InterCHANGE Study Protocol
August 3, 2010

International Cancer of the Head and Neck, Genetics and Environment (InterCHANGE) Study

Draft Protocol
August 2010

1. Background ........................................................................................................................................... 2
2. Objectives ............................................................................................................................................... 8
3. Study design .......................................................................................................................................... 9
4. Centers ................................................................................................................................................ 10
5. Case/control recruitment ....................................................................................................................... 11
6. Questionnaires ...................................................................................................................................... 14
7. Training of interviewers and interview procedure ................................................................................ 14
8. Log sheets ............................................................................................................................................ 15
9. Coding of data and computerized data entry ....................................................................................... 16
10. Collection of biological samples ......................................................................................................... 17
11. Collection of diagnostic and follow up information ............................................................................ 18
12. Laboratory analyses ............................................................................................................................. 19
13. Sample size/Power ............................................................................................................................... 19
14. Timetable ............................................................................................................................................. 21
15. Pilot study ........................................................................................................................................... 21
16. Study organization and management ................................................................................................. 22
17. Reports / Publications / Use of data ................................................................................................. 23
18. Ethical approval .................................................................................................................................. 23
19. References .......................................................................................................................................... 24

Appendices:
1. Questionnaires
2. Additional descriptive background
3. Eligible diseases for hospital controls
4. Log sheets
5. Biological sample preparation
6. Consent forms
1. BACKGROUND

Head and neck cancers are malignancies that arise in the oral cavity, pharynx, and larynx. It is the sixth most common type of cancer in the worldwide, representing about 6% of all cancer cases (Parkin DM, 2005). Worldwide, more than half a million head and neck cancer cases and 320,000 deaths due to head and neck cancer are estimated to occur each year (Ferlay 2008). The half million cases are comprised of 263,861 oral cavity cancers, 135,685 pharyngeal cancers and 151,219 laryngeal cancers. The age-standardized incidence rates are 12.8 and 3.9 per 100,000 males and females, respectively (Ferlay 2008). Global maps of incidence rates for men and women in Asia are shown in Figures 1. In Asia, high incidence rates of head and neck cancer are observed in India for men (22.6/100,000), in Sri Lanka for men (28.9/100,000) and in Bangladesh for women (13.6/100,000) (Ferlay, 2008).

Figure 1. Incidence rate of head and neck cancer in Asia

The estimated number of cases, deaths and incidence rates by sex for select countries of interest for the study is presented in Table 1. The ASRs for head and neck cancer in China are very low, whereas the ASRs in India are high. Another table with age-standardized incidence rates, number of cases based on cancer registry data for 1998-2002 (CIV-IX) is presented in the Appendix for reference.

<table>
<thead>
<tr>
<th>Country</th>
<th>Cases</th>
<th>Deaths</th>
<th>ASR female</th>
<th>ASR male</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>43269</td>
<td>24045</td>
<td>1.5</td>
<td>4.2</td>
</tr>
<tr>
<td>India</td>
<td>138149</td>
<td>101793</td>
<td>7.6</td>
<td>22.6</td>
</tr>
<tr>
<td>Indonesia</td>
<td>9955</td>
<td>5419</td>
<td>3.1</td>
<td>7.2</td>
</tr>
<tr>
<td>Korea, Republic</td>
<td>3078</td>
<td>1260</td>
<td>1.7</td>
<td>8.6</td>
</tr>
<tr>
<td>Malaysia</td>
<td>1428</td>
<td>680</td>
<td>4.6</td>
<td>9.1</td>
</tr>
<tr>
<td>Nepal</td>
<td>2495</td>
<td>1800</td>
<td>7.4</td>
<td>21.0</td>
</tr>
<tr>
<td>Philippines</td>
<td>3777</td>
<td>2540</td>
<td>4.4</td>
<td>8.1</td>
</tr>
<tr>
<td>Singapore</td>
<td>383</td>
<td>147</td>
<td>2.4</td>
<td>9.6</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>3806</td>
<td>2798</td>
<td>6.9</td>
<td>28.9</td>
</tr>
<tr>
<td>Thailand</td>
<td>6904</td>
<td>3644</td>
<td>7.1</td>
<td>11.6</td>
</tr>
<tr>
<td>Viet Nam</td>
<td>2539</td>
<td>1625</td>
<td>2.0</td>
<td>4.5</td>
</tr>
</tbody>
</table>
Treatment
Primary treatment varies with the anatomic subsite and stage of disease. For most early cancers, surgical resection is the cornerstone of treatment (Forastiere 2001). However, for certain anatomic sites such as the tonsils, the base of the tongue, and the floor of the mouth, as well as for all locally advanced cancers, radiotherapy is used, either alone or combined with surgery. Chemotherapy may be used in addition to radiotherapy.

Survival
Five year survival rates in Asia and Europe (for comparison) are presented in Table 2. Compared to survival rates in Europe, those in Shanghai, China are fairly high, whereas those in Thailand are much lower.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral cavity</td>
<td>19.4</td>
<td>39.3</td>
<td>55.2</td>
<td>49</td>
</tr>
<tr>
<td>Oropharynx</td>
<td>23.3</td>
<td>26.7</td>
<td>55.8</td>
<td>40</td>
</tr>
<tr>
<td>Hypopharynx</td>
<td>21.4</td>
<td>NA</td>
<td>24.8</td>
<td>25</td>
</tr>
<tr>
<td>Larynx</td>
<td>20.2</td>
<td>44.5</td>
<td>52.1</td>
<td>63</td>
</tr>
</tbody>
</table>

1. Sankaranaryanan 1998, 2. EUROCare-4

Recent publications suggest a role for HPV infection in head and neck cancer survival. Genetic polymorphisms in DNA repair genes, cell cycle genes, carcinogen metabolism, and growth factor pathways genes have been examined in relation to the progression and survival of UADT cancer patients (Hopkins 2008). Polymorphisms in XRCC1, FGFR and CCND1 were associated with UADT survival in at least 2 studies. However, Hopkins et al critiqued that these studies were small (<300 cases), often had inadequate descriptions of the study population, did not consider multiple comparisons appropriately and may be subject to publication bias. Thus to our knowledge, large scale follow up studies that would allow the simultaneous assessment of the role of HPV, genetics, treatment information, tobacco and alcohol use on head and neck cancer survival are not available, particularly in Asia.

Multiple primary cancers
Patients with head and neck cancer are at a higher risk of multiple primary cancers (MPCs) than patients with other primary cancers. UADT cancer patients also have higher risk of MPCs if they smoke tobacco or drink alcohol (Dikshit 2005). Treatment for the first primary oral cancer such as radiotherapy has also been suggested to contribute to the risk of MPC (Hashibe 2005a). Associations of several SNPs in metabolic genes for carcinogen activation and detoxication, including CYP1A1, GSTM1, and NAT2, as well as the DNA repair gene XRCC3 with the risk of MPCs following a first primary UADT cancer have been suggested (Gal 2005, Rydzanicz 2005).

An association between HPV status and the risk of MPCs among patients with oropharyngeal squamous cell carcinoma was reported from a retrospective series of 90 cancer patients (Licitra 2006). In a pooled analysis of 13 cancer registries, an excess risk of second primary cervical cancers in oropharyngeal cancer patients was observed (Chuang 2008). Conversely,
cervical cancer patients had an excess risk of second primary tumours of the oropharynx and tongue. These results, consistent with another registry based analysis (Rose Ragin 2008), suggest a shared aetiology of HPV infection for the multiple primary cancers. There is a limitation on the data from the cancer registries, since HPV information is not available at the individual level.

Environmental risk factors
The majority of oral, pharyngeal and laryngeal cancers are squamous cell carcinoma (SCC) in histology. The main risk factors for SCC of the head and neck are tobacco smoking and alcohol drinking. Studies show increased risks for smokers on the order of 3 to 4-fold for oral cavity and pharynx cancer, and 10-fold for laryngeal cancer (Vineis 2004). Tobacco use includes smoking of tobacco products such as cigarettes, cigars, and pipes, chewing smokeless tobacco products and using snuff tobacco. Local smoking tobacco products such as bidi and chutta, hand rolled Indian cigarettes are important. In regions including India and Taiwan, chewing areca nut and betel quid with or without tobacco are major risk factors for most head and neck cancers (IARC 2004b). Additional types of tobacco use include kreteks (clove flavored cigarettes) in Indonesia; and suipa, chilum or hookli (clay pipes) in Southeast Asia.

Alcohol consumption includes drinking of beverages containing ethanol such as wine, liquor, beer, and other local alcohol products. Relative to other alcohol related cancers, the risk conferred by alcohol drinking is strong for head and neck cancers (36). Consuming 50 grams of alcohol per day is thought to increase the risk of oral cavity and pharyngeal cancers by approximately 3-fold and the risk of laryngeal cancer by 2-fold relative to non-drinkers (37).

Together, tobacco and alcohol account for approximately 51% of head and neck cancers in the US, 84% of the disease in Europe and 83% of the disease in Latin America (Hashibe 2009). In India, the proportion of oral cavity cancer cases attributed to smokeless tobacco use is approximately 52.5% among males and 52.6% among females (Boffetta 2008). The combined impact of tobacco (cigarettes/day) and alcohol (drinks/week) consumption is greater than the sum of their individual effects and exceeds a multiplicative effect on the risk of head and neck cancers. Population attributable fraction estimates are not available for Asia at present.

Other risk factors include chewing betel quid and areca nut (without tobacco) for oral cavity cancer (IARC 2004a), HPV infection for oral and pharyngeal cancers (IARC 2007b) and low fruit/vegetable intake for head and neck cancers overall (Boeing 2006; Sapkota 2008). Several occupational substances or circumstances such as isopropanol manufacturing, inorganic acid mists containing sulfuric acid and mustard gas are suspected risk factors for laryngeal cancer (Siemiatycki 2004). Other possible risk factors for head and neck cancer include involuntary smoking (Lee 2008), sexual history (Smith 2004), marijuana use (Hashibe 2005b), poor oral hygiene (Rosenquist 2005), family history of head and neck cancers (Negri 2009) socioeconomic status (Conway 2008) and low body mass index (Kreimer 2006).

Genetics
Head and neck cancer studies on genetic variants or single nucleotide polymorphisms (SNPs) have focused on carcinogen metabolism, DNA repair and cell cycle genes. Of particular interest, are the alcohol dehydrogenase (ADH), and the aldehyde dehydrogenase (ALDH) gene families, both involved in alcohol metabolism. Among approximately 3800 upper aerodigestive cancer cases (head and neck + esophageal cancers) and 5200 controls, the alcohol dehydrogenase 1B (ADH1B) R48H (OR=0.56 (95% CI=0.47-0.66) and the ADH7 G92A alleles
(OR=0.68, 95%CI=0.60-0.78) were associated with a strong decreased risk of upper aerodigestive tract cancer (Hashibe 2008). A significant effect was also observed for a previously unstudied ALDH2 variant at position 248 in a series of approximately 1000 case-control pairs, especially among medium/heavy drinkers with the heterozygous (OR=1.76, 1.13-2.75) or homozygous variant (OR =5.79, 1.49-22.5), with a significant dose response for carrying variant alleles (p=0.0007) (Hashibe 2008). Another region of interest is chromosome 8q24, where several SNPs were associated with the risk of oropharyngeal cancer (rs6983267, OR=1.80, 95%CI=1.30-2.49) and laryngeal cancer (rs6983267, OR=2.04, 95%CI=1.12-3.72; Park 2008).

Several genome-wide association (GWA) studies have been initiated and are currently ongoing for study populations in Europe and the US. GWA studies investigate hundreds of 1000s of SNPs, which cover the majority (~80%) of common human genetic variation for SNPs with minor allele frequency > 5%. These studies have been based on hapmap (http://www.hapmap.org), a database of common SNPs. To our knowledge, GWA studies in Asian and Latin American populations for head and neck cancer are not available at present. Since there are differences in haplotype structures among different ethnic groups, it is vital to investigate common genetic variation in these populations.

Though GWA study results are starting to reveal SNPs that confer risk on various diseases, the number of SNPs that are truly associated with cancer risk appear to be lower than expected and will probably not explain the majority of the genetic cancer risk. A new challenge in the field is to study genetic variation with alleles that are less common (as rare as 0.1 - 0.5%). Currently, an international collaborative project to identify such rare alleles is underway (1000 genomes project; http://www.1000genomes.org) and has just started to release data. The statistical power necessary for examining genetic variants that are rare is expected to be large. An epidemiologic study that provides such statistical power will be crucial in understanding the genetics of head and neck cancers in Asian populations.

**Human Papilloma Virus (HPV)**

Human papilloma virus type 16 is now a recognized cause of oropharyngeal cancers (IARC 2007b). The evidence comes primarily from several large epidemiological studies that have analyzed associations of various HPV markers. HPV markers studied were (i) HPV DNA in biopsy tissues or oral cell scraping analyzed by southern blotting or highly sensitive PCR methods, (ii) antibodies to HPV 16 capsids analyzed by ELISA using HPV 16 major capsid protein L1-derived virus-like particles as antigens and (iii) antibodies to HPV 16 E6 and E7 analyzed by ELISA. HPV capsid antibodies are a cumulative marker of past and present HPV infection (Dillner 1999). Young females with new genital HPV 16 infection demonstrated by HPV 16 DNA positivity will seroconvert to about 60% within 6 months. Titers of HPV capsid antibody titers usually are rather low but lasting over many years. Mucosal HPV capsid antibodies are more prevalent in females than in males.

Antibodies to the oncoproteins E6 and E7 of HPV 16 and 18 are markers of invasive cancer expressing these viral oncoproteins (Lehtinen 2003). They are rare in the general population and among women with cervical cancer precursor lesions. In patients with invasive cervical cancer prevalence of these antibodies increases with stage (Zumbach 2000). Antibodies to HPV 16 E6 and E7 also develop in patients with invasive HPV 16 DNA positive head and neck squamous cell carcinoma (Zumbach 2000), particularly in patients with evidence of HPV 16 E6 expression (Kreimer 2005a).
The largest study on the association of HPV and head and neck cancer, coordinated by IARC, involving 1670 case patients (1415 with cancer of the oral cavity and 255 with cancer of the oropharynx) and 1732 control subjects, reported a prevalence of HPV DNA in 3.9% of specimens from the oral cavity and 18.3% of specimens from the oropharynx (Herrero 2003). When cases were compared to controls, a strongly increased risk was observed for antibodies against HPV 16 E6 and E7 proteins, for both cancers of the oral cavity (OR=2.9, 95% CI 1.7-4.8) and oropharynx (OR=9.2, 95% CI 4.8-17.7). A recent US study comprising 204 head and neck cancer cases and 326 controls, a five fold increased risk for HPV 16 E6/E7 antibodies was observed for oral cavity cancer (OR=5.1, 95%CI 1.2-22.4), and a 70-fold increased risk was observed for cancer of the oro-pharynx (OR=72.8, 95%CI 16.0-330) (Smith 2007). In a study including 100 oropharyngeal cases and 200 controls, oropharyngeal cancer risk was increased with HPV 16 (OR=14.6, 95%CI=6.3-36.6) and with oral infection with any of 27 types of HPV (OR=12.3, 95%CI=5.4-26.4) (D’Souza 2007). The exposure profile between HPV positive and HPV negative tumours may differ. For head and neck cancer patients who were HPV positive, a lesser role of tobacco and alcohol and a greater role of oral sex and marijuana were observed (Gillison 2008). The association of HPV and head and neck cancer has not been investigated with antibodies against HPV 16 E6 and E7 in Asia. Studies in Japan, China, Taiwan and India assessed HPV DNA in small case series; the largest series included 188 oral cavity tumors from Okinawa, Japan (Higa 2003). In a systematic review of presence of HPV in DNA in head and neck cancers, HPV 16 accounted for practically all of the HPV infections of the oropharynx whereas HPV 18 was also commonly observed in the oral cavity (Kreimer 2005b).

The presence of HPV infection may also alter the tumor profile and have implications for outcome. A higher prevalence of TP53 mutations has been observed in oral cavity cancers that are negative for HPV, consistent with the disintegration of the TP53 protein by the E6 oncoprotein (Dai 2004). Furthermore, individuals with HPV associated oral cavity cancer appear to have an increased survival but the reasons for this are unclear (Fakhry 2008, Smith 2008). The results were based on a small series of cases (<300 cases) and had limited capacity to explore stratifications by smoking status and other factors. Thus, confounding by tobacco use, differences in the development of second primary tumors and confounding by stage at diagnosis have not been ruled out as potential reasons for the association between HPV and survival.
Rationale for proposed study

- While the head and neck cancer incidence rates are low in China, the smoking prevalence among men is very high. Exploring the reasons behind these statistics may help further elucidate head and neck cancer etiology.
- Large scale follow up studies on HNC patients that would allow the simultaneous assessment of the role of HPV, genetics, treatment information, tobacco and alcohol use on head and neck cancer survival and multiple primary tumor development are not available in Asia.
- Follow up interviews after the initial interview will allow for the assessment of the role of lifestyle habits on prognosis and second primary cancers.
- Geographic variation in the association between HPV and head and neck cancer risk/survival needs to be investigated but there are no large-scale studies in Asia that can assess this association.
- Genome-wide association studies on head and neck cancer have focused thus far on European and US populations. Studies in Asian American populations are necessary to understand the genetics of head and neck cancers in these populations.
- There is a need for new large-scale studies to examine genetic variation that is rare (0.1 to 5%) relative to the currently studied genetic variation with single nucleotide polymorphisms.
2. OBJECTIVES

Our overall objective is to understand the role of lifestyle factors, genetics and HPV infection in the development of head and neck cancer particularly in Asia.

Our specific objectives are to:

- Obtain high quality biological samples (blood for DNA, tumor tissue)
- Follow up head and neck cancer cases with information on treatment and outcomes
- Conduct repeated interviews at check-up visits with updated information on lifestyle, and treatment information
- Recruit a large series of cases/controls (long term target 10,000 cases and 10,000 controls) for future rare allele genetic studies
- Apply a shortened lifestyle questionnaire to focus resources on the samples and the recruitment of a large sample size
3. STUDY DESIGN

Summary

1) Multi-center case-control study
2) Hospital and population controls depending on center
3) Questionnaire based lifestyle interviews
4) Blood samples from cases and controls, repeated sampling if possible
5) Biopsy samples obtained from cases
6) Clinical information on treatment
7) Follow up of patients

This case-control study will be conducted in centers in Asia. Each center will recruit a group of incident cases of head and neck cancer cases, comprising oral cavity, oropharynx and larynx sites. A comparable group of controls will be recruited at the same time. The interview of both cases and controls will be structured to obtain information on current and previous alcohol consumption, dietary habits, tobacco consumption and other lifestyle factors. Blood samples will be collected from all cases and controls. Pathology reports, paraffin embedded or fresh tumour tissue will be obtained from all cases. Written consent for participation will be obtained from all study subjects.
4. **STUDY CENTERS**

**Summary**

Approximately 8 centers are able to initiate the pilot study first and others will start later depending on the available fund.

![Figure 2. Map of centers in Asia](image-url)
5. CASE/CONTROL RECRUITMENT

Case selection:

<table>
<thead>
<tr>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Age ranges from 18 to 80.</td>
</tr>
<tr>
<td>2) Interview and biological sample collection may initially be based on clinical suspicion</td>
</tr>
<tr>
<td>3) Final diagnosis will be based on histological or cytological confirmation</td>
</tr>
<tr>
<td>4) Early identification will be achieved by periodical visits to the relevant hospital departments</td>
</tr>
<tr>
<td>5) Cases must be interviewed within 6 months of diagnosis</td>
</tr>
</tbody>
</table>

Cases will be identified from the participating hospitals as soon as possible after the diagnosis is made. In the presence of a strong clinical suspicion, a case will be interviewed and biological samples taken, pending subsequent confirmation. Early identification will be achieved through active search, i.e., by means of periodical visits to the hospital departments where cases are diagnosed or treated. The frequency of the visits will depend on the cancer burden of each department. A case may be included even if they have started treatment before recruitment, in such circumstances details of the treatment being given, and the duration of the treatment should be recorded. It is preferable if the biological samples can be collected before treatment for cases.

The local coordinator will maintain a record of cases identified and indicate the reasons for lack of interview or material collection (e.g., refusal, treatment).

All cases will be histologically or cytologically confirmed. Information on histological type will be collected and the largest series of histological types (squamous cell carcinoma and adenocarcinoma) will be analysed separately. A case cannot be counted twice, even if they present with 2 primary tumours within the period of recruitment.

Some patients may explicitly state that they would like to take part in the study but do not want to donate a blood samples; such patients can be included in the study.

Definition of cases
First primary tumors of the head and neck with the ICD-O sites specified below are eligible for the study. Second primary tumors and metastasis sites in the head and neck are not eligible. [Percy C, Van Holten V and Muir C. ICD-0: International classification of disease for oncology, Second Edition, World health Organisation, Geneva (1990)]

- C00 Lip
- C01 Base of the tongue
- C02 Other and unspecified parts of the tongue
- C03 Gum
- C04 Floor of mouth
- C05 Palate
C06 Other and unspecified parts of the mouth
C09 Tonsil
C10 Oropharynx
C12 Pyriform sinus
C13 Hypopharynx
C14 other sites in lip, oral cavity and pharynx
C32 (including all subcategories)

Sites NOT included are: cancers of the salivary gland (C07-08), nasopharynx (C11), thyroid (C79.3)

For centers including esophageal cancer, the following ICD-O sites are eligible:

C15.0 Cervical oesophagus
C15.3 Upper third of oesophagus
C15.4 Middle third of oesophagus
C15.5 Lower third of oesophagus
C15.8 Overlapping lesion of oesophagus
C15.9 Oesophagus unspecified

Control selection:

Summary

1) Frequency-matched by sex, 5 year age group (ex. 40-44 years, 45-49 years etc),
   ethnicity and residence area
2) Population controls will be selected where this is feasible
3) Hospital controls chosen from list of acceptable diseases
4) The proportion of hospital controls within a particular diagnostic group should not
   exceed 33%
5) Hospital controls should have been in hospital for less than 1 month when recruited

1-3 controls will be selected for each case recruited, frequency-matched by sex, age,
ethnicity, and residence area (same province and rural vs. urban). The number of controls
per case is left to the discretion of the center, based on the number of cases that can be
recruited (if the number of cases is small, higher number of controls should be recruited).
If the center has many cases from various provinces, then the controls will also need to be
from various provinces. In selecting the source of controls the study group has carefully
considered the advantages and disadvantages of the two main sources: hospital and
population controls.

Population controls are theoretically preferable and they will be selected in the centers
where it is feasible, although they often suffer from a low response rate and differential
recall from the cases. Population controls should be from the same city/county of the cases.

Hospital controls have the advantage that participation rates are usually very high and
recall of past lifestyle habits is usually similar to the cases. The suitability of a hospital
control group will be ensured by the following means:
Controls will be selected from a strictly defined list of non-chronic diseases unrelated to alcohol, tobacco or dietary practices (see appendix 3). This will include 1) endocrine and metabolic, 2) genito-urinary, 3) skin, subcutaneous tissue, and musculoskeletal disorders 4) trauma, 5) gastro-intestinal, 6) circulatory disorders, 7) ear, eye and mastoid disorders, 8) Plastic surgery cases, 9) nervous system diseases, 10) minor surgery, 11) ophthalmic and ear conditions, 11) lower back pain, and 12) urinary tract infection.

- The proportion of controls within a specific diagnostic group will not exceed 33%.
- They will be randomly chosen from among subjects admitted as in-patients or out-patients in the same hospital as the cases and only controls with a recently diagnosed disease will be accepted.
- Control’s current hospital admission was for a condition diagnosed within one month of the interview to avoid over sampling of long stay patients.

Using these selection criteria for controls, the main possible biases in this study, recall bias of past alcohol and dietary habits and participation bias of controls, will be minimized.
6. QUESTIONNAIRES

Initial Lifestyle questionnaire
All cases and controls will undergo an identical interview during which they will complete a lifestyle questionnaire (LQ). The lifestyle questionnaire is structured and will be used to collect the following information:

- Demographic details (age/sex/ethnicity/residence/education)
- Tobacco smoking history
- Passive smoking history
- Alcohol drinking history
- Drinking tea history
- Dietary habits
- Oral cavity health
- History of various diseases
- Family history of cancer
- Occupational history (brief)
- Residential history
- Sexual history

A copy of this lifestyle questionnaire is available in Appendix 1. Blood samples and tumor tissue if possible, will be taken at the time of interviews upon consent of the patient. Viable lymphocyte collection will be optional for the centers.

Follow up questionnaires for patients
All cases will undergo another interview after the initial interview, depending on the feasibility at the center:

- 1 year later
- 2 years later
- 3-5 years later

The follow up questionnaire is available in appendix 1 and includes questions on continued habits after diagnosis for tobacco/alcohol, as well as questions on quality of life.

7. TRAINING OF INTERVIEWERS AND INTERVIEW PROCEDURE

Summary

1. Training of interviewers will be the responsibility of the principal investigator in each center based on a common training procedure
2. Training will concentrate on maximizing completeness and accuracy
3. Specific guidelines for conducting the interview are provided for the interviewers

Before starting the interviews, each interviewer must attend a training course organized by their center. This training course will be based on a common training procedure implemented
by the study coordination team. Interviewers will receive instruction on the general aspects of the study, e.g. relationship with hospital personnel, choice of controls etc.

The training course must include information on the aim of the study, its organization, the participating countries and the time schedule of the study. Legal aspects for epidemiological studies and the necessity for confidentiality should also be discussed. Each interviewer should sign a document stating that he/she accepts the professional secrecy for all data obtained while working in this study. The interviewers should also be familiar with the methodology for selecting cases and controls.

The local questionnaire (LQ) is structured and so training for the interview will concentrate on maximising completeness and accuracy. This will be achieved by minimizing the extent of missing data and also by probing of subject responses to ensure that the answers received are clear and correct. The only reason missing data should occur is if a subject refuses to answer a question. If this happens then the interviewer should write “answer refused” by where the answer should be. Interviewers may ask additional probing questions to ensure that an answer is complete. This may include simply repeating the question or confirming the reply, or just pausing and giving the subject an opportunity to provide extra information. Interviewers may also seek clarification from the subject although the tone of the probe should be neutral. These aspects of interviewing will be introduced to the interviewer during sample interviews which are conducted in the feasibility stage of the study.

The interviewer may also write any comments which are relevant to the interview, e.g. reason for missing questions and comments on the quality of the questionnaire.

8. LOG SHEETS:

<table>
<thead>
<tr>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 A separate logsheet must be completed by the interviewer for all eligible cases and controls (including participating and non-participating patients)</td>
</tr>
<tr>
<td>2 A logsheet will also be completed with more detailed information for participating cases and controls</td>
</tr>
<tr>
<td>3 Logsheets will be stored separately to the interviews</td>
</tr>
</tbody>
</table>

The main record of all eligible cases and controls, including participating and non-participating patients, will be kept on a logsheet. This will be completed by the interviewer throughout the recruitment process. The logsheets include basic information on the patient, tumour histology (for cases) and collection of biological samples for participating patients. The logsheets can be updated with information about the collection of biological material as it is accumulated. The logsheets include basic information on the patient and the reason of non-participation for non-participating patients.

A copy of the log sheets to be used in the study can be obtained in Appendix 4.
9. CODING OF DATA AND COMPUTERISED DATA ENTRY

Summary

1. Each study subject receives a unique identification number
2. A common web-based data entry program will be used for storing data on the LQ
3. A second web-based data entry program will store information from the log sheets

Each individual in the study will receive a unique study number. This will be made up from a country code, a center code, and a subject code. The format for coding will be provided by the study coordination team, in the web-based data entry program that will be prepared and distributed to the centers. It will be noted at the beginning of the LQ. The case/control status of an individual will not be directly apparent from their study number. Instead this information will be stored on their log sheet.

As soon as possible after the interview, and preferably before another interview has occurred, the interviewer will read through all of the interview material and confirm that all responses given are clear and legibly recorded, and that all questions have been answered. If there does appear to be any missing data which are not due to patient refusal the interviewer should arrange to return to the subject at the earliest opportunity and obtain the missing information. All questionnaires will be administered as paper-based. Information from the interviews will be entered into a web-based data entry program.

Data entry program 1
If the information from the LQ is entered onto a database by someone other than the interviewer, the data entry person should have access to the interviewer, so that they can ask any questions if the questionnaire is unclear. The questionnaires should be typed in using the standard data entry package supplied by the study coordination team. This is a web-based data entry program and specific instructions on using the package will be supplied. If any center cannot access the web-based data entry program they should contact the study coordination team to discuss alternative programs or solutions.

Data entry program 2
A separate database will be supplied to store information from the log sheets including the identification number, their case/control status, the identification number of their matched pair, all clinical details, date of interview, and all the information relevant to the collection of biological samples. This information should be entered by the interviewer or the data entry personnel at the local study centers.
10. Collection of biological samples

Summary

1. Biosamples for collection include blood, biopsies, fresh-frozen tissue, paraffin-embedded tissue.
2. The 2×5ml EDTA tubes of blood will be obtained from all cases and controls, and stored at -80°C.
3. Whole blood will be used for the DNA extractions then frozen at -80°C.
4. Fresh biopsy samples will be obtained, when possible.
5. Viable lymphocyte collection will be optional for the centers.
6. Investigators will consult with local pathologists to ensure the availability of paraffin embedded blocks.
7. Investigators will try to ensure that histological slides are available from each case.
8. Central storage will be in the CIHCAMS for the Chinese centers. The study center should keep approximately half of the samples (if 2 tubes of blood are collected). Storage at 2 sites will ensure backup samples are available.
9. All biological samples will have barcode labels that will be provided by the study coordination team.

More detailed information on the preparation, storage and transportation of biological samples is set out in appendix 5, however centers should note that:

(i) Standardized tubes are to be used to minimize the amount of storage space needed in the centralized freezer.

(ii) Tubes are color-coded to facilitate identification.

(iii) Printed bar-code identification labels will be provided by the study coordination team and should be placed on all tubes containing biological samples.

Two 5-ml samples of blood will be collected with anticoagulant (EDTA) for. All samples will be stored at a temperature of -80°C (or at least -70°C depending on type of freezer, non-frost-free freezer preferred). The amount of time between collection and freezing should not exceed 12 hours.

When possible, fresh tumor tissue should be obtained from the cases. The details, including whether the tissue samples are obtained before or after chemo- or radiation therapy, should be documented. If fresh samples are obtained during biopsy or surgery they should be snap frozen and stored at -80°C within 20 minutes of being taken and be stored preferably in liquid nitrogen. Whether tumor tissue samples are snap frozen or not will be recorded on the log sheet. Investigators will also consult with local pathologists on the availability of paraffin-embedded tissue from all cases. Either a full block, part of a block or shavings from a block may be obtained. Note: shavings must be at least 10μm thick, this is 2-3 times the usual thickness. The investigator in each center should also try to obtain from the local pathologist a histological as well as all immuno-histochemistry reports. One central pathologist will use this to conduct a central review and reclassification of all cancers.

Assuming local ethical committee approval, at least 50% of samples from all cases and controls will be sent to the central laboratory for central storage in nitrogen vapor (-120°C) and extraction of DNA. This will be done using an automated robot in place in the laboratory.
A record sheet should be completed to record the shipment of all biological samples to the study coordination team (appendix 6). All samples that are stored centrally will remain the property of the individual center and can be returned upon request. Members of the study group may propose further projects based on these samples, although written consent will be required from all centers before any samples are released.

Each center will have the opportunity to consult with their local ethical committee for approval of each specific project if this is felt necessary.

11. Collection of diagnostic and follow up information

Summary
1. Initial diagnosis may be based on clinical suspicion.
2. Final diagnosis will rely on histological or cytological confirmation.

Initial diagnosis of cases will be based on clinical suspicion of UADT cancer. This will be sufficient to enroll a case into the study, identify a matched control, and proceed with the interview. A final diagnosis of UADT will however be dependent on a histological or cytological confirmation of disease. In some centers histological/ cytological information will only become available after several months. Cases who do not have a diagnosis confirmed by these means will be included in the final analysis although separate analysis will also be conducted after excluded them. The collection of this extra diagnostic information will be the responsibility of one individual within each center.

This diagnostic information from all the cases will be included on their logsheets (appendix 4) and also added to the web-based data entry system.

The comparability of the diagnostic and histologic information across centers within each country will be assessed in the pilot study. In the case that this information is considered to vary substantially across centers, we will consider conducting a validation study by collecting pathology reports for a proportion of cases (~10%) across centers within a country.

Follow up

Follow up information will be obtained by linking the patient’s data with vital statistics, cancer registry files and medical records, depending on the available sources at each center. The linking procedure of the information will depend on each center. Data linking will be by personal ID number where available, by name, date of birth, and sex of the patient, or possibly by medical ID.

The patient will be considered lost to follow-up if he/she cannot be tracked by any of the vital statistics records, cancer registry, medical chart/path report at the hospital, or any contact information.

The study coordination team will provide a web-based database to enter follow up information. Some of the fields will require translation into English by the abstractor at the time of data entry. The format of the database will be discussed with the study group.
12. LABORATORY ANALYSES

Genetics
We will conduct genome-wide association analysis in a later stage. We will also start planning a study based on the 1000 genomes project, to examine rare genetic variants (0.1 to 5%). The data from the 1000 genomes project is expected to be released quarterly.

HPV
The HPV antibody analysis will be conducted under supervision of Dr. Michael Pawlita at the DKFZ (German Cancer Research Center, Heidelberg). At least 100 µl of plasma will be required for the analysis. The HPV types and proteins for which antibodies will be analyzed are shown in Table 4.

<table>
<thead>
<tr>
<th>HPV Type</th>
<th>Proteins</th>
<th>Genus</th>
<th>Species</th>
<th>No Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L1</td>
<td>µ</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>L1</td>
<td>γ</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>L1 E6 E7 E1 E2</td>
<td>α</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>L1 E6 E7</td>
<td>β</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>L1 E6 E7</td>
<td>α</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>L1 E6 E7 E1 E2 E4</td>
<td>α</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>18</td>
<td>L1 E6 E7 E1 E2 E4</td>
<td>α</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>31</td>
<td>L1 E6 E7</td>
<td>α</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>33</td>
<td>L1 E6 E7</td>
<td>α</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>35</td>
<td>L1 E6 E7</td>
<td>α</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>38</td>
<td>L1 E6 E7</td>
<td>β</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>45</td>
<td>L1 E6 E7</td>
<td>α</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>49</td>
<td>L1</td>
<td>β</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>52</td>
<td>L1 E6 E7</td>
<td>α</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>58</td>
<td>L1 E6 E7</td>
<td>α</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>77</td>
<td>L1</td>
<td>α</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>**Total</td>
<td></td>
<td></td>
<td></td>
<td>48</td>
</tr>
</tbody>
</table>

13. SAMPLE SIZE (POWER)

The statistical power to detect the risk of oral cancer with different ratios, 1:1, 1:2 and 1:3, of cases:controls is shown in the figures below. We will be able to detect an OR of 2.0 for risk factor, for the range of proportion ‘exposed’ from 1-95%. With the large sample size, we should have considerable power to examine our primary outcomes.
Figure 4. Sample size required for detecting ORs, for case:control ratio of 1:1, by proportion exposed (power=0.80, alpha=0.05)

Figure 5. Sample size required for detecting ORs, for case:control ratio of 1:2, by proportion exposed (power=0.80, alpha=0.05)

Figure 6. Sample size required for detecting ORs, for case:control ratio of 1:3, by proportion exposed (power=0.80, alpha=0.05)
Estimates on the number of eligible cases in Asia, by center

<table>
<thead>
<tr>
<th>Study center</th>
<th>Estimated number of eligible cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing - Cancer Institute and Hospital</td>
<td>100</td>
</tr>
<tr>
<td>Beijing - Union Medical College hospital</td>
<td>200</td>
</tr>
<tr>
<td>Henan</td>
<td>50</td>
</tr>
<tr>
<td>Tianjing</td>
<td>150</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>70-80</td>
</tr>
<tr>
<td>Shanghai</td>
<td>80-90</td>
</tr>
<tr>
<td>Yunnan</td>
<td>300</td>
</tr>
<tr>
<td>Heilongjiang</td>
<td>200</td>
</tr>
<tr>
<td>Hubei</td>
<td>200</td>
</tr>
<tr>
<td>Gansu</td>
<td>200</td>
</tr>
<tr>
<td>Liaoning</td>
<td>100</td>
</tr>
<tr>
<td>Fujian</td>
<td>150</td>
</tr>
<tr>
<td>Shanxi</td>
<td>200</td>
</tr>
<tr>
<td>Hainan</td>
<td>90</td>
</tr>
<tr>
<td>Sichuan</td>
<td>400</td>
</tr>
<tr>
<td>Taiwan</td>
<td>353</td>
</tr>
<tr>
<td>Korea</td>
<td>129</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>Approximately 3000</strong></td>
</tr>
</tbody>
</table>

14. TIMETABLE

Now: IRB submissions

July 21, 2010: Web-based data entry system was available for testing.

September 15, 2010: target date for IRB approval at centers, Cancer center in Beijing and University of Utah. Biosample collection tubes will be sent from the Cancer Center in Beijing and barcoded labels will be sent from the University of Utah. Training of interviewers to start.

October 1, 2010: target date for centers to start pilot study

November 2010: pilot study to continue

November 1, 2010: University of Utah Health Studies Fund grant due

December 1, 2010: target date for centers to submit their pilot data for grant application

February 5, 2011: NCI grant due

15. PILOT STUDY

A pilot phase will be conducted in each center over a one-two-month period at the beginning of the study. The objectives of the pilot study are as follows:

(i) Verify that translated questionnaires are comparable with the central questionnaire

(ii) Implement a common training procedure for interviewers

(iii) Ensure that the tools to identify and recruit cases and controls are working efficiently

(iv) Confirm that the interview procedure for cases and controls works efficiently
Assess whether the diagnostic and histologic information will be comparable across centers

Ensure that collection of biological samples of cases and controls is viable

Construct and test databases for storage of all interview and related information

Assure that the follow up interviews with additional blood sample collection are feasible

The questionnaire should be translated into the local language at each center. These will be returned to the co-ordinating center, which will review the translation in order to verify the content of the local language version. The training for the interviewer will concentrate on maximising completeness and accuracy of the questionnaires.

At the end of the pilot phase each center should report the following information:

- The number of potential cases
- Total number of cases identified and asked to participate
- Response rates among cases and controls
- Details of particular problems with interview
- The percentage of cases and controls with biological samples
- Processing of biological samples

A web-based data entry system will be developed by the study coordination team and tested in each center. These will comprise a database for all questionnaire information, as well as a database for logsheets.

16. STUDY ORGANISATION AND MANAGEMENT

The study group will be comprised of a principal investigator from each institution and the study coordination team. The study group will be responsible for all decisions regarding the conduct of the study. It will approve the final version of the protocol, and will jointly agree on the work required for each center including specific responsibility for the individual project components, and the project planning and timetable. The study coordination team will be responsible for overall co-ordination of the study and will act as secretariat of the Study Group. The study coordination team will be responsible for monitoring the progress of the study, organization of Study Group meetings, preparation of minutes and study reports, and also financial co-ordination and distribution of funds. A joint review of the study progress, (including methodological problems, recruitment and results) will be conducted by the entire Study Group at the study group meetings.

Site visits will be carried out by one of the study coordination team staff each year, approximately mid-way between study meetings. The purpose of this site visit will be to review center-specific progress with data collection and other project components, as well as anticipate future problems.
17. REPORTS / PUBLICATIONS / USE DATA

The study coordination team will prepare and circulate two newsletters each year to all members of the study group which will include: progress reports submitted from each center, minutes of meetings, updates and changes to the protocol. A webpage will be created for the study and all documents distributed to the study group will be available for download.

The Study Group will take on the responsibility for publishing results, it is anticipated that several papers will be submitted for publication in high impact international medical journals. The ownership of the data will remain in the individual contributing centers, which will retain the right to perform separate analyses on their part of the data.

Authorship guideline

Manuscripts prepared with the overall study data (i.e., including data from all centers) will include 2 coauthors from each center. The 2 authorship positions for each center can be circulated among the center study group, at the discretion of the center PI (i.e., for each paper, different members of the center study group can be listed for the 2 positions). The writing committee for each paper (i.e., authors taking the lead on the manuscript) will decide the order of the authorship (for the first, corresponding and last positions). The order of the coauthors from the centers will be varied across papers (by number of subjects contributed to study, alphabetical, randomized) for variation; this will be overseen by the study coordination team.

The authorship guideline for manuscripts prepared at the center level (i.e., data from one center) is left to the discretion of the center PI (i.e., number of coauthors is not restricted).

18. ETHICAL APPROVAL

Each local center will obtain ethical approval for the conduct of the study from their local ethical committee and will provide the study coordination team with a copy of the written approval. In addition the international study will also be submitted to the University of Utah and CICAMS ethical committee for approval.

A draft consent form (appendix 6) has been provided as a guide. This consent form does not need to be used if consent forms used in previous studies are available at the center institute. The final version of the consent form that will be used for the study, should be forwarded to the study coordination team for the ethical approval application.
19. REFERENCES


IARC monographs. Tobacco smoke and involuntary smoking. IARC Monogr Eval Carcinog Risks Hum 2004b; volume 83.

IARC monographs. Smokeless tobacco and some tobacco-specific N-nitrosamines. IARC Monogr Eval Carcinog Risks Hum 2007a; volume 89.

IARC monographs. Human Papillomavirus. IARC Monogr Eval Carcinog Risks Hum 2007b; volume 90.


