

CHAPTER 4

MATERIAL AND METHOD

4.1 Study design

This is a cross sectional study with case-control design involving hospital-based subjects matched for age and sex to identify the risk factors including genetic polymorphism (GSTM1, GSTT1 and CYP1A1) in the development of oral cancer.

4.2 Conduct of study

This study consists of two parts. The first part consists of data collection on the subjects' socio demographic characteristics, risk habits and dietary pattern using a structured questionnaire. The second part consists of laboratory work to assess the polymorphism of GSTM1, GSTT1 and CYP1A1. The flowchart depicting data collection and genetic analysis is illustrated in Figure 4.1.

4.3 Recruitment of subjects

4.3.1 Cases

All new patients with oral cancer attending the selected centers in Jakarta were included as cases. Oral cancers are oral malignancies diagnosed histologically as squamous cell carcinoma. The histological classifications of oral cancer used in this study followed IARC (2005) as follows as: "a malignant epithelial neoplasm exhibiting squamous differentiation as characterized by the formation of keratin

and/or the presence of intercellular bridges”. The type of oral cancer for all cases in this study is Squamous Cell Carcinoma (SCC).

All patients with newly diagnosed squamous cell carcinoma of the tongue, gingiva, floor of the mouth, lip and other parts of the oral cavity (ICD-10 C00-C06, thereafter referred to “oral cancer”) during the period 1st January 2005 to 1st April 2006 (15 months) at selected centers in Jakarta (deemed as research centers) were included in this study. The research centers were 5 hospitals in Jakarta (Approval letters to conduct the study at the selected centers are attached in Appendix A):

- a. Cipto Mangunkusumo Hospital (Central Jakarta)
- b. Dental Hospital of Trisakti University (West Jakarta)
- c. Fatmawati Hospital (South Jakarta)
- d. Lembaga Kedokteran Gigi TNI AL (LADOGI) (East Jakarta)
- e. Gatot Subroto Hospital (RSPAD Gatot Subroto) (North Jakarta)

The patients, who had oral cancer but did not satisfy the criteria of SCC histologically, were excluded from this study.

4.3.2 Controls

Controls were selected among non cancer patients who attended the selected centers for minor ailments. Two control patients were matched to each case on the basis of sex and age within 5 years.

Non-Indonesian citizens and those who had cancer, cardiovascular disease and were currently undergoing treatment were excluded from this study.

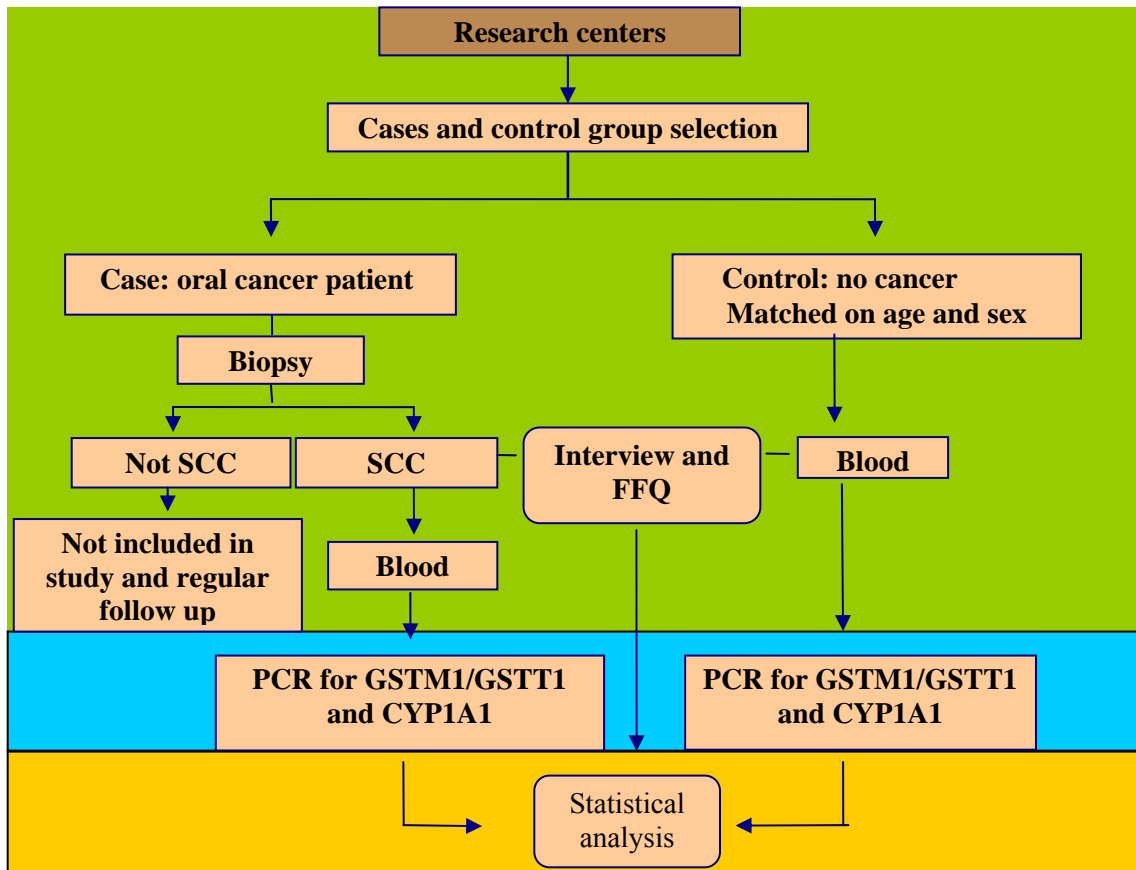


Figure 4.1 Flowchart of data collection

4.3.3 Number of subjects in study population

The number of subject was obtained from sample size calculation for cross-sectional, cohort and clinical trials by OpenEpi version 3.04.04, with significance level (α) = 0.05; test of power (β) 80% and Confidence Interval 95%. Based on the above formula, it was calculated that the required number of matched (1:1) case and controls were 89 pairs (89 cases and 89 controls). In order to ensure that the power of study is at least 80% in the event that the number of cases obtained is less than 89, it was decided that the number of control be increased to two for every case. The sample size calculation is attached in Appendix B.

4.4 Questionnaire survey

4.4.1 Components in questionnaire survey

There were two major components in the questionnaire that were investigated :

(1) Risk factors namely smoking, alcohol consumption and betel quid chewing and (2) Dietary habits namely daily food pattern, food preparation and food frequency. The complete questionnaire is as in Appendix C.

4.4.1.1 Risk factors

This component includes 3 major risk factors to be investigated:

4.4.1.1.1 Smoking (nominal scale).

Never smoker/non smoker were defined as those who never smoked or smoked less than 100 cigarettes in their life time.

Smokers were those who currently practiced the habit or had quit the habit for more or less than 1 year before the interview date.

The smoking duration was obtained from the number of years of smoking, or number of years smoked prior to stopping. The smoking duration was divided into 4 subgroups based on five year intervals. **Stick-years of exposure** followed the **Brickman index**, which was calculated as numbers of cigarettes smoked per day multiplied by smoking duration. A light smoker refers to a consumption of 1-299, moderate was 300-599, and heavy smoker was > 599. **Number of sticks per day** was classified into 3 categories: low for those who smoked 0-10 sticks/day, medium 11-20 sticks/day and high >20 sticks/day. **The type of tobacco** was classified as *kretek* (mixture of tobacco, cloves and sauces),

cigarettes, cheroots, *kawung* (hand rolled tobacco), pipe, *menyan* (hand rolled tobacco and inflorescence).

4.4.1.1.2 Alcohol consumption (nominal scale).

Never drinkers/non drinkers were those who ingested alcohol less than once a month. **Drinkers** were defined based on the same criteria employed for smoking. **Number of bottles per day** was calculated and divided into 3 categories: none for non alcohol drinker, 1-4 bottles/week and >5 bottles per week. **The duration of drinking alcohol** was obtained from number of years of drinking or number of years drank prior to stopping. Thereafter the duration of drinking alcohol was divided into 4 subgroups based on five year intervals. **Type of alcohol** was divided into: beer (including branded, traditional or local beer), wine and spirits (including whisky).

4.4.1.1.3 Quid chewing (nominal scale).

Never chewers/non chewers were defined as those who chewed less than once a month. **Chewers** were defined based on the same criteria employed for smoking. **Number of betel quids per day** was divided into 2 categories: none for never chewed betel quid and chewed 1-10 quids/day. **The duration of quid chewing** was obtained from number of years of chewing, or number of years chewed prior to stopping. Thereafter the duration of quid chewing was divided into 2 groups: none for non chewers and those who chewed more than 1 year. **The type of quid** was classified into 7

types based on the mixtures of quid: chewing areca nut only quid, tobacco only quid, areca-tobacco quid, betel leaf with tobacco, betel leaf with areca nut, betel leaf with areca nut and tobacco and betel leaf, tobacco, areca, lime and other constituents.

4.4.1.1.4 Dietary habit

This component includes 3 major factors to be investigated:

a. Daily food pattern (nominal scale)

Ovo-vegetarian refers to patient who is a vegetarian but consumes egg products. **Lacto vegetarian** refers to patient who is vegetarian but consumes milk products. **Ovolacto vegetarian** refer to patient who is a vegetarian but consumes both egg and milk products. **Pure vegetarian** refers to a patient who only consumes vegetables and **non vegetarian** refers to a patient who consumes meat products and may include vegetables. Furthermore, this component also included **the frequency of intake of home-cooked food, fast food, and frozen food**. For the last three questions, the frequency was classified into low (if subject consume 2-3/month, seldom or never), medium (2-3 week) and high category (>3 times/week).

b. Food preparation method (nominal scale)

There were 4 questions on the method of food preparation. The first question relate to how food was prepared based on 5 categories: steamed, fried, barbequed, roasted, microwaved and others. Secondly, the temperature of food usually consumed was

classified into four categories: very hot, hot, normal, cold. The type of oil usually used to prepare food was enquired and divided into 3 groups: (1) animal fat/butter, (2) vegetable/soy/sunflower oil/palm oil, and (3) margarine. The last question enquired on the frequency of oil reused/reheated, and the response was divided into 3 categories: 1-2 times, 3-4 times, > 4 times.

c. Food Frequency (nominal scale)

Food frequency (FF) consisted of 103 foods, food groups or preparations. Each participant in the study was asked to respond to each item of the FFQ, and the mean daily equivalent was computed. Once a day scored 1, 2-3/day scored 2, 1/week scored 0.14, 2-3/week scores 0.29, 1/month scored 0.03, 2-3/month scored 0.07 and less than once a month, seldom, and never scored 0. The foods in the questionnaire were then classified into the following 15 food groups:

1. Grain (rice, bread, noodle, sweet potato, cassava, sago and corn).
2. White meat (chicken, duck, bird, rabbit).
3. Red meat (beef and lamb).
4. Seafood (fish, prawn, squid).
5. Cooked vegetables (leafy green, beans, cruciferous, brinjal, pumpkin, bamboo shoots, and radish).
6. Raw vegetables (cucumber, others (*petai, jengkol*)).
7. Dairy product (cheese, margarine, yoghurt, skim and powder milk).

8. Fruits (banana, papaya, water melon, apple, honeydew, mango, pineapple, jackfruit, guava, orange and fruit juices).
9. Fermented food (salty egg, salty fish, fermented prawn, soya ketchup, fermented fruit, and pickles).
10. Snack which contains high sugar and fat (desert, sweets and cakes).
11. Canned food (fish, tomato paste, mushroom).
12. Processed food (sausages, nugget, and meat/fish balls).
13. Fast food (chips, hamburger, pizza, instant noodle, ready made spices).
14. Other: Monosodium glutamate (MSG).
15. Drinks (coffee, tea, and canned or carbonated drink).

These grouping were based on the similarity of nutrient content. Subsequently, the dietary pattern was analyzed using factor analysis (Marchioni, 2007). Subjects were assigned scores to indicate the degree to which their diet correlated to each factor retained from factor analysis. Factor scores were categorized into tertile based on the entire study population. Furthermore, the correlation of each component retained from factor analysis and oral cancer was analyzed using chi square and logistic regression to obtain the risk.

4.4.2 Validation of questionnaire

Prior to data collection, two specialists in Oral Medicine validated the questionnaire. A copy of the questionnaire as well as the aims and objectives of

the study were given to each of them and they were required to identify the information obtainable from each question for its content. The questionnaire was validated to:

- a. Ensure agreement between the specialist and the researcher with regards to the information that could be retrieved from each question.
- b. Identify irrelevant and ambiguous statements and instructions.

Following the validation process, the questionnaire was modified.

4.4.3 Pre-test of questionnaire

The validated questionnaire was then pre-tested on 15 patients who attended Trisakti Dental Hospital. The purpose of the pre-test was to:

- a. Clarify the questions in the questionnaire.
- b. Identify the inadequacy of the responses provided.
- c. Check the sequencing and flow of the questions.
- d. Assess the time needed to complete the questionnaire.
- e. Check the overall conduct of the survey.

There were no further modifications made following the pre-test.

4.4.4 Conduct of questionnaire survey

All new patients suspected of having oral cancer attending the selected centers in Jakarta were invited to join the study. Patients were firstly explained about the rationale and objectives of the study. Patients who agreed to participate will sign a consent form, witnessed by one of their family members or neighbours. An incisional biopsy was carried out, following which the tissue was sent to the pathologist (VY) at the Department of Pathology Anatomy, Faculty of Medicine,

University of Indonesia for histopathological assessment. A confirmatory assessment was undertaken by Prof Rosnah Mohd Zain (RBZ) at Faculty of Dentistry University of Malaya to confirm the diagnosis. The patient was included as a case if the histopathological result satisfied the diagnosis of SCC.

Subjects who conformed to the histopathological criteria of SCC, were subjected to an interview using the questionnaire. The interviewer obtained detailed information on sociodemographic, risk habits and diet in both cases and controls. All interviews of the case subjects were carried out prior to any major medical procedures that could potentially affect patient's ability to communicate or to recall the information

The same procedures of data collection were carried out among potential controls after they have read the patient's information sheet and signed the consent form (Appendix D).

4.5 Laboratory procedure

4.5.1 Tissue collection and storage

Tissue obtained by incisional biopsy from the case subject was stored in 10% formalin solution and sent for histopathological assessment. Histopathology diagnosis was undertaken by a pathologist (VY) at University of Indonesia and reconfirmed by RBZ at University of Malaya. The remaining paraffin blocks were kept in the Department of Pathology Anatomy, Faculty of Medicine, University of Indonesia for future study.

4.5.2 Blood collection and DNA Isolation

Five mL of intra venous blood was drawn from each case and control subjects and placed in EDTA tubes. The whole blood was centrifuged and separated from the buffy coat (second layer) for DNA extraction. This was performed at the Health Research and Development Department of Ministry of Health Indonesia in Jakarta. In case that blood samples were not processed immediately, the DNA extraction was performed using the whole blood. Thereafter the PCR assay was undertaken at the Cancer Research Initiatives Foundation (CARIF), Subang Jaya Medical Centre, Subang Jaya, Malaysia. In cases of partially damaged or lysed blood samples, they were still centrifuged and DNA extraction was carried out from 300 µl whole blood. The remaining blood in EDTA tubes were kept at -20°C at University of Trisakti Indonesia (as the main centre) for future investigations.

Genomic DNA was extracted from peripheral blood of case and control subjects using QIAamp DNA mini kit (Qiagen, USA). The DNA extraction protocol using Qiagen kit is attached in Appendix E. Subsequently, the DNA concentration was measured using Nanodrop and kept in 1.5 mL Eppendorf tube at -80°C for the genotyping assay. The genomic concentration of samples is illustrated in Appendix F.

4.5.2.1. Genotyping Assay of GSTM1 and GSTT1

PCR assay set up

A “Master Mix” was prepared using 1.5 mL tubes, consisting of two kind of mixing, the first master mix was 45 µl and the 2nd master mix

was 5 μ l. All the reagents were thawed and stored in ice. The amount of reagents required for each assay was:

1st Master Mix

| | |
|----------------------------------|-------------|
| GSTM primer forward (10 μ M) | 1.5 μ l |
| GSTM primer reverse (10 μ M) | 1.5 μ l |
| GSTT primer forward (10 μ M) | 1 μ l |
| GSTT primer reverse(10 μ M) | 1 μ l |
| Albumin forward (10 μ M) | 0.5 μ l |
| Albumin reverse (10 μ M) | 0.5 μ l |
| dNTPs (10 μ M) | 1 μ l |
| MgCl ₂ (25mM) | 3 μ l |
| 10x Buffer | 4 μ l |
| dH ₂ O | 28 μ l |
| | ----- |
| | 42 μ l |
| DNA | 3 μ l |
| | ----- |
| | 45 μ l |

2nd Master Mix

| | |
|------------------------------|-------------|
| Taq Polymerase (5U/ μ l) | 0.5 μ l |
| 10x Buffer | 1 μ l |
| dH ₂ O | 3.5 μ l |
| | ----- |
| | 5 μ l |

Template: 3 μ l of DNA from sample extract was added to each sample tube. For the negative control, dH₂O was used instead of DNA. The homozygous null polymorphisms GSTM1 and GSTT1 were determined using multiplex PCR, slightly modified from Nair et al.

(1999), for simultaneous replication of both genes. The co-amplification of an albumin gene fragment served as an internal positive control for a successful amplification reaction.

For analysis of n number of sample extracts, $2n+2$ assays were necessary to catch for pipeting error. Aliquots of 45 μ l of the 1st Master Mix were aliquoted into 2 $n+2$ PCR tubes. The reaction mix was preheated in a Thermocycler for 2 minutes, followed by the addition of 5 μ l from the 2nd Master Mix to each sample tube.

PCR was performed in a 50 μ l reaction volume containing 20mM Tris-HCL pH 9.0, 50 mM KCl, 1.5 mM MgCl₂, 200 μ M dNTPs, GSTM1 primers at 3 μ g/ml each, GSTT1 primers at 2 μ g/ml each, albumin primers at 1 μ g/ml each, and 50–100 ng of genomic DNA. It was placed in a preheated PCR machine at 99.9°C for 2 minutes. Taq DNA-Polymerase (2.5 U; Go Taq, Promega, USA) was added to each tube, and the PCR was performed in the GeneAmp PCR system 9700 (Applied Biosystems, USA). After an initial denaturation at 95°C for 5 minutes, amplification was carried out for 40 cycles at 94°C for 1 minute, 64°C for 1 minute and 72°C for 1 minute, followed by final elongation at 72°C for 7 minutes. The samples were then ready for amplification. The sequences for the primer pairs are as follows: GSTM1 sense GAACTCCCTGAAAAGCTAAAGC and antisense GTTGGGCTCAAATATACGGTGG. GSTT1 sense TTCCTTACTGGTCCTCACATCTC and antisense TCACCGGATCAGGCCA

GCA. Albumin sense GCCCTCTGCTAACAAAGTCCTAC and antisense GCCCTAAAGAAA ATCGCCA ATC.

Gel Electrophoresis

The PCR products were electrophoresed in 2% agarose gel. A 100bp DNA ladder was used to range the sizes of the PCR amplicons. The preparation is attached in Appendix G.

Criteria for evaluation of GSTM1 and GSTT1

The null polymorphism of GSTM1 and GSTT1 were determined using a modified multiplex PCR approach for simultaneous replication of both genes for molecular analysis. The co-amplification of an albumin gene fragment served as an internal positive control for a successful amplification reaction. The genomic DNA was checked using 2% agarose gel. A 350 bp DNA fragment corresponding to the albumin gene product provided as an internal positive control for each reaction and can be seen in all PCR reactions. A 220 bp product was present only in those samples containing the GSTM1 gene (wild type allele), while 480 bp product was present only in those samples containing the GSTT1 gene (wild type) samples (Figure 5.5).

The criteria of GSTM1 and GSTT1 in statistical analysis in this study are divided into 2 categories:

1. Wild type if the genes are present
2. Polymorphism if the genes are not present or it is called null.

4.5.2.2. Genotyping Assay of CYP1A1

Isoleucine/valine polymorphism of CYP1A1 was determined using PCR and RFLP, modified from a method done by Sreelekha et al, (2001). The master mix was prepared as below:

PCR assay set up

1st Master Mix

| | |
|------------------------------------|------------|
| CYP1A1 primer forward (10 μ M) | 1 μ l |
| CYP1A1 primer reverse (10 μ M) | 1 μ l |
| dNTPs (10 μ M) | 1 μ l |
| MgCl ₂ (25 mM) | 3 μ l |
| 10x Buffer | 4 μ l |
| dH ₂ O | 32 μ l |
| | ----- |
| | 42 μ l |
| DNA | 3 μ l |
| | ----- |
| | 45 μ l |

2nd Master Mix

| | |
|------------------------------|-------------|
| Taq Polymerase (5U/ μ l) | 0.5 μ l |
| 10x Buffer | 1 μ l |
| dH ₂ O | 3.5 μ l |
| | ----- |
| | 5 μ l |

PCR was performed in a 50 μ L reaction volume containing 20mM Tris-HCL pH 9.0, 50 mM KCl, 25 mM MgCl₂, 200 μ M dNTPs, and 100-200 ng of genomic DNA. It was placed in a preheated PCR machine at 94°C for two minutes. At amount of 0.5 μ l of Taq DNA-

Polymerase (5U; Pormega, USA) was added to each tube, and the PCR was performed in a GeneAmp PCR system 9700 (Applied Biosystem, USA). After an initial denaturation at 94°C for 5 minutes, amplification was carried out for 30 cycles at 94°C for one minute, 68°C for one minute and 72°C for one minute, followed by final elongation at 72°C for ten minutes. The sequences for the primer pairs are as follows: CYP1A1 sense AAAGGCTGGGTC CACCCTCT and antisense CCAGGAAGAGAAAGACCTCCC AGCGGGGCCA. The PCR product of CYP1A1 was run in 2% agarose gel and was evident by the presence of band at 370bp (Figure 5.6) and lane B showed absence of band which was the negative control.

Restriction Fragment Length Polymorphism Analysis (RFLP)

Restriction enzyme (RE) digest was used to distinguish the genotype of CYP1A1. Samples successfully amplified for CYP1A1 were used for restriction enzyme (Nco1) digestion. After RE digest, DNA purification was done by using QIAamp DNA purification kit (Qiagen, USA). The DNA purification protocol using Qiagen kit (Qiagen, USA) is attached in Appendix I. Upon purification, digestion was done in 20µl reaction with Nco1 (5 U/µl) at 37°C for one hour and the enzyme was incubated by heat at 65°C for 10 minutes. The Master Mix for restriction enzyme was prepared as listed below. Lambda DNA was used as a positive control for restriction digestion. The digestion products (polymorphisms) were electrophoresed on a 4% agarose gel

(Figure 5.7) stained with ethidium bromide and visualized using Chemmimager (Alpha Innotech, USA).

“Master Mix” for Restriction Enzyme (Nco1)

| | |
|--------------------|-------------|
| 10x Buffer | 2 μ l |
| Nco1(10U/ μ l) | 0.5 μ l |
| dH ₂ O | 2.5 μ l |
| DNA | 15 μ l |
| | ----- |
| | 20 μ l |

Criteria for evaluation of CYP1A1 polymorphisms

Polymorphisms of CYP1A1 were shown in the PCR product which was digested by Nco1 restriction enzyme. Single band of 232 bp product indicates the wild type allele (isoleucine/isoleucine) of CYP1A1. The presence of 232 bp and 263 bp (isoleucine/valine) or a single band of 263 (valine/valine) indicates a polymorphism (Figure 5.6 and Figure 5.7).

The term criterion for CYP1A1 in statistical analysis in this study is also divided into 2:

1. Wild type is considered if the band present as a single band at 232 bp (isoleucine/isoleucine).
2. Polymorphism is considered if the band present as either two band at 232 and 263 bp (isoleucine/valine) or single band at 263 bp (valine/valine).

4.6 Research Ethics

This study has been approved by the Ethics Committee, Faculty of Dentistry of the University of Malaya [MEC approval number: DF OP413/42/(P)] and Komisi Etik Riset Biomedik pada Manusia/Hewan (ethics committee) of Trisakti Univeristy [approval number 07/KE/FKG/07/2004], Appendix I.

4.7. Data Analysis

Data obtained from the questionnaires were entered into SPSS version 12. A set of new variables Total Risk Factor (TRF), Total Dietary Habit (TDH), Total Food Preparation (TFP) and Total FFQ (TFFQ) was created by adding total scores for each subject under each construct in order to facilitate further analysis. The new variables were not tested for normality of distribution of study population, due to nominal scale entered in each scales. In order to ensure that the employed scales measure consistently what they are intended to measure, the cronbach alpha coefficient was computed to check reliability. The value of normality and reliability test are illustrated in Table 5.3.

Odds ratio (OR's) for oral cancer was estimated to measure the relationship of risk (OR) of disease due to each study factor by using conditional logistic regression. The univariate model was undertaken using conditional logistic regression to the confounding effect of each variable. Multivariate analysis model (adjusted odds ratio) was performed for variables which have $p > 0.250$ (Schezelman, 1994). A variable was considered to be a confounder if its inclusion in a model changed the OR estimation by more than 10%. The variables that fulfill these criteria as confounders are included in the models presented with OR estimated by method of maximum likelihood. The 95%

confidence intervals (95%CI) were based on the standard error of coefficient estimated. Observation results are shown in 2 x 2 tables for those with oral cancer and control patients as well as with and/or without exposure. Pearson Chi-square test was used for the comparison of proportions, Fisher's exact test for the distribution of the genotypes and a multiple logistic regression model was calculated and conditional logistic regression analysis used to obtain odd ratios (ORs) and their 95% confidence intervals (CIs). The association between the genotype frequencies of GSTM1, GSTT1 and CYP1A1 of cases and control group were assessed using OR and confidence interval (95% CI). The GSTM1, GSTT1 and CYP1A1 distribution were compared between groups using χ^2 test. The statistical computation was carried out by using the STATA 8 version software and SPSS for windows 12 version software.

Factor analysis was used to identify dietary factors of combinations of foods consumed by the subjects. Factor analysis is a generic name given to multivariate statistical analysis that is applied to the identification of factors in a set of performed measurements. Such factors would correspond to indicators (Pallant, 2005; Heppner and Heppner, 2004). In this method, all variables are considered simultaneously, each one in relation to the other.

Principal component analysis was used for the extraction of the factors. This method studies the spatial distribution of the items so as to identify groupings and the relationships between them. The first factor extracted accounts for the maximum possible variance in the data set. The second component, independent from the first, explains as much of the remaining variance as possible, and so on, without any correlation between the components. When determining the number

of factors to retain, it is common practice to consider all factors with Kaiser's criterion or eigenvalues greater than 1, which indicates that the factor at hand describes more of the variability in the data than does an original variable for the factor individually (Pallant, 2005; Heppner and Heppner, 2004). However, it was decided to select only factors lying above the inflection point on the curve, since these correspond to the factors with greater joint variance.

Factor loadings measure the correlation between derived factors and the original measures. These were analyzed after orthogonal rotation using the varimax method (Pallant, 2005). That is, each factor is independent of the others, maintaining axes at 90°. This operation allows for simpler structure, by means of the distribution of the explained variance among the individual components, thus increasing the number of higher and lower factors. Identifying significant factor loadings based on sample size was done by BMDP statistical software (1993). The factor loading 0.40 for 200-249 sample size was considered as significantly contributing to the factor. Within a factor, negative loadings indicate that the food group is inversely associated with the factor, while positive loadings indicate direct association. The higher the factor loading of a food group, the greater the contribution of that group to the factor, since the square of the factor loading corresponds to the percentage variance of the food group that is explained by the factor. Labels were assigned to each factor on the basis of an approximate description of the food items most highly represented. Factor scores for each component retained were calculated for cases and controls. Subjects were assigned scores to indicate the degree to which their diet adhered to each factor retained. Factor scores were categorized into tertile based on the distribution of the entire population and the risk estimates only at highest tertile

(Franceschi et al, 1999). Factor analysis and subgrouping the principle component into tertiles (by means visual bander) were performed using SPSS 12 version software and STATA version 8 for the multivariate and OR (95%CI).

To determine the associations between dietary pattern from factor analysis model and oral cancer, chi square was done. For multivariate analysis, control variables with p values less than or equal to 0.25 in univariate analysis were selected. Furthermore odds ratios and 95% confidence intervals were calculated using conditional logistic regression models to obtain the risk. The variables were then introduced into the regression model (multivariate) using a stepwise and enter procedure. Variables that remained significant after adjustment for the other variables were kept in the model.