Silver, M. I., Viscidi, R. & Gravitt, P. (2016), 'Race is associated with sexual behaviors and modifies the effect of age on human papillomavirus serostatus among perimenopausal women', *Sex. Transm. Dis.* **43**(4), 231–237.

Revelle, W. (2017), 'psych: Procedures for psychological, psychometric, and personality research'.

Shi, R., Devarakonda, S., Liu, L., Taylor, H. & Mills, G. (2014), 'Factors associated with genital human papillomavirus infection among adult females in the united states, NHANES 2007-2010', *BMC Res. Notes* **7**, 544.

Tachezy, R., Klozar, J., Rubenstein, L., Smith, E., Saláková, M., Smahelová, J., Ludvíková, V., Rotnáglová, E., Kodet, R. & Hamsíková, E. (2009), 'Demographic and risk factors in patients with head and neck tumors', *J. Med. Virol.* **81**(5), 878–887.

Urban, D., Corry, J. & Rischin, D. (2014), 'What is the best treatment for patients with human papillomavirus–positive and–negative oropharyngeal cancer?', *Cancer* **120**(10), 1462–1470.

WHO (2016), 'CoDQL - cause of death query online', `http://apps.who.int/healthinfo/statistics/mortality/causeofdeath_query/`. Accessed: 2016-12-28.

Wickham, H. (2009), 'ggplot2: Elegant graphics for data analysis'.

Wolfbane Cybernetic Ltd. (2001), 'International classification of diseases', `http://www.wolfbane.com/icd/icd9a.htm`. Accessed: 2016-12-28.

zur Hausen, H. (2009), 'Papillomaviruses in the causation of human cancers - a brief historical account', *Virology* **384**(2), 260–265.

A. Appendix

Figure A.1: Regions of the Czech Republic

Retrieved from EU2009.cz (2011)