DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ASSESSMENT OF INTERACTION AND CONFOUNDING EFFECTS IN LOGISTIC REGRESSION MODEL: AN APPLICATION IN A CASE-CONTROL STUDY OF STOMACH CANCER

by
Özgül VUPA

September, 2009
İZMİR

ASSESSMENT OF INTERACTION AND CONFOUNDING EFFECTS IN LOGISTIC REGRESSION MODEL: AN APPLICATION IN A CASE-CONTROL STUDY OF STOMACH CANCER

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics Program

by
Özgül VUPA

September, 2009
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"ASSESSMENT OF INTERACTION AND CONFOUNDING EFFECTS IN LOGISTIC REGRESSION MODEL: AN APPLICATION IN A CASE-CONTROL STUDY OF STOMACH CANCER"** completed by **ÖZGÜL VUPA** under supervision of **PROF. DR. GÜL ERGÖR** and we certify that in ur opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Gül ERGÖR

Supervisor

Asc. Prof. Dr. C. Cengiz ÇELİKOĞLU

Thesis Committee Member

Asc. Prof. Dr. Ali Kemal ŞEHİRLİOĞLU

Thesis Committee Member

Prof. Dr. Serdar KURT

Examining Committee Member

Prof. Dr. Ergun KARAAĞAOĞLU

Examining Committee Member

Prof. Dr. Cahit HELVACI
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Prof. Dr. Gül ERGÖR for her guidance and helping me to successfully complete throughout this dissertation. She always give me interest, enthusiasm, tenacity, applicable criticism and encouragement throughout this dissertation.

I would like to thank my dissertation committee member, Assoc. Prof. Dr. C. Cengiz ÇELİKOĞLU who made many valuable suggestions and gave constructive advice.

I would like to thank my other dissertation committee member, Assoc. Prof. Dr. Ali Kemal ŞEHİRLİOĞLU who spent part of their time and made many valuable suggestions.

I would like to express deeply felt thanks to Prof. Dr. Serdar KURT for his supports, helpful suggestions, important advice and constant encouragement during my academic life.

I would like to thank Prof. Dr. Seymen BORA, Assistant Prof. Dr. Elçin BORA and Assoc. Prof. Dr. Ayfer ÜLGENALP for helping and guiding me during collection my thesis data and blood samples and also detecting the presence or absence of the genotypes.

I would like to thank my close friend Research Ass. Özlem GÜRÜNLÜ ALMA for her continual encouragement and support throughout this dissertation. Also, thank you all of department's staff for supporting me all time.

I would like to express my deepest gratitute to my family for their encouragement and support during my dissertation.

Özgül VUPA

ASSESSMENT OF INTERACTION AND CONFOUNDING EFFECTS IN LOGISTIC REGRESSION MODEL: AN APPLICATION IN A CASE-CONTROL STUDY OF STOMACH CANCER

ABSTRACT

Stomach cancer (SC) is a major cause of cancer death worldwide. Stomach cancer is the second most common cancer in men and third in women in Turkey. Glutathione S-transferases (GSTs) appear to play a critical role in the protection from the effects of carcinogens. The contribution of GSTM1 and GSTT1 genotypes to susceptibility to the risk of SC and their interaction with cigarette smoking are still unclear in Turkish population. The aim of this study was to determine whether there was any association between genetic polymorphisms of GSTM1 and GSTT1 and SC as well as any interaction between polymorphisms and smoking.

The case-control study was carried out in İzmir, Turkey. The data were collected by questionnaire from 127 SC cases and 101 healthy controls. The relationships between SC and determined risk factors were assessed using ORs and 95 percent CIs derived from univariate, stratified and multivariate analyses.

The finding of the study showed that the prevalences of GSTM1 and GSTT1 null genotypes were 58.2 percent and 22.8 percent in cases, 46.5 percent and 22.2 percent in controls, respectively. In stratified analysis, we found that gender and age were confounder. There were no interactions in all multivariate analysis. This study revealed that GSTM1 polymorphism in SC has a potential role for interaction between this polymorphism and smoking. Our data suggested an increased risk for GSTM1 genotype although a significant association was not found. There was no association for GSTT1 genotype in cases and controls.

Keywords: Confounding, Interaction, GSTM1 and GSTT1 Genotypes, Stomach Cancer.

LOJİSTİK REGRESYON MODELİNDE ETKİLEŞİM VE KARIŞTIRICI ETKİLERİNİN DEĞERLENDİRİLMESİ: MİDE KANSERİ ÜZERİNE BİR OLGU-KONTROL ÇALIŞMASINDAKİ UYGULAMASI

ÖZ

Mide kanseri tüm dünyada kanser ölümlerinin başlıca nedenidir. Türkiye'de mide kanseri vakaları, erkeklerde ikinci sırada iken kadınlarda üçüncü sırada yer alır. Glutathione S-transferases (GSTs) enzimleri, kanserojenlerin etkilerinden korunmada önemli bir rol oynar. GSTM1 ve GSTT1 genlerinin sigara içme ile etkileşimlerinin mide kanseri riskine katkısı Türk populasyonunda hala tam olarak bilinmemektedir. Bu çalışmanın amacı, GSTM1 ve GSTT1 genlerine ait bozulumların sigara içme ile arasındaki etkileşimleri de göz önüne alındığı durumda bu bozulmalar ile mide kanseri arasındaki ilişkinin var olup olmadığını belirlemektir.

İzmir ilinde gerçekleştirilen bu olgu kontrol çalışmasına ait veri seti, 127 mide kanserli hastadan ve 101 sağlıklı kontrolden anket çalışması ile toplanmıştır. Belirlenmiş risk faktörleri ve mide kanseri arasındaki ilişkiler, tek değişkenli, tabakalandırma ve çok değişkenli analizlerden elde edilen odds oran değerleri ve bu oranların yüzde 95 güven aralıkları bulunarak incelenmiştir.

Bu çalışmada GSTM1 ve GSTT1 bozuk genlerinin prevelansı sırasıyla olgularda yüzde 58,2 ve 22,8, kontrollerde ise yüzde 46,5 ve 22,2 olarak bulunmuştur. Tabakalandırma analizlerinde yaş ve cinsiyet karıştırıcı etki olarak bulunmuştur. Çok değişkenli analizlerde ise etkileşim bulunmamıştır. Mide kanseri olmada GSTM1 bozulumu ile sigara arasındaki etkileşimin potansiyel bir rolde olduğu bulunmuştur. Olgu ve kontrollerde GSTM1 geni anlamlı olarak bulunmamasına rağmen, bu veri setinde GSTM1 geninin mide kanseri için artan bir risk faktörü olduğu bulunmuştur. Diğer yandan olgu ve kontrollerde GSTT1 geni için bir ilişki bulunmamıştır.

Anahtar Kelimeler: Karıştırıcı Etki, Etkileşim, GSTM1 and GSTT1 Genleri, Mide Kanseri.

CONTENTS

Page

CHAPTER ONE
INTRODUCTION

The logistic regression analysis is one of the statistical techniques and it is used in predictive probability modeling. This logistic regression model is a member of a general class of models called log-linear models. This model is used for categorical dependent (response, outcome) variables. It describes the relationship between the categorical dependent variable and any types of independent (explanatory, exposure) variables. This model is particularly useful when studying contingency tables. For this reason this model is used in many different sciences. Logistic regression model is used extensively and successfully in the medical sciences to describe the probability or risk of developing a condition (Le, 2003) and it is used in the social sciences (Jaccard, 2001).

In recent times, logistic regression model is used in epidemiologic studies connected with gene-environment association. Of course, there are some reasons for these associations in the epidemiologic studies. These are bias, confounding and interaction effects. An essential aim of the design and analysis phases of any study is to prevent, reduce and assess bias and confounding effect (Jepsen et al., 2004). On the other hand, interaction effect can not be prevented, but it can be controlled with statistical methods. Besides, the interaction and confounding effects are used for model building in the statistical models.

In the statistical models, the interaction effect is said to exist when the effect of an independent variable on a dependent variable differs depending on the value of a third variable. This third variable is commonly called a “moderator variable” or “risk factor”. But confounding exists if meaningfully different interpretations of the relationship of interest result when a third variable is ignored or included in the data analysis. Interaction effect is firstly investigated before confounding effect.

Interaction and confounding effects can be investigated in gene-environment relation. In this thesis, Glutathione S-transferases (GSTs) genotypes (enzymes) will be investigated with interaction and confounding effects in gene-environment relation. GSTs genotypes are involved in the detoxification of many potential carcinogens. The contribution of GSTM1 and GSTT1 genotypes to susceptibility to the risk of stomach cancer and their interaction with cigarette smoking are not clear in many ethnic groups. The aim of this thesis is to determine whether there is any relationship that can be defined as interaction and confounding effects between genetic polymorphism of GSTM1 and GSTT1 genotypes and risk factor as smoking status in stomach cancer.

This thesis contains six chapters. In Chapter 1, whole study is summarized shortly. In Chapter 2, literature reviews about logistic regression model, interaction and confounding effects, risk factors and stomach cancer patients with(out) GSTs genotypes are summarized. In Chapter 3, basic features of a logistic regression model are described. In Chapter 4, interaction and confounding effects in the multiple logistic regression model are examined. In Chapter 5, investigation of interaction and confounding effects in stomach cancer patients with(out) GSTs genotypes are examined with statistical methods (Univariate Analysis (χ^2 test), Stratified Analysis (Breslow Day test, Crude Odds Ratio, Stratified Odds Ratio, Mantel&Haenszel Odds Ratio), Multivariate Analysis (Logistic Regression). In this chapter, applications and results about the study are given. In last chapter, the conclusion of the study is discussed.

CHAPTER TWO
LITERATURE REVIEWS

2.1 Literature Review of Logistic Regression Model

The general informations that are the interpretation of coefficients, model building strategies, some diagnostic measures of the multiple logistic regression models were investigated by Hosmer & Lemeshow (2000). In addition, some main titles that are binomial distribution, Hosmer–Lemeshow test, likelihood, likelihood ratio test, logit function, maximum likelihood estimation, odds, odds ratio, predicted probability, Wald test were investigated by Bewick, Cheek & Ball (2005) in logistic regression model. Rousseeuw & Christmann (2003) studied about outliers in logistic regression model. Also logistic regression model has been used extensively and successfully in medical sciences to describe the probability or risk of developing a condition that can be disease over a specified time period as a function of certain risk factors (Le, 2003). In addition, the logistic regression model has been used in the social sciences (Jaccard, 2001; Pampel, 2000). Nowadays, logistic regression model is used in the gene-environment relation. For example, the interaction effects between some null genotypes as GSTM1 and GSTT1 and risk factors as smoking, alcohol drinking, nutritional and medical factors in stomach cancer were investigated with logistic regression model by Setiawan et al. (2000), Gao et al. (2002), Boccia et al. (2007).

2.2 Literature Review of Interaction and Confounding Effects

Interaction effect is used by social, medical and scientific scientists. The most popular scientists about interaction effects in literature are as follows: Fisher (1926), Rothman et al. (1980; 1998), Kopman (1981), Greenland (1983; 1993), Smith & Day (1984), Kleinbaum et al. (1988), Thompson (1991), Kleinbaum (1994), Assmann et al. (1996), Figueiras et al. (1998), Jaccard (2001), Skrondal (2003), Preacher (2004),

Rodriguez & Llorca (2004), Royston & Saurbrei (2004), Jepsen et al. (2004), Ahlbom & Alfredsson (2005), Kalilani & Atashili (2006), respectively.

Confounding effect is commonly used by medical scientists. The most popular scientists about confounding effects in literature are as follows: Miettinen & Cook (1981), Boivin & Wacholder (1985), Greenland & Robins (1985), Grayson (1987), Solis (1998), McNamee (2003; 2005), Jepsen et al. (2004), Rodriguez & Llorca (2004), Ylöstalo & Knuuttila (2006), Bhopal (2007), Dorak (2007), Schneider (2007), respectively.

Nowadays, interaction effect between GSTM1 and GSTT1 genotypes and smoking risk factor was investigated by Setiawan et al. (2000), Gao et al. (2002), Tamer et al. (2005) and Schneider et al. (2006). Confounding effect with risk factors as gender and age was investigated by Chow et al. (1997). Interaction and confounding effects between GSTM1 and GSTT1 genotypes and possible risk factors (age, gender, smoking (pack year), education, alcohol drinking, salt intake, fruit intake, BMI) was investigated by Setiawan et al. (2000).

2.3 Literature Review of GST's Genotypes and Stomach Cancer

According to WHO, stomach cancer is a major cause of cancer death worldwide. It is very common in certain Asian, Central European, Central and South American countries. Each year there are 59,300 cases in the USA, 2,800 in Canada, 2,000 in Australia and 9,100 in the UK. 50 years ago stomach cancer was the most common type of cancer. Now it is number 5 or 6 in most western countries. For example, stomach cancer is now the 7th common cancer among adults in the UK. Generally, out of every 100 cancers diagnosed, 3 are cancer of the stomach. Worldwide, there are nearly 800,000 cases each year.

According to the prevalence or incidence, Italy has a high prevalence of stomach cancer, affecting about 50% of the "normal" population. It also has a moderately high incidence of stomach cancer, in the range of 30 cases per 100,000 persons per year. In comparison, USA incidence is less than 10 (the world's lowest) cases per 100,000 persons per year and Japan's rate is about 60 (competing with Korea as the world's highest) cases per 100,000 persons per year. San Marino is known for quite a high incidence of stomach cancer 50-100 cases per 100,000 persons per year. Korea and Japan have the highest rates, ten times the rate in the USA.

Stomach cancer depends on many factors. These are gender, age, diet status, body mass index, smoking, family history, intake of food, intake of alcohol, environmental exposure etc. Some of these risk factors were investigated by Hirayama (1984), Jedrychowski et al. (1986), Hu et al. (1988), Dyke et al. (1992), Nazario et al. (1993), Hansson et al. (1994), Lee et al. (1995), Tredaniel et al. (1997), Terry et al. (1998), Setiawan et al. (2000) and Yalçın et al. (2006).

GSTs genotypes are involved in the detoxification of many potential carcinogens. Several GST gene families have been identified: alpha, mu, theta and pi. GSTM1 and GSTT1 are major members of the GST family. The null genotypes of GSTM1 and GSTT1 genes may be associated with an increased risk of stomach cancer. These genes are absent in 10%-60% of different ethnic populations (35-60% for GSTM1, 10-60% for GSTT1). Prevalences of these GSTs genotypes in literature are given as follows (Ca: cancer, Co: control, -: null genotype):

Table 2.1 Prevalences of GSTs genotypes in different ethnic populations

	Ca group		Co group	
Population	GSTM1 -	GSTT1 -	GSTM1 -	GSTT1 -
English	52.9%		54.8%	
China	48.0%	54.0%	50.0%	38.0%
Japan	-	54.0%	-	45.0%
Italian	56.0%	37.0%	53.0%	22.0%

Few studies have correlated environmental factors and genetic susceptibility with the risk of stomach cancer, especially in the Chinese population, which has one of the highest incidences of stomach cancer in the world. There is no information between environmental factors and genetic susceptibility with the risk of stomach cancer in Turkey for interaction and confounding effects. Both GSTM1 and GSTT1 genotypes can be catalyze the detoxification of compounds in cigarette smoke. In this study, we aimed to evaluate the association between GSTM1 and GSTT1 genotypes and the risk factor as smoking in stomach cancer. In addition, we aimed to find possible interaction and confounding effects between GSTM1 and GSTT1 genotypes with smoking risk factor.

CHAPTER THREE
GENERAL INFORMATION ABOUT LOGISTIC REGRESSION MODEL

Logistic regression (LogR) is used when the dependent variable (Y_i) is nominal or ordinal scale and the independent variables (X_i) are of any type of scale (i = 1, 2, …, p, p is the number of independent variables). Logistic regression is popular to overcome many of the restrictive assumptions of ordinary least square (OLS) regression. These assumptions are ordered as follows:

* LogR does not assume a linear relationship between the dependent and the independent variable(s).
* The dependent variable need not be normally distributed.
* The dependent variable need not be homoscedastic for each level of the independents. It means that there is no homogeneity of variance.
* Normally distributed error terms are not assumed.
* LogR does not require that the independents be interval scale.
* LogR does not require that the independents are unbounded.

3.1 Regression Model with Binary Dependent Variable

The dependent variable of interest is not on a continuous scale and it may have only two possible outcomes and therefore it can be represented by a binary indicator variable taking on values 0 and 1. This dependent variable is measured on a binary scale. For example, the dependent variable may be alive or dead, present or absent, cancer group or control group.

The simple linear regression model is written as: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $i = 1, 2, \ldots, n$. Where Y is the dependent variable, X is the independent variable, β_0 is a constant term and β_1 is a slope coefficient. The expected value of dependent variable, $E\{Y_i\}$, has a special meaning in this case. Since $E\{\varepsilon_i\} = 0$, it is written as: $E\{Y_i\} = \beta_0 + \beta_1 X_i$. When Y_i is a Bernoulli random variable, there are two

probabilities $(Y_i = 0,1)$. π_i is the probability that Y_i=1 and $(1-\pi_i)$ is the probability that Y_i=0. The expected value of a Bernoulli random variable is $E\{Y_i\} = \pi_i$. So, $E\{Y_i\}$ is written as $E\{Y_i\} = \beta_0 + \beta_1 X_i = \pi_i$. In addition, the variance of a Bernoulli random variable, $V(Y_i)$, for the simple linear regression model is $V(Y_i) = E(Y_i - E(Y_i))^2 = \pi_i(1-\pi_i)$.

3.2 Special Problems When Dependent Variable is Binary

According to the linear regression model, the error terms are assumed to have a normal distribution with a constant variance for all levels of X_i. However, when the dependent variable is 0 or 1 binary indicator variable, error terms are not only distributed normal but also they don't have constant variance. The error term $\varepsilon_i = Y_i - (\beta_0 + \beta_1 X_i)$ can take on only two values. If Y_i=1, then the error term takes the value as $\varepsilon_i = 1 - \pi(x_i) = 1 - \beta_0 - \beta_1 X_i$ with the probability $\pi(x_i)$. If Y_i=0, then the error term takes the value as $\varepsilon_i = -\pi(x_i) = -\beta_0 - \beta_1 X_i$ with probability $1 - \pi(x_i)$. Thus, the assumption of normality does not hold for this model. It is not appropriate (Neter et al., 1996). Another problem with the error terms (ε_i) is that they do not have equal variances. The variance of Y_i, $V(Y_i)$, for the simple linear regression model is $\pi_i(1-\pi_i)$. Also, the variance of the error terms (ε_i) is the same as that of Y_i, because ε_i is equal to $(Y_i - \pi_i)$ and π_i is a constant. The last problem is related with constraints on dependent (response) function. Since the response function represents probabilities, the mean responses should be constrained as follows: $0 \le E(Y_i) = \pi_i \le 1$

3.3 Logistic Response Function

The conditional mean, $\pi(x_i)$, is shown as:

$$\pi(x_i) = E(Y|x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \qquad (3.1)$$

This specific form is called logistic response function. A transformation of $\pi(x_i)$ is the logit transformation. This transformation is expressed as follows:

$$g(x_i) = \ln\left[\frac{\pi(x_i)}{1-\pi(x_i)}\right] = \ln(e^{\beta_0+\beta_1 x_i}) = \beta_0 + \beta_1 x_i \tag{3.2}$$

The importance of this transformation is that $g(x_i)$ has many of the desirable properties of a linear regression model. The logit transformation is linear in its parameters and it may be continuous. In addition, the logit may have range from $-\infty$ to ∞, depending on the range of x_i (Hosmer & Lemeshow, 2000).

3.4 Fitting of Multiple Logistic Regression Model

Multiple logistic regression model for the case of more than one independent variable is fitted. In this setting, the vector $\tilde{x} = (x_1, x_2, \ldots, x_p)$ represents the collection of p independent variables for this model. The equations for the probability and the logit transformation can be expressed as follows:

$$\pi(\tilde{x}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)} = \frac{\exp(g(\tilde{x}))}{1 + \exp(g(\tilde{x}))} \tag{3.3}$$

$$g(\tilde{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \tag{3.4}$$

There is a sample of n independent observations and it is expressed as $(\tilde{x}_i, y_i)$. Where y_i denotes the value of a dichotomous response variable and $\tilde{x}_i$ is the value of the independent variables for the ith subject. The estimates of these parameters are shown as: $\tilde{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$.

The general methods of estimation in simple or multiple logistic regression model are investigated in three main concepts. These are the Maximum Likelihood Method, Iteratively Reweighted Least Squares Method and the Minimum Logit Chi-Square Method.

3.4.1 Likelihood Function

Likelihood function express the probability of the observed data as a function of the unknown parameters. For pairs $(\tilde{x}_i, y_i)$, since $y_i = 1$, the contribution to the likelihood function is $\pi(\tilde{x}_i)$. Since $y_i = 0$, the contribution to the likelihood function is $1 - \pi(\tilde{x}_i)$. Since Y_i's have a Bernoulli distribution, the probability density function can be defined as follows:

$$P(Y = y_i) = f_i(y_i) = \pi(\tilde{x}_i)^{y_i} (1 - \pi(\tilde{x}_i))^{1-y_i} \tag{3.5}$$

Where $y_i = 0$ or $y_i = 1$ for $i = 1, 2, \ldots, n$. Since the observations Y_i are assumed to be independent, the likelihood function can be defined as follows:

$$L(\tilde{\beta}) = \prod_{i=1}^{n} \pi(\tilde{x}_i)^{y_i} (1 - \pi(\tilde{x}_i))^{(1-y_i)} \tag{3.6}$$

In order to maximize this function, the derivative must be taken with respect to each of the parameters. Then, the resulting equations would be set equal to zero and solved simultaneously. These equations are called likelihood equations. In this case, there are (p+1) likelihood equations which are obtained by differentiating the log-likelihood function with respect to the (p+1) coefficients. In addition, this process can be simplified by performing the same analysis on the natural log of the likelihood function (Kleinbaum, 1998). Obtaining the likelihood equations are expressed as:

$$\sum_{i=1}^{n} \left[y_i \, \pi(\tilde{x}_i) \right] = 0 \tag{3.7}$$

$$\sum_{i=1}^{n} x_{ij}\left[y_i - \pi(\tilde{x}_i)\right] = 0 \qquad j = 1, 2, \ldots, p \tag{3.8}$$

Likelihood equations are not linear, solving these equations simultaneously requires an iterative procedure that is normally left to a software package. By using these packages (SPSS, NCSS, etc…) programs, maximum likelihood estimates of the parameters are obtained easily.

3.4.2 Maximum Likelihood Estimation Method

The maximum likelihood estimation method (MLE) is used to calculate the logit coefficients. This method yields values for the unknown parameters which maximize the probability of obtaining the observed set of data. In order to apply this method, the likelihood function is constructed. This method uses the logistic function and an assumed distribution of Y to obtain estimates for the coefficients that are most consistent with the sample data.

The sum of the observed values of Y_i is equal to the sum of the expected values. This is shown as:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{\pi}(\tilde{x}_i) \tag{3.9}$$

$\hat{\tilde{\beta}}$ denote the solution of likelihood equations. In other words, $\hat{\tilde{\beta}}$ is the maximum likelihood estimate of $\tilde{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$. $\hat{\pi}(\tilde{x}_i)$ is the maximum likelihood estimate of $\pi(\tilde{x}_i)$ and it estimates the conditional probability that Y_i is equal to 1, given $X = x_i$. In other words, $\hat{\pi}(\tilde{x}_i)$ is the fitted multiple logistic response function for the $\underline{i}$th case and the value of

$$\hat{\pi}(\tilde{x}_i) = \frac{\exp(\hat{g}(\tilde{x}_i))}{1 + \exp(\hat{g}(\tilde{x}_i))} \tag{3.10}$$

is computed using $\hat{\tilde{\beta}}$ and $\tilde{x}_i$.

MLE is a iterative algorithm and this procedure is complex and usually requires numerical search methods. Hence MLE of the logistic regression is done on a computer.

3.4.3 Dummy Variable

If some of the independent variables are discreate, ordinal or nominal scaled variable (categorical variable) with more than two levels, then the model differs from general formula in equation (3.4). For example, education, smoking status, race, sex, regions of Turkey, number of treatment groups etc...can be given. If the number of variable categories is equal to k, then (k-1) dummy variables must be created. For example, one of the independent variables is education and that is coded as "no education", "primary, middle and high school" or "university". Here, two dummy variables are necessary. When the respondent or reference variable is "university", the two dummy variables, D_1 and D_2, would both be set equal to zero; when the respondent is "primary, middle and high school", D_1 would be set equal to 1 while D_2 would still equal 0; when the respondent is "no education", D_2 would be set equal to 1 while D_1 would still equal 0 (Hosmer & Lemeshow, 2000). It is shown in Table 3.1.

Table 3.1 The coding of dummy variables for education

	Dummy Variable	
Education Variable	**D_1**	**D_2**
university	0	0
primary, middle and high school	1	0
no education	0	1

The notation to indicate dummy variables is more different than the logistic regression model. Suppose that the jth independent variable x_j has k_j levels. The $(k_j - 1)$ dummy variables are needed and they are denoted as D_{jm}. In addition, the coefficients for these dummy variables are denoted as β_{jm}, $m = 1, 2, \ldots, (k_j - 1)$. The logit for a model with p independent variables and the jth independent variable being discrete is expressed as:

$$g(\tilde{x}) = \beta_0 + \beta_1 x_1 + \ldots + \sum_{m=1}^{k_j - 1} \beta_{jm} D_{jm} + \beta_p x_p \qquad (3.11)$$

3.5 Testing for the Significance of the Coefficients

After estimating the coefficients, an assessment of significance of the variable in the fitted model is concerned. This involves formulation and testing of statistical hypothesis to determine whether the independent variable in the model is significantly related to the response variable (Hosmer & Lemeshow, 2000). The approach in testing for the significance of the coefficient of a variable in the model is related with the following question "Does the model which includes the variable in question tell us more information about the response variable than does a model which does not include that variable?". This question is answered by comparing the observed values of the response variable to those predicted by each of two models. If the predicted values with the variable in the model are better or clearer, than when the variable is not in the model, then the variable in question is said to be significant. The comparison is based on the log-likelihood. In addition, it is not important question of whether the predicted values that are obtained from saturated model have accurate relation or representation of the observed values of response variable in an absolute sense or not. This is concerned in goodness of fit. In logistic regression model, there are three commonly used tests for hypothesis testing. These are Likelihood Ratio Test, Wald Test and Score Test.

3.5.1 Likelihood Ratio Test

Comparison of observed to predicted values is based on the log-likelihood function in logistic regression. The model which includes all possible terms (including interactions) is called as saturated model. In addition, a saturated model is one that contains as many parameters as there are data points. The current model is the subset of the saturated model. The current model does not include the variable investigated by the researcher. The likelihood ratio test statistic is (–2) times of the difference between the log likelihoods of saturated and current model. The distribution of the likelihood ratio test statistic is closely approximated by the chi-square distribution for large sample sizes. The degress of freedom of the approximating chi-square distribution is equal to the difference in the number of regression coefficients in the two models (NCSS, 2004).

The comparison of observed to predicted values is based on the log likelihood function. The log likelihood equation takes the form as follows:

$$\ln L(\beta_0, \beta_1, \ldots, \beta_p) = \ln L(\tilde{\beta})$$

$$= \sum_{i=1}^{n} \left\{ y_i (\beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}) - \ln(1 + \exp(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi})) \right\} \qquad (3.12)$$

To better understand this comparison, it is helpful conceptually to think that an observed value of the response variable as also being a predicted value resulting from a saturaed model. A saturated model is one that contains as many parameters as there are data points. This comparison is obtained as follows:

$$D = -2 \ln \left[\frac{\text{likelihood of the current model}}{\text{likelihood of the saturated model}} \right] \qquad (3.13)$$

This expression is called the deviance (D). The deviance for logistic regression model plays the same role as sum of squares error (SSE) in linear regression. Using minus twice its log is necessary to obtain a quantity whose distribution is known.

Also, this procedure can be used for hypothesis testing purposes. This test is called Likelihood Ratio Test. In order to determine whether the parameter is significant to the model or not, the deviance of the model containing the independent variable is compared with the deviance of the model without the independent variable. This change in D is called G statistic. This statistic in logistic regression plays the same role as the numerator of the partial F test in linear regression. The test statistic is expressed as follows:

$$G = D(\text{for the model without the variable(s)}) - D(\text{for the model with the variable(s)})$$

$$G = -2\ln\left[\frac{\text{likelihood without the variable(s)}}{\text{likelihood with the variable(s)}}\right] \tag{3.14}$$

In checking the significance of the model, the following null and alternative hypotheses are written as follows:

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$$
$$H_1 : \text{At least one of the } \beta_p \neq 0 \tag{3.15}$$

The statistic G has a chi-square distribution with $(\nu_2 - \nu_1)$ degrees of freedom (df). Here, ν_2 equals to the number of variables in the saturated model plus 1 and ν_1 equals to the number of variables in the current model plus 1. For this test, the decision rule requires that p-value is $P\left\{\chi^2_{(1-\alpha,df=(\nu_2-\nu_1))} > G\right\}$. If this p-value is less than α-value, H_0 is rejected. This means that the model would be deemed significant. Here, any or all of the coefficients are nonzero. α-value is usually accepted as 0.05. For this reason, p-value is compared with $\alpha = 0.05$ level. On the other hand, if p-value is greater than α-value, then the current model is as good as the saturated model and the null hypothesis (H_0) is failed to reject. In addition, if the statistic G is greater than $\chi^2_{(1-\alpha,df=(\nu_2-\nu_1))}$, then H_0 is rejected. The model is accepted as significant.

3.5.2 Wald Test

After testing the significance of the model, at least one or perhaps all p coefficients can be different from zero. The Wald test statistics are used to see which variables are significant. These statistics have the standard normal distribution and they are evaluated as follows:

$$W_j = \frac{\hat{\beta}_j}{S\hat{E}(\hat{\beta}_j)} \sim Z(\alpha/2) \tag{3.16}$$

Under the hypothesis that $\beta_j = 0$, two tailed $p-value$ is evaluated by $P(|Z| > W)$. Standard error of $\hat{\beta}_j$ is provided by the square root of the corresponding diagonal element of the covariance matrix $V(\hat{\beta}_j)$. Where, Z denotes a random variable following the standard normal distribution. If this p-value is less than given α-value, then the null hypothesis is rejected. For this test, p-value can be defined by $\text{p-value} = 2P(Z > \text{the observed test statistic})$.

For multivariate case, Wald test is used in statistical package programs. This W value is then squared, yielding a Wald statistic with a chi-square distribution. However, several authors have identified problems with use of the Wald statistics. Menard warns that for large coefficients, standard error is inflated, lowering the Wald statistic (chi-square) value (Menard, 1995). Agresti states that the likelihood ratio test is more reliable for large sample sizes than the Wald test. (Agresti, 2002) The Wald test is obtained from the following vector-matrix calculation.

$$W = \hat{\tilde{\beta}}' \left[\sum \hat{\tilde{\beta}} \right]^{-1} \hat{\tilde{\beta}} \tag{3.17}$$

W has a chi-square distribution with (p+1) degrees of freedom under the hypothesis that each of the (p+1) coefficients are equal to zero. A similar situation

can be done with excluding $\hat{\beta}_0$ from the analysis, then W will be distributed as chi-square with p degrees of freedom.

3.5.3 Score Test

Score test is based on the conditional distribution of the p derivatives of $L(\tilde{\beta})$ with respect to $\tilde{\beta}$. The computation of the score test is as complicated as the Wald test.

3.6 Interpretation of the Coefficients

The estimated coefficients for the independent variables give the slope or rate of change of a function of the dependent variable per unit of change in the independent variable. The function of the dependent variable yields a linear function of the independent variables. This is called a link function. In linear regression model, it is the identity function. In logistic regression model, the link function is the logit.

In linear regression model, the slope coefficient, β_1, is equal to the difference between the value of the dependent variable at $(x+1)$ and the dependent variable at x. It is expressed as follows:

$$\beta_1 = \left| y(x = x+1) - y(x = x) \right| \tag{3.18}$$

In logistic regression, model it is expressed as follows:

$$g(x+1) = \ln\left\{\pi(x+1)/\left(1-\pi(x+1)\right)\right\} = \beta_0 + \beta_1(x+1) = \beta_0 + \beta_1 x + \beta_1 \tag{3.19}$$

Here, the logit difference is equal to β_1 and it is evaluated as follows:

$$g(x+1) - g(x) = g(x+1) - (\beta_0 - \beta_1(x)) = \beta_1 \tag{3.20}$$

3.6.1 Dichotomous Independent Variable

In this case, independent variable (x) can take only two values and it is coded as 0, 1. In logistic regression model, there are two values of $\pi(x)$ and two values of $1-\pi(x)$. The odds of the outcome being present among individuals with $x=1$ and $x=0$ are expressed respectively.

$$\frac{P(y=1|x=1)}{P(y=0|x=1)}=\frac{\pi(1)}{1-\pi(1)} \qquad \frac{P(y=1|x=0)}{P(y=0|x=0)}=\frac{\pi(0)}{1-\pi(0)} \tag{3.21}$$

The logit is defined to be the logarithm (natural exponential) of the odds. They are defined by $g(1)$ and $g(0)$ for dichotomous independent variable. The "odds ratio = OR" is defined as the ratio of the odds for $x=1$ to the odds for $x=0$ and it is expressed as follows:

$$OR=\frac{\pi(1)\big/(1-\pi(1))}{\pi(0)\big/(1-\pi(0))} \tag{3.22}$$

$$OR=\exp(\beta_1) \tag{3.23}$$

The log of OR is called logit difference (log odds ratio) and it is expressed as: $\ln(OR)=\ln[\pi(1)/(1-\pi(1))]-\ln[\pi(0)/(1-\pi(0))]=g(1)-g(0)=\beta_1$. OR can take any value between 0 and ∞. OR gives us the effect of a one-unit change in X on the probability that $Y=1$. If OR equals 1, the effect is estimated to equal 0. If OR is greater than 1, for example $\hat{OR}$ equals 1.8, a one-unit increase in X raises the probability of $Y=1$ by 0.8, or 80%. On the other hand, If OR is less than 1, for example $\hat{OR}$ equals 0.2, the effect of X on Y is negative: a one-unit increase in X leads to a 80% reduction in the probability of $Y=1$.

The variance is evaluated by $V(\hat{\beta}_1) = \left[(1/a) + (1/b) + (1/c) + (1/d)\right]$. Where a, b, c, d are cell frequencies in the 2×2 table of $Y \times X$. The distribution of the estimate of OR tends to be skewed to the right. Thus, confidence interval is usually based on $\hat{\beta}_1$ which is closer to being normally distributed. $\hat{\beta}_1 \sim N(\beta_1, V(\hat{\beta}_1))$ The confidence interval for the odds ratio is $\exp\left\{\hat{\beta}_1 \pm Z_{1-\alpha/2}\ SE(\hat{\beta}_1)\right\}$.

3.6.2 Polytomous Independent Variable

In this case, if the independent variable takes three or more levels, then, it is called polytomous independent variable. For example, nominal scale variable X is coded at 4 levels. Thus, (4-1)=3 dummy variables are created.

3.6.3 Continuous Independent Variable

In this case, when there is an independent continuous variable in the model, the unit of this variable should be defined. Most often the value of "1" is not biologically very interesting. For example, increased risk for 1 additional year of age or mmHg in systolic blood pressure or mg/100 ml of cholesterol are not very interesting. But, A change of 10 years or 5 mmHg or 25 mg/100 ml may be more meaningful. The log odds ratio for a change of c units in X, OR and variance of the variable are expressed respectively as follows:

$$x = g(x + c) - g(x) = c\beta_1 \qquad OR(x + c, x) = e^{c\,\beta_1}$$
$$V\left\{\ln(\hat{OR}(x + c, x))\right\} = c^2 V(\hat{\beta}_1) \tag{3.24}$$

100% confidence interval is evaluated as:

$$\exp(c\,\hat{\beta}_1 - Z_{1-\alpha/2}\ c\ SE(\hat{\beta}_1)) \le OR \le \exp(c\,\hat{\beta}_1 + Z_{1-\alpha/2}\ c\ SE\ (\hat{\beta}_1)) \tag{3.25}$$

3.7 Model Building Procedures in Logistic Regression Model

If there are more variables included in the model, then standard errors of estimates become greater. While there are many independent variables in the model, model building and devoloping include more complex situations. For this reason, to select less variables is very important. There are different ways used for variable selection in logistic regression model. These are the univariate analysis and the multivariate analysis. Multivariate analysis consists on two methods. These are stepwise logistic regression methods (Forward Selection, Backward Elimination) and best subset logistic regression method.

The variable selection process begins with univariate analysis of each variable. The variables are selected for the multivariate analysis after fitting the univariate analysis. Any variable whose univariate test has a p-value ≤ 0.20 is considered as candidate for the multivariate model along with all variables of known clinical importance. Otherwise, if any variable's p-value is greater than 0.20, then this variable is excluded from the model. The importance of each variable included in the multivariate logistic regression model should be verified. Variables that do not contribute to the model are eliminated from the model and the new model is constructed. The new model are compared to the old model through the likelihood ratio test (Hosmer & Lemeshow, 2000). Stepwise logistic regression is an extremely popular method for model building. Stepwise procedures assume an initial model and then use rules for adding or delating terms to arrive at a final model (Cristensen, 1997). There are two procedures for model building in the stepwise logistic regression method. The forward selection process adds variables sequentially to the model until further additions do not improve the fit. At each stage, the variable giving the greatest improvement in the fit is selected. The maximum p-value for the final model is a sensible criterion. A stepwise variation of this procedure retests, at each stage, variables added at previous stages to see if they are still needed. The backward elimination process begins with a complex model and sequentially removes variables. At each stage, the variable with least damaging effect on the

model is removed. The process stops when any further deletion leads to a significantly poorer-fitting model.

It is so clear that modeling is a useful process both for prediction of future observables and for describing the relationship between variables. Large models reproduce the data on which they were fitted better than smaller models. The saturated model provides a perfect fit of the data. However, smaller models have more powerful interpretations and are often better predictive tools than large models. Often, the main goal is to find the smallest model that fits the data (Cristensen, 1997).

3.8 Validation in Logistic Regression Model

Regression models are powerful tools frequently used to predict a dependent variable from a set of independent variables. An important problem is whether results of the regression analysis on the sample can be extended to the population the sample has been chosen from. If this happens, then the model is said to be a good fit. This is investigated in the topic “model validation analysis”. Model validation analysis is used in logistic regression model with some statistical tests and methods. After fitting the logistic regression model, it is useful to test its effectiveness by using goodness of fit tests. In addition, it is decided whether the fit of the model is adequate by using goodness of fit tests or not. One of them is deviance test and the other is Hosmer-Lemoshow test. Here, the null hypothesis is that the model of interest fits well. The observed values of the outcome variable in vector form is denoted as y where $y^* = (y_1, y_2, \ldots, y_n)$ and the fitted values of the outcome variable in vector form as $\hat{y}$ where $\hat{y}^* = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$. $(y_i - \hat{y})$ is defined to be residual and its value must be small $(i = 1, 2, \ldots, n)$.

3.8.1 The Hosmer-Lemeshow Test

The aim of the Hosmer-Lemeshow test is to make a group of the values of the estimated probabilities. 10 groups are created $(g = 10)$. The first group contains $n_1^* = n/10$ subjects having the smallest estimated probabilities. The last group contains $n_{10}^* = n/10$ subjects having the largest estimated probabilities. The each group's n_k^* equals to $n/10$ $(k = 1, 2, \ldots, 10)$. For the $y = 1$ row, the estimates of the expected values are found by summing the estimated probabilities over all subjects in a group. For $y = 0$ row, the estimates of the expected values are found by subtracting from 1 (1-the estimated probabilities over all subjects in a group). The Hosmer-Lemeshow goodness of fit statistics is denoted by $\hat{C}$ and it is evaluated as follows:

$$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n_k^* \overline{\pi}_k)^2}{n_k^* \overline{\pi}_k (1 - \overline{\pi}_k)} \tag{3.26}$$

Where n_k^* is the number of covariate patterns in the $k\underline{th}$ group.

$$o_k = \sum_{j=1}^{n_k^*} y_j \tag{3.27}$$

Where o_k is the number of responses among n_k^* covariate patterns. In addition, $\overline{\pi}_k$ is the average estimated probability and it is calculated as

$$\overline{\pi}_k = \sum_{j=1}^{n_k^*} \frac{m_j \hat{\pi}_j}{n_k^*} \tag{3.28}$$

The distribution of the statistic $\hat{C}$ is well approximated by the chi-square distribution with $(g - 2)$ degrees of freedom, , when j is equal to n and the fitted logistic regression model is the correct model. If the value of the Hosmer-Lemeshow goodness of fit statistic computed from "deciles of risk" table is less than the

corresponding p-value computed from the chi-square distribution with 8 degrees of freedom, then the model is accepted to fit quite well.

The Hosmer-Lemeshow goodness of fit statistic is easily interpretable and it can be easily applied to data. It is illustrated as follows:

Table 3.2 Observed and estimated expected frequencies

		Decile of Risk				
Y		1	2	…	10	Total
Y=1	Obs	o_{11}	o_{12}	…	o_{110}	n_1
	Exp	$\overline{\pi}_{11}$	$\overline{\pi}_{12}$	…	$\overline{\pi}_{110}$	
Y=0	Obs	o_{01}	o_{02}	…	o_{010}	n_0
	Exp	$\overline{\pi}_{01}$	$\overline{\pi}_{02}$	…	$\overline{\pi}_{010}$	
Total		n/10	n/10	…	n/10	n

3.9 Multicollinearity in Logistic Regression Model

A set of variables are exactly collinear if one of them is a linear function of the others. In other words, two variables are collinear if they are highly correlated. Multicollinearity in logistic regression models is a result of strong correlations between independent variables. The existence of multicollinearity inflates the variances of the parameter estimates. That may result, particularly for small and moderate sample sizes, in lack of statistical significance of individual independent variables while the overall model may be strongly significant. Multicollinearity may also result in wrong signs and magnitudes of regression coefficient estimates, and consequently in incorrect conclusions about relationships between independent and dependent variables. Multicollinearity can be detected in high correlation coefficient and R^2 values. The problem of multicollinearity can be overcomed by using Ridge Logistic Regression Method.

CHAPTER FOUR
INTERACTION AND CONFOUNDING EFFECTS IN LOGISTIC REGRESSION MODEL

In clinical epidemiology, the two basic components of any study are exposure and outcome. The exposure can be a risk factor or a treatment. The outcome is usually death or disease. Risks, rates, prevalences and odds are common measures of the frequency of an outcome. Comparing them between groups yields relative frequency measures that are relative risks, rate ratios, prevalence ratios and odds ratios. The main study designs in observational studies are cohort, case-control and cross sectional studies. In a cohort study, patients with different levels of exposure are followed forward in time to determine the incidence of the outcome in question in each exposure group. With this design, the investigator can study several outcomes within the same study. The most common frequency measures are relative risks and incidence rate ratios. In a case-control study, the first step is to identify the outcome of interest or the cases. That makes it a good design for studying rare outcomes. Having identified the cases, the investigator selects the controls from the source population. The level of exposure is compared between cases and controls. The relative frequency measure is the odds ratio. The estimate is better if the disease is rare. In a cross sectional study, exposure and outcome are measured simultaneously. Prevalence rates can be compared between groups (Jepsen et al., 2004).

There can be noncausal associations in epidemiologic studies. These are bias, confounding and interaction effects. An essential aim of the design and analysis phases in any study is to prevent, reduce and assess bias and confounding effect. Interaction should be treated differently from confounding. Interaction can not be prevented or reduced, it can be assessed with statistical methods.

In this chapter, interaction and confounding effects in the multiple logistic regression model are examined.

4.1 Definition of Interaction and Confounding Effects

The interaction effect is said to exist when the effect of an exposure variable on an outcome variable differs depending on the value of a risk factor variable. Jaccard (2001), called the exposure variable and the risk factor as “focal variable” and “moderator variable”, respectively. For example a researcher wants to determine whether a clinical treatment for depression is more effective for males than females. It is evident in this case that gender is the moderator variable and the presence versus absence of the treatment is the focal variable (Jaccard, 2001).

Confounding effect differs between the comparison groups and this confounder may affect the outcomes. Confounding exists if meaningfully different interpretations of the relationship of interest result when a risk factor variable is ignored or included in the data analysis. Confounding is known as “mixing of the effect” of the exposure-outcome relationship of interest with that of a third factor that is called “confounder”. Confounding occurs when the exposed and non-exposed groups in the source population are not comparable, because of inherent differences in background outcome. Confounding can also be introduced into a study through selection factors (response bias) or misclassification of exposure or outcome.

As seen in Figure 4.1, smoking is associated with drinking alcohol but it is not the result of drinking alcohol. Smoking is a significant risk factor for lung cancer. Smoking is correlated with alcohol consumption and a risk factor even for those who do not drink alcohol. Alcohol consumption may be correlated with smoking but is not a risk factor in non-smokers. In addition, it can be considered how strongly the confounder is associated with the outcome (Dorak, 2006).

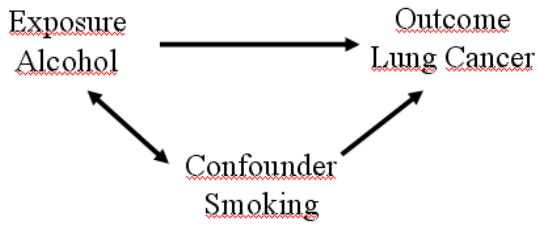

Figure 4.1 The relationship between alcohol, lung cancer and smoking

The interaction and confounding effects are used for model building in the statistical models. In addition, the relationship between categorical outcome variable and any types of exposure variables are measured using the OR and their 95% CIs derived from logistic regression analysis determining for interaction factor and controlling for possible confounding factor using any software (SPSS, NCSS, etc…). Crude and stratified ORs are calculated for exposure variables. Dummy variables are used to estimate the OR for each category of exposure variables in logistic regression analysis.

4.2 Additive and Multiplicative Interaction Effects

Departures from additive and multiplicative interaction effects between exposure and risk factors are evaluated. The null hypotheses of additivity and multiplicativity can be tested easly. A more than additivity interaction is indicated when:

$$OR_{11} > OR_{10} + OR_{01} - 1 \quad (4.1)$$

Where OR_{11} = OR when both factors are present, OR_{10} = OR when only factor 1 is present and OR_{01} = OR when only factor 2 is present. A more than multiplicativity interaction is suggested when:

$$OR_{11} > OR_{10} \times OR_{01} \quad (4.2)$$

The departures from additivity and multiplicativity effects are assessed by including main effect variables and their product terms in logistic regression model.

4.3 Modeling of Interaction Effect in Logistic Regression Model

The most common approach to modeling interactions in logistic regression model is to use product terms. Logistic regression model with two continuous independent variables (X and Z) is

$$\text{logit}(\pi) = \beta_0 + \beta_1 X + \beta_2 Z \tag{4.3}$$

Where Z is a moderator variable and X focal variable. There is an interaction effect such that the effect of X on the outcome variable differs depending on the value of Z. One way of expressing this is to model β_1 as a linear function of Z.

$$\beta_1 = \beta_0' + \beta_3 Z \tag{4.4}$$

According to this formulation, for every 1 unit that Z changes, the value of β_1 is predicted to change by β_3 units. The expression in equation (4.4) for β_1 is substituted in equation (4.3).

$$\text{logit}(\pi) = \beta_0 + (\beta_0' + \beta_3 Z)X + \beta_2 Z \tag{4.5}$$

Equation (4.6) is evaluated with multiplying equation (4.5)

$$\text{logit}(\pi) = \beta_0 + \beta_0' X + \beta_3 XZ + \beta_2 Z \tag{4.6}$$

The interaction model with a product term is obtained after assigning new labels to the coefficients and rearranging term. This model is obtained as

$$\text{logit}(\pi) = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ \tag{4.7}$$

4.4 Testing of Interaction Effect in Logistic Regression Model

Interaction effect is tested using hierarchical logistic regression and the homogeneity test of odds ratios in logistic regression model. This homogeneity test is called Breslow Day test.

4.4.1 Hierarchical Logistic Regression

Kleinbaum (1994) suggests that the interaction effect in logistic regression is found by hierarchically well formulated models. A hierarchically well formulated model is one in which all lower order components of the highest order interaction term are included in the model. For example, if interest is an a two way interaction between two continuous variables (X, Z), then a hierarchically well formulated model includes X, Z and XZ as independent variables. For a categorical independent variables, D_1 and D_2, and continuous variable Z, hierarchically well formulated interaction model includes D_1, D_2, Z, D_1Z and D_2Z (Jaccard, 2001).

Interaction effect with hierarchically well formulated model is tested using hierarchical logistic regression in which one determines whether the product terms significantly improve model fit over and above the case where no product terms are included in the model. This approach involves estimating χ^2 values for each of the (4.3) and (4.7) equations. Equation (4.3) is "no interaction" model and equation (4.7) is "interaction" model. In another words, the interaction between two continuous variables (X, Z) is represented by a single product term as equation (4.3) and it is a single degree of freedom interaction. In such cases, the statistical significance of the interaction can be determined either by conducting a hierarchical test of changes in χ^2 values reflecting model fit or by examining the significance test of the logistic coefficient associated with the single product term. If the logistic coefficient for the product term is not statistically significant, then this implies that the interaction effect is not statistically significant. Hierarchical test uses differences in χ^2 results based

on likelihood ratio statistics. The alternative criterion at the level of coefficients is also used. This criterion is called Wald test.

For example, suppose that χ^2 value of the "no interaction" model is 24.75 with df = 3 and χ^2 value of the " interaction" model is 34.19 with df = 5. The difference in the χ^2 value is 34.19 - 24.75 = 9.44, which is distributed as a χ^2 with df equal to the difference in their df, 5 – 3 = 2. Consulting a table of critical χ^2 values for $\alpha = 0.05$ and df = 2, the χ^2 difference is statistically significant and this implies that there is a significant interaction effect.

4.4.2 Breslow Day Test

Rothman and Greenland (1998) suggest the use of stratified data as a temporary tool in data analysis. They suggest that in stratified data, stratum-specific estimates should be calculated first and if interaction is present, stratum-specific estimates should be reported since summary estimates do not convey information on the pattern of variation of stratum-specific estimates. In a situation where data are reasonably consistent, a singular estimate should be calculated either by summarizing stratum-specific estimates or by ignoring the stratification variable, depending on the situation and the p-value for this should be calculated. However, Breslow and Day suggests the procedure to identify interaction. This procedure can be ordered as 3 steps.

- Calculate the appropriate crude measure of association between exposure and outcome. This measure can be risk ratio (RR) or odds ratio (OR).
- Calculate RR's or OR's for the association when data has been stratified according to levels of the third variable (one for each level).
- Use Breslow Day (BD) test.

Breslow-Day statistics is used for stratified analysis of 2×2 tables. BD statistics tests the null hypothesis of homogeneous odds ratios. BD tests the null hypothesis that the odds ratios for the "s" strata are all equal. When the null hypothesis is true, this statistics has an asymptotic chi-square distribution with "s-1" degrees of freedom. Hypothesis of BD test is shown as follows:

$$H_0: OR_1 = OR_2 = \ldots = OR_s \qquad (4.8)$$
$$H_1: OR_1 \neq OR_2 \neq \ldots \neq OR_s$$

OR and RR can be evaluated in 2×2 table as follows:

(exposure (+), outcome (+): a; exposure (+), outcome (-): b; exposure (-), outcome (+): c; exposure (-), outcome (-): d)

$$OR = (a \times d)/(b \times c) \qquad (4.9)$$

$$RR = (a/a+b)/(c/c+d) \qquad (4.10)$$

BD statistics is computed as

$$BD\ \chi^2 = \sum_{i=1}^{s} \frac{[a_i - E(a_i | \text{crude OR})]^2}{V(a_i | \text{crude OR})} \qquad i = 1, 2, \ldots, s \qquad (4.11)$$

Where E, V and i denote expected value, variance and the number of stratum, respectively. The summation does not include any table with a zero row or column. BD test statistics distributes χ^2 with "s-1" degrees of freedom. The BD test requires a large sample size within each stratum, and this limits its usefulness. When BD test is investigated in epidemiology studies, it is said that BD tests whether OR between exposure and outcome is the same as in different risk factor categories. If Breslow-Day p-value is less than 0.05 then H_0 is not rejected. In this situation, there is an interaction between the exposure and risk factor variables.

4.5 Interaction Effect Between Categorical (Qualitative) and Continuous (Quantitative) Independent Variables in Logistic Regression Model

The interaction effect is investigated for categorical or continuous independent variables. Analyses require the use of dummy variables for categorical independent variables. In this section, for the logistic model with two independent variables (categorical or continuous), X and Z, and a product terms, XZ, let X be focal variable and let Z be the moderator variable. Which one (X or Z) is categorical or continuous independent variables will be demostrate in related section.

4.5.1 Interaction Effect Among Categorical Independent Variables

The interaction effect of interest involves categorical independent variables. In this section, such analyses require the use of dummy variables (X and Z).

For an interaction logistic model with two categorical independent variables, the logistic coefficient for any dummy variable for X is conditioned to the reference group for Z. The exponent of the logistic coefficient for any dummy variable for X is the odds ratio that divides the predicted odds for the reference group on X, for the case where the dummy variables on Z equal zero. The exponent of the logistic coefficient for a product term is a ratio of predicted odds ratios. It focuses on the predicted odds for the group scored 1 on the dummy variable for X divided by the predicted odds for the reference group on X and divides this odds ratio when computed for the group scored 1 on the dummy variable for Z by the corresponding odds ratio for the reference group on Z (Jaccard, 2001).

As noted in previous section, interaction effect is tested using hierarchical logistic regression in which one determines whether the product terms significantly improve model fit over and above the case where no product terms are included in the model. This approach involves estimating a model χ^2 values for each of the "no interaction" model and "interaction" model. In addition, BD test can be used to detect interaction effect.

4.5.2 Interaction Effect Between Categorical and Continuous Independent Variables

The interaction effect of interest involves a mixture of categorical and continuous independent variables.

For an interaction logistic model with a continuous variable, X, a categorical variable, Z, and a product term, XZ, for the case of dummy coding on Z, the exponent of the logistic coefficient for X is the multiplicative factor by which the predicted odds change given a 1 unit increase in X for the reference group on Z. The exponent of the logistic coefficient for the product term, XZ, is the ratio of the multiplicative factor by which the predicted odds change given a 1 unit increase in X for the group scored 1 on the dummy variable for Z divided by the corresponding multiplicative factor for the reference group on Z (Jaccard, 2001.

For an interaction logistic model with a categorical variable, X, a continuous variable, Z, and a product terms, XZ, for the case of dummy coding on X, the exponent of the logistic coefficient for a dummy variable of X, is the ratio of the predicted odds for the group scored 1 on the dummy variable divided by the predicted odds for the reference group on X, conditioned on Z=0. The exponent of the logistic coefficient for a product term indicates the multiplicative factor by which the odds ratio comparing the predicted odds for the group scored 1 on X changes given a 1 unit increase in Z (Jaccard, 2001.

In addition, interaction effect is tested using hierarchical logistic regression.

4.5.3 Interaction Effect Among Continuous Independent Variables

The interaction effect of interest involves continuous independent variables. For an interaction logistic model with two continuous variables, X and Z, and a product term, XZ, the exponent of the logistic coefficient for X equals a multiplicative factor by which the predicted odds change given a 1 unit increase in X when Z=0. The exponent of the logistic coefficient for the product term is the multiplicative factor by which the multiplicative factor of X changes given 1 unit increase in Z (Jaccard, 2001.

As section 4.5.2, interaction effect is tested using hierarchical logistic regression.

4.6 Assessment of Interaction Effect

There are some strategies for assessment of interaction effect in design and analysis phases. Restriction is used in the design phase. Stratification and multivariate analysis (model fitting) are used in the analysis phase.

Stratification: Reporting measures of association for each category/level of potential interaction variables (BD Test). Interaction can be determined by stratification.

Restriction: Restricting study subjects to only one category/level of a potential interaction variable. Interaction can be determined by restricting the study population to those with a specific value of the interaction variable. This method also known as specification.

Multivariate Analysis (Model Fitting): If proper model variables are selected then interaction among variables can be determined. Multivariate models can be ordered as logistic regression, conditional logistic regression, poisson regression, Cox's proportinal hazards model, log-linear models, multiple linear regression etc…

4.7 Evaluation of Confounding Effect

Confounding can occur in every epidemiological study (Rodriguez & Llorca, 2004). Confounding in epidemiology is mixing of the effect of the exposure under study on the outcome with that of a third factor that is associated with the exposure and an independent risk factor for the outcome (Dorak, 2006). The consequence of confounding is that the estimated association is not the same as true effect (Dorak, 2006).

The relationship between exposure, outcome and confounder is shown in Figure 4.2.

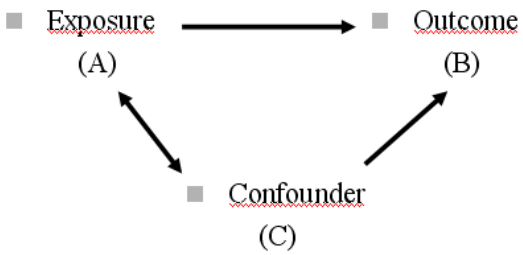

Figure 4.2 The relationship between exposure, outcome and confounder

According to Figure 4.2, the confounder (C) is associated with the exposure of interest (A) but not a consequence of it and the confounder is an independent risk factor for the outcome (B). These are essential characteristics of a confounding variable (Dorak, 2006).

There are some steps for evaluating confounding effect. These can be ordered as follows:

- Stratify the data into subgroups.
- Calculate the effect estimate within each subgroup.
- Calculate a summary effect estimate across strata.

Stratification means that the study population is divided into a number of strata, so that strata within a stratum share a characteristic and each stratum is analysed separately. If the study population is to be divided into more than a few strata, it has to be large to begin with to yield conclusive results. Stratified analyses are the best way to evaluate confounding. Confounding can be adjusted for if the strata are recombined with the Mantel-Haenszel method or a similar method. Mantel-Haenszel method tests the null hypothesis that the odds ratios for the "s" strata are all equal. When the null hypothesis is true, the statistics has an asymptotic chi-square distribution with "s-1" degrees of freedom. In another words, each of stratum specific estimates is unconfounded by risk factor, since there is no variability of the confounding variable within the stratum. In addition, it is necessary to report the unconfounded OR estimate for each stratum and calculate a confidence interval around each estimate. It is also useful to calculate a single overall estimate of the association between exposure and outcome variables, once the effect of the confounding factor has been taken into account. A single overall estimate of the association between exposure and outcome variables that is unconfounded by risk factor is derived from the stratified data by calculating a weighted average of the stratum specific estimates. The method of calculating the overall estimate of effect is often referred to as pooling. A simple method for calculating a pooled summary OR estimate from a series of 2 ×2 tables was proposed by Mantel and Haenszel (Hennekens, C. & Buring, J., 1987). Hypothesis of Mantel-Haenszel test and evaluation of this statistics are shown as follows:

$$H_0: OR_1 = OR_2 = \ldots = OR_s \qquad (4.12)$$

$H_1: OR_i \neq OR_j$ (at least one different)

$$MH - OR = \frac{\sum_{i=1}^{s} a_i d_i / N_i}{\sum_{i=1}^{s} b_i c_i / N_i} \qquad i = 1, 2, \ldots, s \qquad (4.13)$$

Where a: exposure (+), outcome (+); b: exposure (+), outcome (-); c: exposure (-), outcome (+); d: exposure (-), outcome (-)

4.8 Reducing of Confounding Effect

There are some strategies for reducing confounding effect in observational studies. These strategies can be reduced in design and analysis phases. Restriction, matching and randomizing are used in design phase. Stratification and multivariate analysis are used in analysis phase (Jepsen et al., 2004).

- Restriction: Restricting study subjects to only one category/level of a potential confounding variable. Confounding can be reduced by restricting the study population to those with a specific value of the confounding variable. This method, also known as specification, makes examinations of the association between the confounder and the outcome invalid, and the findings can not be generalised to those who were left out by the restriction.

- Matching: Choosing subjects in one comparison group that match on key characteristics of subjects in the other group. Matching constrains subjects in different exposure groups to have the same value of potential confounders, often age and gender. However, with increasing numbers of matching variables, the identification of matched subjects becomes progressively demanding, and matching does not reduce confounding by factors other than the matching variables. Matching is most commonly used in case–control studies, but it can be used in cohort studies as well.

- Stratification: Reporting measures of association for each category/level of potential confounding variables (Mantel Haenszel Method).

- Multivariate Analysis: Statistical methods are used for controlling the effects of one or multiple potential confounding variables. These methods require that the confounding variables are known and measured. These methods can be ordered as logistic regression, conditional logistic regression, poisson regression, Cox’s proportional hazards model, log-linear models and multiple linear regression.

In conclusion, if outcome variable is categorical then logistic regression is used and the relationship between outcome variable and exposure variables are measured using the OR and their 95% CIs derived from logistic regression analysis determining for interaction effect and controlling for possible confounding effect. Crude and stratified ORs are calculated for exposure variables. BD statistics and MH OR are used for interaction effect. Breslow-Day statistics tests whether the OR between investigated exposure and outcome is similar in different risk factor categories. After interaction investigation, confounding effect is investigated. If MH OR is less than the specific evaluated value (crude OR - crude OR×%10) or greater than the other specific evaluated value (crude OR + crude OR×%10) then the risk factor is confounder and MH OR is used instead of crude OR or stratum specific ORs (Boccia et al., 2007).

Dorak (2006) showed some differences between interaction and confounding effects in Table 4.1

Table 4.1 Some differences between interaction and confounding effects

Interaction Effect	**Confounding Effect**
Belongs to nature	Belongs to study
Different effects in different strata	Adjusted OR/RR different from crude OR/RR
Useful	Creates confusion in data
Increases knowledge of biological mechanism	Prevent (design)
Allows targeting of public health action	Control (analysis)

According to this table, confounding effect is related in study design but interaction effect is related in nature of variables. Interaction does not been prevented but it is determined with restriction, stratification and multivariate analysis. Confounding can be prevented and controlled with restriction, matching, randomizing, stratification and multivariate analysis.

CHAPTER FIVE
APPLICATION AND RESULTS

The case-control study is the most common design to investigate gene-environment interactions (Goldstein et al., 1997). Controls are used as the reference group. A growing number of epidemiologic evidence suggests that cancer risk as well as stomach cancer susceptibility results from the combined effect of genes and environment. Several genotypes involved in xenobiotic metabolism are subject to genetic polymorphisms and many researchers have reported an association between these polymorphisms and suspectibility to cancer. Among these genotypes, GSTM1 and GSTT1 are extensively studied. Genetic polymorphism of GSTM1 and GSTT1 genotypes are due to quite frequent presence of null alleles.

In many epidemiologic studies, stratified and logistic regression analysis are usually used to estimate odds ratios of results from the combined effect of genes and environment for cases and controls. Stomach cancer investigated in epidemiologic studies is a disease with complex origins. The roles and associations of the genetic and environmental factors in stomach cancer should be recognized for Turkish population. Because there has been no detailed study of the association of GSTM1&GSTT1 and stomach cancer suspectibility in a Turkish population.

The aim of this case-control study is to determine whether there is any relationship that can be defined as interaction and confounding effects between genetic polymorphism of GSTM1 and GSTT1 genotypes and risk factor as smoking status in stomach cancer between 2006-2009 years in Dokuz Eylül University in İzmir.

This chapter consists on two main sections. Firstly, a detailed description of the data source is given. This section outlines the study population, stomach cancer cases and controls, variables of risk factors (demographic, medical, drinking and

nutritional factors) and epidemiological data collection. Secondly, material and methods of the study are given. This section outlines statistical analyses for stomach cancer to determine the association of genetic (GSTM1 and GSTT1 genotypes) and environmental factors (smoking, age, gender) in accordance with interaction and confounding effects.

5.1 Study Population

A case-control study of stomach cancer was conducted in İzmir, Turkey between 2006-2009 years. İzmir is Turkey's third most populous city. The city of İzmir is composed of nine metropolitan districts (Balçova, Bornova, Buca, Çiğli, Gaziemir, Güzelbahçe, Karşıyaka, Konak and Narlıdere), each with its own distinct features. This study was conducted in Dokuz Eylül University Hospital (DEUH) in Balçova.

5.2 Cancer Cases and Controls

The study population consisted of 127 stomach cancer cases who attended the General Surgery Department at DEUH in İzmir. These eligible stomach cancer cases were examined from January 1, 2006, to December 30, 2008, with pathologically confirmed diagnoses of stomach adenocarcinoma. The control group consisted of 101 cancer-free individuals who applied to outpatient clinics of DEUH with acute conditions. Eligible controls were healthy and cancer-free individuals. The control group was restricted in specific years old. Younger than eighteen years individuals were not allowed to participate in this group. Garcia & Lubin (1999) calculated power and sample size in case-control studies of gene-environment interactions. According to Garcia & Lubin's study, study sample size was calculated in Epi-Info package program. Components of this sample size were power (80% or 90%), confidence interval (95%), prevalance of genetic polymorphism (20%, 30%, 40%, 50%), odds ratio (2.35 and 3.71) and control/case ratio (1:1 and 2:1). The sample size calculations were given in Appendix 1. According to Appendix 1, components of this sample size were 80% power, 95% confidence interval, 50% prevalance of genetic polymorphism, 2.35 odds ratio and 1:1 control/case ratio.

5.3 Variables of Risk Factors

Dependent variable was to be case or control. Risk factors were independent variables of this study. Risk factors for stomach cancer were investigated in six category as follows:

i) Demographic Factors (Gender, Age, Place of Residence and Birth, Education Status, Marital Status, Job Status, Health Care Insurance, Socioeconomic Status)
ii) Medical Factors (Body Mass Index (BMI), Weight Loss, Health Problem, Family Cancer History, Smoking and Cumulative Cigarette Smoking (Pack-years))
iii) Drinking Habits (Alcohol Consumption, Tea, Herbal Tea, Instant Coffee, Turkish Coffee and Soft Drinks)
iv) Nutritional Factors (Salt, Margarine, Oil, Butter, Sugar, Hot Food, Spicy Food, Red Meat and Poultry Intakes)
v) Pathological Factors (Location of Tumor, Histological Type)
vi) Genotype Factors (Polymorphisms of the GSTM1 and GSTT1)

Independent variables in stomach cancer cases and controls for demographic characteristics were gender, age, birth place, residence place, education, marital status, job status, insurance and income. All of them were categorical variables. Age variable was divided into two groups as old&young. Older individuals were greater than 61 years old. The categorization of age variable was based on the most recently published literature. Birth place and residence place variables were divided into two groups as “other city” or “İzmir”. Education variable was divided into four groups as “no education”, “primary school”, “middle&high school” and “university”. Marital status variable was divided into two groups as single&married. Job status variable was divided into four groups as “not working”, “working”, “retired” and “house wife”. Insurance variable was divided into three groups as “Emekli Sandığı”, “SSK” and “others”. Income variable was divided into three groups as “less than 499 TL”, “between 500 and 1000 TL” and “greater than 1001 TL”.

Independent variables in stomach cancer cases and controls for medical history were BMI, weight loss, health problem, family cancer history, smoking and pack-years. BMI variable (weight/height^2) was divided into four groups as “thin (0-20)”, “normal (21-25)”, “fat (26-30)” and “obese (31 and more)”. Weight loss and health problem variables were divided into two groups as present&absent. Weight loss variable is loosing weight situation of individulas in last six month. Family cancer history variable was divided into two groups as yes/no. The categorization of smoking variable was based on the most recently published literature. Status of smoking variable was divided into two groups as “never” and “ever-smokers (current and former)”. Former smokers were defined as persons who quit smoking 1 year or more before being diagnosed in cases. Information was collected on the usual number of cigarettes smoked per day, the age at which the subject started smoking and the age at which the subject stopped smoking if the person was a former smoker. Pack-years were calculated for the cumulative cigarette smoking. One pack-year was defined as smoking 20 cigarettes daily for one year. For example, a patient who has smoked 15 cigarettes a day for 40 years has a (15×40)/20 = 30 pack year smoking history. Pack-years variable was divided into two groups as >20 & ≤ 20.

Independent variables in stomach cancer cases and controls for different drinking characteristics were alcohol, tea, herbal tea, instant coffee, Turkish coffee and soft drink intakes. All of these variables were divided into two groups as yes/no.

Independent variables in stomach cancer cases and controls for nutritional characteristics were salt intake, margarine intake, oil intake, butter intake, sugar intake, sugar intake, hot food intake, spicy food intake, red meat intake and poultry intake. All of these variables were divided into two groups as user&non-user.

Genetic polymorphism of GSTM1 and GSTT1 genotypes are due to quite frequent presence of null alleles. GSTM1 and GSTT1 genotype frequences in stomach cancer cases according to histological types were divided into three groups as “Adenocarcinoma”, “Diffuse Adenocarcinoma” and “Intestinal Adenocarcinoma”. GSTM1 and GSTT1 genotype frequences in stomach cancer cases according to

location of stomach cancer were divided into three groups as “upper tumors (cardia)”, “middle tumors (fundus)” and “distal tumors (corpus, antrum)”.

5.4 Data Collection

Data included interview-questionnaire data, medical record review and blood samples. A written informed consent was obtained from the entire case and control group and the study was approved by the local Ethical Committee of Dokuz Eylül Medical Faculty. The blood samples collected from both cases and controls were processed at the study laboratory in DEUH. The medical records for patients were abstracted for relevant clinical data including endoscopy and pathology examinations. In medical records, the location of tumor was defined according to International Classification of Disease for Oncology, second edition (ICD-O-2). Blood samples were used to obtain for GSTM1 and GSTT1 assays. A polymerase chain reaction (PCR) method was used to detect the presence or absence of the GSTM1 and GSTT1 genotypes. In other words, GSTM1(+) & GSTT1(+) (wild type or normal genotype) and GSTM1(-) & GSTT1(-) (null genotype) were analyzed by using PCR method. Absence of GSTM1(-) & GSTT1(-) specific PCR products indicated that GSTM1&GSTT1 genotypes were deleted on both alleles (-/-). In contrast, the presence of gene specific PCR product indicated at least 1 functional allele (+/-, +/+). Genotype data were available for 79 cases and for 99 controls.

Cases and controls were interviewed by using a standard epidemiological questionnaire with 20 questions consisted on these risk factors of stomach cancer in Appendix 2. According to the most recently published literature, the information was collected on different risk factors in epidemiological questionnaire and medical record as demographic factors, medical factors, drinking habits, nutritional factors, pathological factors and genotype factors.

5.5 Statistical Analysis

The relationships between stomach cancer and determined risk factors (demographic, medical, drinking, nutritional and genotype factors) were assessed using ORs and 95% CIs derived from univariate, stratified and multivariate analyses using SPSS and Epi-Info software. The categorizations of risk factors were based on the most recently published literature.

In univariate analysis, crude ORs and their 95% CIs were calculated for all of risk factors. The associations between different risk factors and cases&controls were analyzed in this analysis.

In stratified analysis, interaction and confounding checks were performed. Interaction effect was investigated before confounding effect. An heterogenity test (BD test) was used to test differences among the strata for interaction. Biological interaction between GSTM1&GSTT1 genotypes and some risk factors (age, gender and smoking) was estimated using departure from additivity or multiplicative of effects as the criterion of interaction as suggested by Rothman & Greenland (1998). If MH OR is less or greater than the specific evaluated value then the risk factor is confounder and MH OR is used instead of crude OR or stratum specific ORs (Boccia et al., 2007).

In multivariate analysis, dummy variables were used to estimate ORs for each category of exposure. The association among GSTM1 and GSTT1 genotypes, smoking, gender, age and stomach cancer was modelled through logistic regression analysis.

5.6 Results

5.6.1 General Characteristics of the Study Population

General characteristics (demographic, medical, drinking and nutritional) of the study population were presented in Tables 5.1, 5.2, 5.3 and 5.4, respectively.

The distribution of variables in stomach cancer cases and controls for demographic characteristics was presented in Table 5.1.

Table 5.1 Distribution of variables in stomach cancer cases and controls for demographic characteristics

Demographic Characteristics	Cases		Controls		p-value [b]	Crude OR	OR CI (95%)
	n	%	n	%			
Gender							
Male	72	56.7	37	36.6	0.0030*	2.26	1.32-3.87
Female [a]	55	43.3	64	63.4		1.00	
Age							
Old ($\geq$61)	77	60.6	25	24.8	0.0001*	4.68	2.63-8.32
Young (<61) [a]	50	39.4	76	75.2		1.00	
Birth Place							
Other City	87	68.5	69	31.7	0.9760	1.01	0.57-1.77
Izmir [a]	40	31.5	32	68.3		1.00	
Residence Place							
Other City	42	33.1	25	24.8	0.1710	1.50	0.84-2.69
Izmir [a]	85	66.9	76	75.2		1.00	
Education							
No Edu.	22	17.3	5	5.0	0.0001*	10.06	2.82-38.26
Primary Sch.	40	31.5	39	38.6		2.34	1.02-5.44
Middle&High Sch.	51	40.2	25	24.8		4.66	1.98-11.14
University [a]	14	11.0	32	31.6		1.00	
Marital Status							
Single	19	15	22	21.8	0.1830	0.63	0.32-1.25
Married [a]	108	85	79	78.2		1.00	
Job Status							
Not Working	0	0	4	1.8	0.0001*	-	-
Working	35	27.6	44	34.6		0.65	0.33-1.28
Retired	47	37.0	16	27.6		2.42	1.12-5.27
House Wife [a]	45	35.4	37	36.0		1.00	
Insurance							
Emekli Sandığı	62	48.8	40	39.6	0.3350	1.28	0.58-2.82
SSK	42	33.1	42	41.6		0.83	0.37-1.85
Others [a, c]	23	18.1	19	18.8		1.00	
Income (TL)							
0-499	33	32.7	19	19.0	0.0860	2.07	0.93-4.65
500-1000	37	36.6	44	44.0		1.00	0.50-2.02
1001 [a, d]	31	30.7	37	37.0		1.00	

a: Referent

b: It is obtained from χ^2 test

c: Bağkur, Private Insurance et al.

d: The information for one person was not obtained in income variable

***: p-value < 0.05**

Lots of cases were males (56.7%), whereas lots of controls were females (63.4%). There was statistically significant difference in gender between two groups (p-value = 0.003). Gender was the important risk factor for cases. Case group was older than control group. 60.6% of the case group was old and 75.2% of the control group was young. There was statistically significant difference in age between two groups (p-value = 0.0001). 66.9% of the cases were living in İzmir and 75.2% of the controls were living in İzmir. The proportion of “no education” status was higher among cases (17.3%) than those of the controls (5.0%) as expected. In addition, indivudals (n = 51) in “middle&high school” category were more among cases but indivudals (n = 39) in “primary school” category were more in controls. There was statistically significant difference in education between two groups (p-value = 0.0001). Individuals who had “no education” were 10.06 times more likely to have cancer than individuals who graduated from university. According to marital status variable, most indivuduals were married both cases (85.0%) and controls (78.2%). Working individulas also were more in every two groups. There was statistically significant difference in job status variable between two groups (p-value = 0.0001). “Emekli Sandığı” was most common health care provider among cases (48.8%) but “SSK” was among controls (41.1%). However, “Emekli Sandığı” (39.6%) was following health care provider among controls after “SSK”. Socieconomic status was lower in cases than controls. There were no statistically significant difference in birth place, residence place, marital status, insurance and income variables between two groups.

The distribution of variables in stomach cancer cases and controls for medical history was presented in Table 5.2.

Table 5.2 Distribution of variables in stomach cancer cases and controls for medical history

Medical History	Cases		Controls		p-value [b]	Crude OR	OR CI (95%)
	n	%	n	%			
BMI							
0-20: Thin	22	17.3	11	10.9	0.0001*	0.93	0.37-2.33
21-25: Normal [a]	69	54.3	32	31.7		1.00	
26-30: Fat	26	20.5	33	32.7		0.37	0.18-0.75
31-: Obese	10	7.9	25	24.7		0.19	0.07-0.46
Weight Loss							
Present	96	75.6	38	37.6	0.0001*	5.13	2.90-9.09
Absent [a]	31	24.4	63	62.4		1.00	
Other Health Problem							
Present	71	55.9	61	60.4	0.4950	0.83	0.49-1.41
Absent [a]	56	44.1	40	39.6		1.00	
Family Cancer History							
Yes	45	35.4	31	30.7	0.4510	1.24	0.71-2.16
No [a]	82	64.6	70	69.3		1.00	
Smoking							
Smoker	71	55.9	44	43.6	0.0640**	1.64	0.97-2.78
Non-Smoker [a]	56	44.1	57	56.4		1.00	
Pack-Years							
(>20)	52	73.2	18	40.9	0.001*	3.95	1.78-8.78
($\leq$20)	19	26.8	26	59.1		1.00	
Pack-Years							
(>31)	39	54.9	12	27.3		4.45	1.70-11.82
(21-30)	13	18.3	6	13.6	0.001*	2.96	0.84-10.80
($\leq$20) [a]	19	26.8	26	59.1		1.00	
a: Referent							
b: It is obtained from χ^2 test							
***: p-value < 0.05**							
****: p-value < 0.10**							

According to BMI variable, lots of cases were normal (54.3%). But lots of controls were fat (32.7%). There was statistically significant difference in BMI between two groups (p-value = 0.0001). The proportion of weight loss variable was significantly higher among cases (75.6%) than those of the controls (37.6%) as expected. There was no difference in family cancer history variable between two groups (p-value = 0.451). Smoking prevalence were 55.9% and 43.6% in cases and controls, respectively. The difference was not statistically significant in smoking variable between two groups (p-value = 0.064). But this variable was investigated

with a different categorized variable. This categorized variable was the cumulative cigarette smoking and this variable was called pack-years in this study. According to this variable, there was significant difference in pack-years variable between two groups (p-value = 0.001). Using cigarette more than 20 pack-years was increasing the risk 3.95 times for stomach cancer compared to less than 20 pack-years category.

The distribution of variables in stomach cancer cases and controls for different drinking characteristics was presented in Table 5.3.

Table 5.3 Distribution of variables in stomach cancer cases and controls for different drinking characteristics

Drinking Characteristics	**Cases**		**Controls**		**p-value** [b]	**Crude OR**	**OR CI (95%)**
	n	**%**	**n**	**%**			
Alcohol							
Yes	31	24.4	23	22.8	0.7730	1.09	0.59-2.03
No [a]	96	75.6	78	77.2		1.00	
Tea							
Yes	121	95.3	98	97.0	0.3750[c]	0.62	0.15-2.53
No [a]	6	4.7	3	3.0		1.00	
Herbal Tea							
Yes	43	33.9	40	39.6	0.3700	0.78	0.45-1.34
No [a]	84	66.1	61	60.4		1.00	
Instant Coffee							
Yes	32	25.2	35	34.7	0.1190	0.63	0.36-1.13
No [a]	95	74.8	66	65.3		1.00	
Turkish Coffee							
Yes	37	29.1	53	52.5	0.0001*	0.37	0.21-0.64
No [a]	90	70.9	48	47.5		1.00	
Soft Drinks							
Yes	33	26.2	52	51.5	0.0001*	0.33	0.19-0.58
No [a]	93	73.8	49	48.5		1.00	
a: Referent							
b: It is obtained from χ^2 test							
c: Fisher Exact Test p-value							
***: p-value < 0.05**							

There were significant differences in Turkish coffee intake (p-value = 0.0001) and soft drink intake (p-value = 0.0001) variables between two groups. This difference was probably due to digestive problems of the cases thus they were consuming these

drinks less than the controls. In adddition, there were no significant differences in all other variables.

The distribution of variables in stomach cancer cases and controls for nutritional characteristics was presented in Table 5.4.

Table 5.4 Distribution of variables in stomach cancer cases and controls for nutritional characteristics

Nutritional Characteristics	Cases		Controls		p-value [b]	Crude OR	OR CI (95%)
	n	%	n	%			
Salt							
User	97	76.4	55	54.5	0.0001*	2.70	1.53-4.76
Non-User [a]	30	23.6	46	45.5		1.00	
Margarine							
User	90	70.9	31	30.7	0.0001*	5.49	3.10-9.72
Non-User [a]	37	29.1	70	69.3		1.00	
Oil							
User	119	93.7	97	96.0	0.4320	0.61	0.18-2.10
Non-User [a]	8	6.3	4	4.0		1.00	
Butter							
User	107	84.3	40	39.6	0.0001*	8.16	4.38-15.20
Non-User [a]	20	15.7	61	60.4		1.00	
Sugar							
User	108	85.0	61	60.4	0.0001*	3.73	1.98-7.00
Non-User [a]	19	15.0	40	39.6		1.00	
Hot Food							
User	86	67.7	58	57.4	0.1100	1.55	0.90-2.67
Non-User [a]	41	32.3	43	42.6		1.00	
Spicy Food							
User	113	89.0	64	63.4	0.0001*	4.67	2.35-9.28
Non-User [a]	14	11.0	37	36.6		1.00	
Red Meat							
User	99	78.0	58	57.4	0.0010*	2.62	1.47-4.66
Non-User [a]	28	22.0	43	42.6		1.00	
Poultry							
User	113	89.0	84	83.2	0.2040	1.63	0.76-3.50
Non-User [a]	14	11.0	17	16.8		1.00	
a: Referent							
b: It is obtained from χ^2 test							
***: p-value < 0.05**							

There were significant differences in all variables except oil intake, hot food intake and poultry intake variables between two groups.

5.6.2 The Association of GSTM1 and GSTT1 Genotypes and Stomach Cancer

Differences in the distribution between cases and controls were tested using χ^2 test for some characteristics of the study populations in univariate analysis. In addition, OR values and CIs between cases and controls groups were evaluated. After these calculations, comparison of genotype frequencies between stomach cancer cases and population controls was also performed by χ^2 test. The association between the GSTM1 and GSTT1 genotypes and the risk of stomach cancer was estimated by ORs and their 95% CIs in Table 5.5. In this data set, the GSTM1 and GSTT1 genotypes of 79 cases and 99 controls could be determined by PCR analysis.

Table 5.5 GSTM1 and GSTT1 genotype and risk of stomach cancer

	Cases	**Controls**	**Total**	**p-value** [b]	**Crude OR**	**OR CI (95%)**
	n (%)	**n (%)**	**N**			
GSTM1						
Null	46 (58.2)	46 (46.5)	92	0.119	1.61	0.88-2.92
Normal [a]	33 (41.8)	53 (53.5)	86		1.00	
Total	79 (100.0)	99 (100.0)	178			
GSTT1						
Null	18 (22.8)	22 (22.2)	40	0.929	1.03	0.51-2.10
Normal [a]	61 (77.2)	77 (77.8)	138		1.00	
Total	79 (100.0)	99 (100.0)	178			
a: Referent						
b: It is obtained from χ^2 test						
Only Cases and Controls by Blood Sample Availability						

The GSTM1 and GSTT1 genotype prevalences in stomach cancer cases and controls were shown. The prevalence of GSTM1 null genotype was 58.2% in cases, 46.5% in controls. The prevalence of GSTT1 null genotype was 22.8% in cases, 22.2% in controls. Using normal GSTM1 and GSTT1 genotypes as referent, the crude ORs for GSTM1 and GSTT1 were found as 1.61 (95% CI, 0.88-2.92) and 1.03 (95% CI, 0.51-2.10), respectively. Although the frequency of GSTM1 null genotype was higher in cases it was not significant (p-value = 0.119). In addition, no difference in the frequency of GSTT1 null genotype between cases and controls was also observed (p-value = 0.929).

GSTM1 and GSTT1 genotype frequences in stomach cancer cases according to histological types and location of stomach cancer were given in Tables 5.6 and 5.7, respectively (p-value is obtained from χ^2 test).

Table 5.6 GSTM1 and GSTT1 genotype frequences in stomach cancer cases according to histological types

	GSTM1			**GSTT1**		
	Null	**Normal**	**Total**	**Null**	**Normal**	**Total**
	n (%)	**n (%)**	**N (%)**	**n (%)**	**n (%)**	**N (%)**
Adenocarcinoma	19 (57.6)	14 (42.4)	33 (100)	9 (27.3)	24 (72.7)	33 (100)
Diffuse	17 (65.4)	9 (34.6)	26 (100)	6 (23.1)	20 (76.9)	26 (100)
Intestinal	10 (50.0)	10 (50)	20 (100)	3 (15.0)	17 (85.0)	20 (100)
Total Cases	46 (58.2)	33 (41.8)	79 (100)	18 (22.8)	61 (77.2)	79 (100)
p-value		0.574			0.586	

Lots of histological types of cases were “Adenocarcinoma” for GSTM1 and GSTT1 genotypes. There was no relavant association between GSTM1 genotype and histological types of stomach cancer cases (p-value = 0.574). Also for GSTT1 genotype, there was no relavant association between this genotype and histological types of stomach cancer cases (p-value = 0.586).

Table 5.7 GSTM1 and GSTT1 genotype frequences in stomach cancer cases according to location of stomach cancer

	GSTM1			GSTT1		
	Null	**Normal**	**Total**	**Null**	**Normal**	**Total**
Location	**n (%)**	**n (%)**	**N (%)**	**n (%)**	**n (%)**	**N (%)**
Antrum&Corpus	23 (62.2)	14 (37.8)	37 (100)	7 (18.9)	30 (81.1)	37 (100)
Cardia	5 (50.0)	5 (50.0)	10 (100)	1 (10.0)	9 (90.0)	10 (100)
Fundus	7 (77.8)	2 (22.2)	9 (100)	2 (33.3)	6 (66.7)	9 (100)
Total Cases	35 (62.5)	21 (37.5)	56 (100)	11 (33.3)	45 (80.4)	53 (100)
p-value		0.457			0.434	

In Table 5.7, lots of tumors were located in distal stomach (corpus, antrum). There were no differences in GSTM1 and GSTT1 genotype frequencies in respect to the location of stomach cancer.

5.6.3 Stratified Analysis for Interaction and Confounding

Interaction and confounding effects were analyzed in stratified analysis.

The interaction and confounding between smoking status and gender in stomach cancer were examined by stratification in Table 5.8. Gender was a potential confounder.

Table 5.8 Association of smoking status and stomach cancer stratified by gender

Gender	Smoking Status	Case n (%)	Control n (%)	p-value [b]	OR	OR CI (95%)
Male	**Smoker**	48 (66.7)	21 (56.8)	0.309	1.52	0.67-3.44
	Non-Smoker [a]	24 (33.3)	16 (43.2)		1.00	
	Total	72 (100)	37 (100)			
Female	**Smoker**	23 (41.8)	23 (35.9)	0.511	1.28	0.61-2.69
	Non-Smoker [a]	32 (58.2)	41 (64.1)		1.00	
	Total	55 (100)	64 (100)			
Crude OR For Smoking					1.64	0.97-2.78
BD p-value				0.754		
MH OR					1.38	
MH CI (95%)						0.80-2.39
a: Referent						
b: It is obtained from χ^2 test						

The risk of stomach cancer related to smoking status was examined by stratification of gender. There was no statistically significant association between smoking status and stomach cancer in male individuals (p-value = 0.309). For female individuals, there was also no statistically significant association between smoking status and stomach cancer (p-value = 0.511). According to BD p-value, the null hypothesis of homogeneity ORs was not rejected. This means that ORs for the gender strata were similar. This analysis stratified by gender showed that the association between smoking status and the risk of stomach cancer development was not more evident among male (OR = 1.52, 95% CI = 0.67-3.44) and female (OR = 1.28, 95% CI = 0.61-2.69). In other words, there was no interaction between gender and smoking status in stomach cancer. After interaction investigation, confounding effect was investigated. If MH OR is less than 1.48 (1.64 – 0.164) or greater than 1.80 (1.64 + 0.164) then gender is confounder and MH OR is used ($1.64 \times 0.1 = 0.164$). MH OR, 1.38, was used instead of crude OR or stratum specific ORs. According to this result, gender was a confounder.

The interaction and confounding between smoking status and age in stomach cancer were examined by stratification in Table 5.9.

Table 5.9 Association of smoking status and stomach cancer stratified by age

Age	Smoking Status	Case n (%)	Control n (%)	p-value [b]	OR	OR CI (95%)
Old	**Smoker**	42 (54.5)	9 (36.0)	0.107	2.13	0.84-5.42
	Non-Smoker [a]	35 (45.5)	16 (64.0)		1.00	
	Total	77 (100)	25 (100)			
Young	**Smoker**	29 (58.0)	35 (46.1)	0.189	1.62	0.79-3.32
	Non-Smoker [a]	21 (42.0)	41 (53.9)		1.00	
	Total	50 (100)	76 (100)			
Crude OR For Smoking					1.64	0.97-2.78
BD p-value				0.645		
MH OR					1.80	
MH CI (95%)						1.02-3.17
a: Referent						
b: It is obtained from χ^2 test						

The risk of stomach cancer related to smoking status was examined by stratification of age. There was no statistically significant association between smoking status and stomach cancer for older individulas (p-value = 0.107). For younger individuals, there was also no statistically significant association between smoking status and stomach cancer (p-value = 0.189). According to BD p-value, strata for age were homogenous (BD p-value = 0.645). There was no statistically significant interaction between the age and smoking status in stomach cancer. MH OR, 1.80, was equal to 1.80 (1.64 + 0.164). Age was treated as a confounder and MH OR is used instead of crude OR or stratum specific ORs.

After investigation of interaction and confounding effects between smoking status and age in stomach cancer cases and controls by stratification, smoking status was investigated in the different form such as pack-years. The interaction and

confounding between pack-years status and gender in stomach cancer were examined by stratification in Table 5.10.

Table 5.10 Association of pack-years status and stomach cancer stratified by gender

Gender	Pack-Years(PY)	Case n (%)	Control n (%)	p-value [b]	OR	OR CI (95%)
Male	**PY > 20**	43 (89.6)	13 (61.9)	0.007	5.29	1.47-18.9
	PY ≤ 20 [a]	5 (10.4)	8 (38.1)		1.00	
	Total	48 (100)	21 (100)			
Female	**PY > 20**	9 (39.1)	5 (21.7)	0.200	2.31	0.63-8.47
	PY ≤ 20 [a]	14 (60 9)	18 (78.3)		1.00	
	Total	23 (100)	23 (100)			
Crude OR For Pack-Years					3.95	1.78-8.78
BD p-value				0.373		
MH OR					3.45	
MH CI (95%)						1.39-8.57
a: Referent						
b: It is obtained from χ^2 test						

The risk of stomach cancer related to pack-years status was examined by stratification of gender. There was statistically significant association between pack-years status and stomach cancer in male individuals (p-value = 0.007). For female individuals, there was an increased risk between pack-years status and stomach cancer but this was not statistically significant (p-value = 0.200, OR = 2.31, 95% CI = 0.63-8.47). According to BD p-value, strata for gender were homogenous (BD p-value = 0.373). There was no statistically significant interaction between pack-years status and gender in stomach cancer. MH OR, 3.45, was less than 3.55 (3.95 - 0.395). For this reason gender was confounder and MH OR is used instead of crude OR or stratum specific ORs. But it was noted that this stratified analysis showed that the association between pack-years status and the risk of stomach cancer development was more evident among male individuals (OR = 5.29, 95% CI = 1.47–18.9).

The interaction and confounding between pack-years status and age in stomach cancer were also examined by stratification in Table 5.11.

Table 5. 11 Association of pack-years status and stomach cancer stratified by age

Age	Pack-Years(PY)	Case n (%)	Control N (%)	p-value [b]	OR	OR CI (95%)
Old	**PY > 20**	36 (85.7)	5 (55.6)	0.039	4.80	0.99-23.1
	PY ≤ 20 [a]	6 (14.3)	4 (44.4)		1.00	
	Total	42 (100)	9 (100)			
Young	**PY > 20**	16 (55.2)	13 (37.1)	0.149	2.08	0.76-5.68
	PY ≤ 20 [a]	13 (44.8)	22 (62.9)		1.00	
	Total	29 (100)	35 (100)			
Crude OR For Pack-Years					3.95	1.78-8.78
BD p-value				0.376		
MH OR					2.58	
MH CI (95%)						1.10-6.03

a: Referent

b: It is obtained from χ^2 test

The risk of stomach cancer related to pack-years status was examined by stratification of age. There was statistically significant association between pack-years status and stomach cancer in older individuals (p-value = 0.039). But the association between pack-years status and stomach cancer in younger individuals was not significant (p-value = 0.149). On the other hand, there was an increased risk between pack-years status and stomach cancer in older individuals, but this incerase did not reach statistical significance (OR = 4.80, 95% CI = 0.99-23.1). According to BD p-value, strata for age were homogenous (BD p-value = 0.376). For this reason, it was expressed that there was no statistically significant interaction between pack-years status and age. MH OR, 2.58, was less than 3.55 (3.95 - 0.395). For this reason, age was confounder and MH OR was used instead of crude OR or stratum specific ORs.

We focused on the effect of GSTM1 and GSTT1 null genotypes in stomach cancer. The stratified and logistic regression analysis were used to assess the interaction and confounding effects between GSTM1 and GSTT1 null genotypes and some other possible risk factors (age, gender, smoking) in stomach cancer.

The interaction and confounding between GSTM1 genotype and gender in stomach cancer were examined by stratification in Table 5.12.

Table 5.12 Association of GSTM1 genotype and stomach cancer stratified by gender

Gender	GSTM1	Case n (%)	Control n (%)	p-value [b]	OR	OR CI (95%)
Male	**Null**	29 (61.7)	21 (60.0)	0.876	1.07	0.44-2.63
	Normal [a]	18 (38.3)	14 (40.0)		1.00	
	Total	47 (100)	35 (100)			
Female	**Null**	17 (53.1)	25 (39.1)	0.190	1.77	0.75-4.16
	Normal [a]	15 (46.9)	39 (60.9)		1.00	
	Total	32 (100)	64 (100)			
Crude OR For GSTM1					1.61	0.88-2.92
BD p-value				0.430		
MH OR					1.39	
MH CI (95%)						0.75-2.58

a: Referent

b: It is obtained from χ^2 test

The risk of stomach cancer related to GSTM1 genotype was examined by stratification of gender. There was no association between GSTM1 genotype and stomach cancer in male individuals (p-value = 0.876) and also in female individuals (p-value = 0.19). According to BD p-value, the null hypothesis of homogeneity ORs was not rejected and strata for gender were homogenous (BD p-value = 0.430). For this reason, it was expressed that there was no statistically significant interaction between gender and GSTM1 genotype in stomach cancer. MH OR, 1.39, was less than 1.45 (1.61 - 0.161). Thus, gender was confounder and MH OR is used instead of crude OR.

After grouping according to gender, the interaction and confounding between GSTM1 genotype and age in stomach cancer were examined by stratification in Table 5.13.

Table 5.13 Association of GSTM1 genotype and stomach cancer stratified by age

Age	GSTM1	Case n (%)	Control n (%)	p-value [b]	OR	OR CI (95%)
Old	**Null**	27 (57.4)	11 (45.8)	0.353	1.59	0.59-4.29
	Normal [a]	20 (42.6)	13 (54.2)		1.00	
	Total	47 (100)	24 (100)			
Young	**Null**	19 (59.4)	35 (46.7)	0.229	1.67	0.72-3.86
	Normal [a]	13 (40.6)	40 (53.3)		1.00	
	Total	32 (100)	75 (100)			
Crude OR For GSTM1					1.61	0.88-2.92
BD p-value				0.945		
MH OR					1.64	
MH CI (95%)						0.86-3.11
a: Referent						
b: It is obtained from χ^2 test						

The risk of stomach cancer related to GSTM1 genotype was examined by stratification of age. There was no statistically association between GSTM1 genotype and stomach cancer in older (p-value = 0.353) or younger individuals (p-value = 0.229). According to BD p-value, strata for age were homogenous (BD p-value = 0.945). There was no interaction between age and GSTM1 genotype in stomach cancer. MH OR, 1.64, was between 1.45 (1.61 - 0.161) and 1.77 (1.61 + 0.161). For this reason, age was not confounder.

After the association between GSTM1 genotype and some risk factors as gender and age were investigated, the interaction and confounding between GSTT1 genotype and gender in stomach cancer were examined by stratification in Table 5.14.

Table 5.14 Association of GSTT1 genotype and stomach cancer stratified by gender

Gender	GSTT1	Case n (%)	Control n (%)	p-value [b]	OR	OR CI (95%)
Male	**Null**	8 (17)	10 (28.6)	0.211	0.51	0.18-1.47
	Normal [a]	39 (83)	25 (71.4)		1.00	
	Total	47 (100)	35 (100)			
Female	**Null**	10 (31.2)	12 (18.7)	0.170	1.97	0.74-5.23
	Normal [a]	22 (68.8)	52 (81.3)		1.00	
	Total	32 (100)	64 (100)			
Crude OR For GSTT1					1.03	0.51-2.10
BD p-value				0.064		
MH OR					1.05	
MH CI (95%)						0.52-2.12
a: Referent						
b: It is obtained from χ^2 test						

The risk of stomach cancer related to GSTT1 genotype was examined by stratification of gender. There was no statistically association between GSTT1 genotype and stomach cancer in male individuals (p-value = 0.211). There was also no statistically significant association between GSTT1 genotype and stomach cancer in female individuals (p-value = 0.17). According to this analysis, the existence of GSTT1 null genotype in male individulas was protective effect (OR = 0.51, 95% CI = 0.18-1.47). The risk of being stomach cancer for female individulas who have GSTT1 null genotype was 1.97 times more effective to female individulas who have GSTT1 normal genotype. OR values for two strata suggested that there was interaction effect of gender on GSTT1 status. However the BD p-value, strata for gender was not significant (p-value = 0.064). Still an interaction effect may be suspected due to the nearly significant p-value.

The interaction and confounding between GSTT1 genotype and age in stomach cancer were examined by stratification in Table 5.15.

Table 5.15 Association of GSTT1 genotype and stomach cancer stratified by age

Age	GSTT1	Case n (%)	Control n (%)	p-value [b]	OR	OR CI (95%)
Old	**Null**	13 (27.7)	5 (20.8)	0.532	1.45	0.45-4.70
	Normal [a]	34 (72.3)	19 (79.2)		1.00	
	Total	47 (100)	24 (100)			
Young	**Null**	5 (15.6)	17 (22.7)	0.409	0.63	0.21-1.89
	Normal [a]	27 (84.4)	58 (77.3)		1.00	
	Total	32 (100)	75 (100)			
Crude OR For GSTT1					1.03	0.51-2.10
BD p-value				0.308		
MH OR					0.93	
MH CI (95%)						0.43-2.01
a: Referent						
b: It is obtained from χ^2 test						

The risk of stomach cancer related to GSTT1 genotype was examined by stratification of age. There was no association between GSTT1 genotype and stomach cancer in older (p-value = 0.532) or younger individuals (p-value = 0.409). As seen in Table 5.15, there was an increased risk between GSTT1 genotype and stomach cancer in older individuals, but this incerase did not reach statistical significance (OR = 1.45, 95% CI = 0.45-4.70). On the other hand, the existence of GSTT1 null genotype in younger individulas was protective effect, but effect was not significant (OR = 0.63, 95% CI = 0.51-2.10). According to BD p-value, strata for age were homogenous (BD p-value = 0.308). For this reason, it was expressed that there was no interaction between the age and GSTT1 genotype in stomach cancer. MH OR, 0.93, was between 0.927 (1.03 – 0.103) and 1.333 (1.03 + 0.103). For this reason, age was not confounder.

After the association between GSTM1 and GSTT1 genotypes and some risk factors as gender and age were investigated, the interaction and confounding between these genotypes and smoking status in stomach cancer were examined by stratification in Tables 5.16 and 5.17, respectively.

Table 5.16 Association of GSTM1 genotype and stomach cancer stratified by smoking

Smoking Status	GSTM1	Case n (%)	Control n (%)	p-value [b]	OR	OR CI (95%)
Smoker	**Null**	29 (64.4)	19 (44.2)	0.056	2.29	0.97-5.39
	Normal [a]	16 (35.6)	24 (55.8)		1.00	
	Total	45 (100)	43 (100)			
Non-Smoker	**Null**	17 (50.0)	27 (48.2)	0.869	1.07	0.46-2.52
	Normal [a]	17 (50.0)	29 (51.8)		1.00	
	Total	34 (100)	56 (100)			
Crude OR For GSTM1					1.61	0.88-2.92
BD p-value				0.219		
MH OR					1.56	
MH CI (95%)						0.86-2.85
a: Referent						
b: It is obtained from χ^2 test						

The risk of stomach cancer related to GSTM1 genotype was examined by stratification of smoking. There was no association between GSTM1 genotype and stomach cancer in smoker (p-value = 0.056) or non-smoker individuals (p-value = 0.869). The risk of being stomach cancer for smokers who have GSTM1 null genotype was 2.29 times more than the smokers who have GSTM1 normal genotype (OR = 2.29, 95% CI = 0.97-5.39). In non-smokers, the null genotype did not increase the risk of being stomach cancer (OR = 1.07, 95% CI = 0.46-2.52, p-value = 0.869). According to BD p-value, strata for smoking were homogenous (BD p-value= 0.219). Although this result suggested an interaction effect the BD p-value was not significant (BD p-value = 0.219). Therefore, it was expressed that there was no statistically significant interaction between smoking and GSTM1 genotype in

stomach cancer. MH OR, 1.56, was between 1.45 (1.61 - 0.161) and 1.77 (1.61 + 0.161). For this reason, smoking was not confounder.

The same prosedure was applied for GSTT1 genotype in Table 5.17.

Table 5.17 Association of GSTT1 genotype and stomach cancer stratified by smoking

Smoking Status	GSTT1	Case n (%)	Control n (%)	p-value [b]	OR	OR CI (95%)
Smoker	**Null**	8 (17.8)	7 (16.3)	0.852	1.11	0.36-3.38
	Normal [a]	37 (82.2)	36 (83.7)		1.00	
	Total	45 (100)	43 (100)			
Non-Smoker	**Null**	10 (29.4)	15 (26.8)	0.787	1.14	0.44-2.93
	Normal [a]	24 (70.6)	41 (73.2)		1.00	
	Total	34 (100)	56 (100)			
Crude OR For GSTT1					1.03	0.51-2.10
BD p-value				0.974		
MH OR					1.13	
MH CI (95%)						0.55-2.32
a: Referent						
b: It is obtained from χ^2 test						

The risk of stomach cancer related to GSTT1 genotype was examined by stratification of smoking. There were no association between GSTT1 genotype and stomach cancer for smoker (p-value = 0.852) or non-smoker individuals (p-value = 0.787). According to BD p-value, strata for smoking were homogenous (BD p-value= 0.974). For this reason, it was expressed that there was no interaction between smoking and GSTT1 genotype. MH OR, 1.13, was less than 1.133 (1.03 + 0.103). Thus smoking was a confounder.

After the association, interaction and confounding effects between GSTM1 and GSTT1 genotypes and smoking status were investigated, smoking status was investigated with an another form as pack-years in Tables 5.18 and 5.19, respectively.

Table 5.18 Association of GSTM1 genotype and stomach cancer stratified by pack-years status

Pack-Years (PY)	GSTM1	Case n (%)	Control n (%)	p-value [b]	OR	OR CI (95%)
PY > 20	**Null**	21 (65.6)	12 (66.7)	0.941	0.95	0.28-3.24
	Normal [a]	11 (34.4)	6 (33.3)		1.00	
	Total	32 (100)	18 (100)			
PY ≤ 20	**Null**	8 (61.5)	7 (28.0)	0.045	4.11	1.00-16.99
	Normal [a]	5 (38.5)	18 (72.0)		1.00	
	Total	13 (100)	25 (100)			
Crude OR For GSTM1					1.61	0.88-2.92
BD p-value				0.123		
MH OR					1.772	
MH CI (95%)						0.73-4.32

a: Referent

b: It is obtained from χ^2 test

The risk of stomach cancer related to GSTM1 genotype was examined by stratification of pack-years status. After grouping according to pack-years status, GSTM1 null genotype was accociated with an increased stomach cancer risk for pack-years ≤ 20 status (OR = 4.11, 95% CI = 1.00-16.99). On the other hand, the existence of GSTM1 null genotype for pack-years > 20 status was protective effect, but the effect was not significant (OR = 0.95, 95% CI = 0.28-3.24). This situation was an unexpected and important result. According to BD p-value, strata for pack-years status were homogenous (BD p-value = 0.123). For this reason, it was expressed that there was no statistically significant interaction between the stomach cancer and GSTM1 genotype according to pack-years status. But, we can say that pack-years status may create differences associated with GSTM1 genotype and stomach cancer. Pack-years status may play an important role to find differences associated with GSTM1 genotype and stomach cancer. Because, the risk of being stomach cancer for smokers who have GSTM1 null genotype was 4.11 times more than the smokers who have GSTM1 normal genotype for pack-years ≤ 20, but the other status (pack-years > 20) was protective effect. For this reason, we may suspect

from the interaction between GSTM1 genotype and pack-years status in stomach cancer.

The stratification of pack-years status for GSTT1 genotype was given in Table 5.19.

Table 5.19 Association of GSTT1 genotype and stomach cancer stratified by pack-years status

Pack-Years (PY)	GSTT1	Case n (%)	Control n (%)	p-value [b]	OR	OR CI (95%)
PY > 20	**Null**	4 (12.5)	5 (27.8)	0.177	0.37	0.08-1.62
	Normal [a]	28 (87.5)	13 (72.2)		1.00	
	Total	32 (100)	18 (100)			
PY ≤ 20	**Null**	4 (30.8)	2 (8.0)	0.068	5.11	0.79-32.97
	Normal [a]	9 (69.2)	23 (92.0)		1.00	
	Total	13 (100)	25 (100)			
Crude OR For GSTT1					1.03	0.51-2.10
BD p-value				0.024		
MH OR					1.06	
MH CI (95%)						0.36-3.08

a: Referent

b: It is obtained from χ^2 test

The risk of stomach cancer related to GSTT1 genotype was examined by stratification of pack-years status. Results in Table 5.19 may be interpreted as Table 5.18. After grouping according to pack-years status, GSTM1 null genotype was accociated with an increased stomach cancer risk for pack-years ≤ 20 status, but this risk was not significant (OR = 5.11, 95% CI = 0.79-32.97). On the other hand, the existence of GSTM1 null genotype for pack-years > 20 status was protective effect, but effect was not significant (OR = 0.37, 95% CI = 0.08-1.62). According to BD p value, there was the interaction between GSTT1 genotype and pack-years status in stomach cancer (BD p-value = 0.024).

5.6.4 Biological Approach of Interaction

In a biological approach, interaction can be defined as follows: when the OR of outcome (stomach cancer) in the presence of exposure (risk factors: smoking, gender, age etc…) differs from the OR expected to result from their individual rates. The effect can be greater than what we whould expect (positive interaction) or less than what we whould expect (negative interaction). The problem is to determine what we whould expect to result from the individual effects of the exposures. With this definition, interaction was investigated in summary statistics as OR values. Firstly, interaction between the GSTM1 genotype and environmental factor as smoking status for stomach cancer was given in Table 5.20.

Table 5.20 OR values of stomach cancer according to presence or absence of two exposures: smoking and GSTM1

	GSTM1 normal	**GSTM1 null**
Non-Smoker	1.00	1.07
Smoker	1.14	2.60

In individuals with neither exposure, the OR was 1. In individuals exposed to factor GSTM1 only and not to factor smoking, the OR was 1.07. In individuals exposed to factor smoking only and not to factor GSTM1, the OR was 1.14. These were individual effects. In individuals exposed to both factors GSTM1 and smoking, the OR was 2.60 which was the combined effect. As seen in Table 5.20, combined effect 2.60 was higher than a multiplicative effect 1.22 (1.07×1.14) and this indicated the presence of interaction between GSTM1 genotype and smoking status in stomach cancer.

Interaction between GSTT1 genotype and smoking status in stomach cancer was investigated with the same method in Table 5.21.

Table 5.21 OR values of stomach cancer according to presence or absence of two exposures: smoking and GSTT1

	GSTT1 normal	GSTT1 null
Non-Smoker	1.00	1.14
Smoker	1.75	1.95

The combined effect 1.95 was lower than a multiplicative effect 1.99 (1.14×1.75) and this indicated the absence of interaction between GSTT1 genotype and smoking status in stomach cancer.

Secondly, interaction between the GSTM1 and GSTT1 genotypes and other environmental factors as gender and age for stomach cancer was given in Table 5.22.

Table 5.22 OR values of stomach cancer according to presence or absence of gender&age and GSTM1&GSTT1 exposures

	GSTM1		GSTT1	
Gender	**normal**	**null**	**normal**	**null**
Female	1.00	1.77	1.00	1.97
Male	3.34	3.59	3.69	1.89

	GSTM1		GSTT1	
Age	**normal**	**null**	**normal**	**null**
Young	1.00	1.67	1.00	0.63
Old	4.73	7.55	3.84	5.58

The combined effect 3.59 was lower than a multiplicative effect 5.91 and this indicated the absence of interaction between GSTM1 genotype and gender in stomach cancer. In addition, combined effect 1.89 was lower than a multiplicative effect 7.27. For this reason, this result indicated the absence of interaction between GSTT1 genotype and gender in stomach cancer. Combined effect 7.55 was lower than a multiplicative effect 7.90 and this indicated the absence of interaction between GSTM1 genotype and age in stomach cancer. On the other hand, combined effect 5.58 was higher than a multiplicative effect 2.42 and this indicated the presence of interaction between GSTT1 genotype and age in stomach cancer.

5.6.5 Multivariate Analysis

After these investigations about interaction and confounding effects in stratified analysis and in a biological approach, it was focused on the effect of GSTM1 and GSTT1 null genotypes in stomach cancer and logistic regression was used to assess the interaction effects between GSTM1 and GSTT1 null genotypes and other risk factors (smoking, age, gender) in stomach cancer. Departures from additive and multiplicative interaction effects between GSTM1, GSTT1 and risk factors for stomach cancer were evaluated. In addition, likelihood ratio test was used to detect interaction effect.

Logistic regression was used to control for potential confounders and to estimate crude and adjusted ORs and 95% CIs. Firstly, the association between GSTM1 and GSTT1 genotypes and risk of stomach cancer was estimated with ORs and their 95% CIs in logistic regression models in Table 5.23.

Table 5.23 Association between GSTM1&GSTT1 genotypes and stomach cancer in logistic regression model

						95% CI for OR	
Variable	**β**	**S. E.**	**Wald**	**p value**	**OR**	**Lower**	**Upper**
GSTM1	0.474	0.304	2.423	0.120	1.61	0.88	2.92
Constant	-0.474	0.222	4.565	0.033	0.62		
GSTT1	0.032	0.361	0.008	0.929	1.03	0.51	2.10
Constant	-0.233	0.171	1.847	0.174	0.79		

Reference categories: normal GSTM1 and normal GSTT1 genotypes

OR values for GSTM1 and GSTT1 genotypes were same results in Table 5.5. Using normal GSTM1 and GSTT1 genotypes as referent, values of crude ORs for these genotypes were 1.61 and 1.03, respectively. Logistic regression analysis did not show any association between GSTT1 null genotype and stomach cancer risk (p-value = 0.929, OR = 1.03, 95% CI: 0.51-2.10). GSTM1 null genotype showed a slightly increase risk for stomach cancer (p-value = 0.120, OR = 1.61, 95% CI: 0.88-2.92), but this was not statistically significant considering CIs for ORs included 1.

In 5.5.3 section, we found that age and gender as confounder. After controlling age and gender, the adjusted ORs and 95% CIs for GSTM1 and GSTT1 were calculated in logistic regression models. Exposure variables were age, gender and GSTM1&GSTT1 in these models.

Table 5.24 Logistic regression analysis in adjusted ORs of stomach cancers with GSTM1 genotype, gender and age

						95% CI for OR	
Variable	β	**S. E.**	**Wald**	**p value**	**OR**	**Lower**	**Upper**
GSTM1	0.383	0.334	1.314	0.252	1.47	0.76	2.82
Gender	0724	0.335	4.684	0.030	2.06	1.07	3.98
Age	1.418	0.337	17.731	0.000	4.13	2.13	7.99
Constant	-1.352	0.308	19.248	0.000	0.26		

Reference categories: normal GSTM1 genotype

Logistic regression model included GSTM1 (normal versus null), gender (female versus male) and age (young versus old) variables. After controlling for gender and age, the adjusted OR for GSTM1 was 1.47 which was not significant considering the 95 % CI (p value = 0.252, 95 % CI: 0.76-2.82).

The same procedure was applied for GSTT1 genotype below:

Table 5.25 Logistic regression analysis in adjusted ORs of stomach cancers with GSTT1 genotype, gender and age

						95% CI for OR	
Variable	β	**S. E.**	**Wald**	**p value**	**OR**	**Lower**	**Upper**
GSTT1	-0.050	0.390	0.016	0.899	0.95	0.44	2.05
Gender	0.784	0.330	5.635	0.018	2.20	1.15	4.18
Age	1.407	0.336	17.568	0.000	4.08	2.11	7.88
Constant	-1.162	0.273	18.115	0.000	0.31		

Reference categories: normal GSTT1 genotype

After controlling for gender and age, the adjusted OR for GSTT1 was 0.95 which was not significant considering the 95 % CI (p value = 0.899, 95 % CI: 0.44-2.05).

After controlling age, gender and smoking, the adjusted ORs and 95% CIs for GSTM1 and GSTT1 were calculated in logistic regression models. Exposure variables were age, gender, smoking and GSTM1&GSTT1 in these models. These were shown in Table 5.26 and 5.27, respectively.

Table 5.26 Logistic regression analysis in adjusted ORs of stomach cancers with GSTM1 genotype, gender, age and smoking

						95% CI for OR	
Variable	**β**	**S. E.**	**Wald**	**p value**	**OR**	**Lower**	**Upper**
GSTM1	0.375	0.337	1.240	0.265	1.45	0.75	2.81
Gender	0.602	0.346	3.032	0.082	1.83	0.93	3.60
Age	1.482	0.344	18.533	0.000	4.40	2.24	8.64
Smoking	0.515	0.344	2.238	0.135	1.67	0.85	3.28
Constant	-1.569	0.346	20.562	0.000	0.21		
-2 Log-likelihood = 212.134							
Reference categories: normal GSTM1 genotype							

Logistic regression model included GSTM1 (normal versus null), gender (female versus male), age (young versus old) and smoking (non-smoker versus smoker) variables. After controlling for gender, age and smoking, the adjusted OR for GSTM1 was 1.45 which was not significant considering the 95 % CI (p value = 0.265, 95 % CI: 0.75-2.81). Logistic regression analysis revealed age to be associated with increase OR value for stomach cancer (p-value = 0.000, OR = 4.40). 95 % CI for OR value did not include 1 (95 % CI: 2.24-8.64).

Logistic regression analysis in adjusted ORs of stomach cancers with GSTT1 genotype, gender, age and smoking is given below.

Table 5.27 Logistic regression analysis in adjusted ORs of stomach cancers with GSTT1 genotype, gender, age and smoking

						95% CI for OR	
Variable	**β**	**S. E.**	**Wald**	**p value**	**OR**	**Lower**	**Upper**
GSTT1	0.025	0.396	0.004	0.949	1.03	0.47	2.23
Gender	0.662	0.341	3.781	0.052	1.94	0.99	3.78
Age	1.467	0.343	18.354	0.000	4.34	2.22	8.49
Smoking	0.523	0.345	2.299	0.129	1.69	0.86	3.32
Constant	-1.407	0.324	18.880	0.000	0.24		
-2 Log-likelihood = 213.373							
Reference categories: normal GSTT1 genotype							

As seen in Table 5.27, after controlling for gender, age and smoking, the adjusted OR for GSTT1 was 1.03 which was not significant considering the 95% CI (p value = 0.949, 95% CI: 0.47-2.23). In addition, logistic regression analysis showed that age was associated with increase OR value for stomach cancer (p-value = 0.000, OR = 4.34).

Interactions between GSTM1 and GSTT1 genotypes and smoking were assessed by likelihood ratio tests comparing multivariate logistic regression models, with and without interaction terms in the presence of individuals genotypes effects. After controlling age, gender and smoking, the adjusted ORs and 95% CIs for GSTM1 and GSTT1 were found in logistic regression models in Tables 5.28 and 5.29, respectively.

Table 5.28 Logistic regression analysis for interaction between GSTM1 and smoking by gender and age

						95% CI for OR	
Variable	β	**S. E.**	**Wald**	**p value**	**OR**	**Lower**	**Upper**
GSTM1	0.110	0.470	0.055	0.814	1.12	0.44	2.81
Gender	0.588	0.347	2.879	0.090	1.80	0.91	3.55
Age	1.468	0.345	18.156	0.000	4.34	2.21	8.53
Smoking	0.235	0.489	0.230	0.631	1.26	0.48	3.30
GSTM1 × Smoking	0.536	0.668	0.644	0.422	1.71	0.46	6.33
Constant	-1.426	0.384	13.767	0.000	0.24		
-2 Log-likelihood = 211.490							
Reference categories: normal GSTM1 genotype							

The adjusted ORs for GSTM1 and GSTM1 × Smoking variables were 1.12 and 1.71, respectively. 95% CIs for ORs values of these variables included 1, respectively (p value = 0.814, 95% CI: 0.44-2.81; p value = 1.71, 95% CI: 0.46-6.33). For this reason, these variables were not statistically significant. In addition, the likelihood ratio test statistic (G) for the hypothesis that the slope coefficient is zero was obtained as minus twice the difference between the log-likelihoods for all variables in the model and the model containing all variables except for the variable GSTM1 × Smoking. Under the null hypothesis, G value followed the chi-square distribution with 1 degrees of freedom. This was denoted by $(\nu_{\text{full}} - \nu_{\text{reduced}}) = 6 - 5 = 1$. The likelihood ratio test for the difference between the models in Tables 5.26 and 5.28 (a test for the significance of GSTM1 × Smoking) yielded a value of $G = [212.134 - 211.490] = 0.644$. Comparing this value to the chi-square distribution with 1 degrees of freedom yielded a value of 3.84 $(\chi^2_{1,0.95})$. Here, 0.644 was less than 3.84. For this reason, GSTM1 × Smoking variable was not significant in this model.

The same procedure was applied for GSTT1 genotype below:

Table 5.29 Logistic regression analysis for interaction between GSTT1 and smoking by gender and age

						95% CI for OR	
Variable	**β**	**S. E.**	**Wald**	**p value**	**OR**	**Lower**	**Upper**
GSTT1	-0.010	0.519	0.000	0.984	1.00	0.36	2.74
Gender	0.662	0.341	3.773	0.052	1.94	0.99	3.78
Age	1.469	0.343	18.359	0.000	4.34	2.22	8.50
Smoking	0.505	0.387	1.697	0.193	1.66	0.77	3.54
GSTT1 × Smoking	0.086	0.804	0.011	0.915	1.09	0.22	5.27
Constant	-1.397	0.336	17.241	0.000	0.25		
-2 Log-likelihood = 213.362							
Reference categories: normal GSTT1 genotype							

As seen in Table 5.29, the adjusted ORs for GSTT1 and GSTT1 × Smoking variables were 1.00 and 1.09, respectively. 95% CIs for ORs values of these variables included 1 value, respectively (p value = 0.984, 95% CI: 0.36-2.74; p value = 0.915, 95% CI: 0.22-5.27). For this reason, these variables were not statistically significant. In addition, the likelihood ratio test statistic (G) was used. Under the null hypothesis, G value followed the chi-square distribution with 1 degrees of freedom. The likelihood ratio test for the difference between the models in Tables 5.27 and 5.29 (a test for the significance of GSTT1 × Smoking) yielded a value of $G = [213.373 - 213.362] = 0.011$. Comparing this value to the chi-square distribution with 1 degrees of freedom yielded a value of 3.84 $(\chi^2_{1,0.95})$. Here, 0.011 was less than 3.84. For this reason, GSTT1 × Smoking variable was not significant in this model.

After analysis of these gene-environment interactions for GSTM1 × Smoking and GSTT1 × Smoking, the other gene-environment interactions were analyzed in logistic regression models. These models included GSTM1 (normal versus null), GSTT1 (normal versus null), gender (female versus male), age (young versus old), smoking (non-smoker versus smoker), and pack-year (less than 20 versus greater than 20) variables. These gene-environment interactions were given in Table 5.30.

Table 5.30 Some gene-environment ınteractions in logistic regression models

			95% CI for OR	
Gene*Environment	**p-value**	**OR**	**Lower**	**Upper**
GSTM1 × Gender	0.414	0.56	0.07	1.59
GSTM1 × Age	0.881	0.90	0.24	3.41
GSTM1 × Pack-Year	0.114	0.20	0.03	1.47
GSTT1 × Gender	0.172	0.34	0.07	1.59
GTTT1 × Age	0.181	3.11	0.59	16.45
GSTT1 × Pack-Year	0.130	0.14	0.01	1.79

Interaction and confounding effects were also investigated in stratified analysis, a biological approach and multivariate analysis. A summary of all analysis was shown in Table 5.31.

Table 5.31 Comparison of stratified analysis, biological approach and multivariate analysis

	Stratified Analysis		**Biological Approach**	**Multivariate Analysis**
	Interaction	**Confounding**	**Interaction**	**Interaction**
Smoking × Gender	No	Gender		
Smoking × Age	No	Age		
Pack-Year × Gender	No	Gender		
Pack-Year × Age	No	Age		
GSTM1 × Gender	No	Gender	No	No
GSTM1 × Age	No	No	No	No
GSTT1 × Gender	Yes	-	No	No
GSTT1 × Age	No	No	Yes	No
GSTM1 × Smoking	No	No	Yes	No
GSTM1 × Pack-Year	Yes	-		No
GSTT1 × Smoking	No	Smoking	No	No
GSTT1 × Pack-Year	Yes	-		No

CHAPTER SIX
CONCLUSION

The case-control study of 127 cases of stomach adenocarcinoma and 101 controls between 2006 and 2008 years in Turkey was designed to evaluate the effect on stomach cancer risk of GSTM1 and GSTT1 gene polymorphisms. In addition, interaction and confounding effects were analyzed for some risk factors in stomach cancer.

The stomach cancer depends on many factors. These are gender, age, diet status, tobacco use, family history, intake of some food types, country of origin, obesity and environmental exposure. In our study, risk factors for stomach cancer were investigated in six category as demographic factors (gender, age, place of residence and birth, education status, marital status, job status, health care insurance, socioeconomic status), medical factors (body mass index, weight loss, health problem, family cancer history, smoking and cumulative cigarette smoking (pack-years), drinking habits (alcohol consumption, tea, herbal tea, instant coffee, turkish coffee and soft drinks intakes), nutritional factors (salt, margarine, oil, butter, sugar, hot food, spicy food, red meat and poultry intakes), pathological factors (location of tumor, histological type) and genotype factors (GSTM1 and GSTT1 genotypes).

The increased consumption of vegetables and fresh fruit has been shown to reduce the risk of stomach cancer (Hansson et al., 1994; Terry et al., 1998), whereas high consumption of salt tends to increase the risk of stomach cancer (Lee et al., 1995; Nazario et al., 1993). Some studies have reported an increased risk of stomach cancer in populations who consumed less milk and dairy product (Hirayama, 1984) and meat (Jedrychowski et al., 1986) or more salted foods (Hu et al., 1988). Tobacco smoking has been considered a potential risk factor for stomach cancer. From the previous epidemiological studies, a risk of stomach cancer among smokers is increased compared with non-smokers (Hansson et al., 1994; Tredaniel et al., 1997; Dyke et al., 1992). Smoking also has been implicated in stomach cancer. In 2004, the Surgeon General issued a report linking smoking to a range of diseases, including

stomach cancer. Like many cancers, stomach cancer is most common in older people. Few cases occur below 50 years of age and the highest rates are in men and women over 70. Men are twice as likely to get stomach cancer as women (Yalçın et al., 2006).

In this study, a polymerase chain reaction (PCR) method was used to detect the presence or absence of the GSTM1 and GSTT1 genotypes. According to this method, genotype data were available for 79 cases and for 99 controls. Blood samples of some cases were not collected and some result did not come out properly from the laboratory which consisted of missing data for genoytpe. The small number of cases was a limitation to estimate OR's precisely. The other limitation of this study was the difference of age and gender in cases and controls. It was more difficult to find older controls or they were more reluctant to be interviewed. We did not used matching according to age or gender. In our study, stomach cancer cases were older than population controls. In addition, male proportions were higher than female proportions in stomach cancer cases. This situation created a potential confounding effect. To minimize the possible confounding effects of these differences, we controlled for age and gender in all of our analyses. We found that age and gender were confounder in stratified analyses. On the other hand, we found possible gene- environment interaction in stratified analyses and also from a biological approach.

The presence of GSTM1 and GSTT1 null genotypes varies in ethic groups. GSTM1 genotype was absent (null) in 35%–60% of individuals (Bell et al., 1993; Chenevix et al., 1995), and GSTT1 genotype was absent in 10%–65% of the human population (Nelsen et al., 1995; Chenevix et al., 1995). Lim et al. (1994) found that the prevelance of GSTM1 null genotype among healthy Chinese to be 49% in Hong Kong and 45% in Taiwan. Katoh et al. (1996) found that the prevelance of GSTM1 null genotype in cases and controls were 57% and 44% in Japanese population, respectively. Deakin et al. (1996) found that the prevelance of GSTM1 null genotype was 53% in cases and 55% in controls in English population. Setiawan et al. (2000) found that the prevelance of GSTM1 and GSTT1 null in controls to be 51% and 46%

in China, respectively. Setiawan et al. (2000) found that the prevelance of GSTM1 and GSTT1 null in cases to be 48% and 54%, respectively. Cai et al. (2001) found that GSTT1 null genotype in cases was 43% in China. Tamer et al. (2005) found that the prevelance of GSTM1 and GSTT1 null in controls to be 43% and 26%, respectively. In addition, Tamer et al. (2005) found that the prevelance of GSTM1 and GSTT1 null in cases to be 57% and 30%, respectively. Palli et al. (2005) found that the prevelance of GSTM1 and GSTT1 null in controls to be 50% and 17% in Italian population, respectively. In addition, Palli et al. (2005) found that the prevelance of GSTM1 and GSTT1 null in cases to be 51% and 23%, respectively. Boccia et al. (2007) found that the prevelance of GSTM1 and GSTT1 null in cases to be 56% and 37%; in controls to be 53%, and 22% in Italian population, respectively. Our results were similar to these results. According to our result, the prevalence of GSTM1 null genotype was 58.2% in cases, 46.5% in controls and the prevalence of GSTT1 null genotype was 22.8% in cases, 22.2% in controls in Turkish population.

Very few studies in the past have studied the associations between GSTM1 and GSTT1 null genotype and the risk of stomach cancer. Harada et al. (1992) showed an association between GSTM1 and stomach cancer in small sample size. Katoh et al. (1996) also showed a weak association between GSTM1, but not GSTT1, genotype and stomach cancer in Japanese population. Deakin et al. (1996) showed no association between GSTT1 and stomach cancer in English population. Setiawan et al. (2000) showed no association between GSTM1 and stomach cancer in China. But Setiawan et al. (2000) showed an association between GSTT1 and stomach cancer. Cai et al. (2001) found that GSTT1 gene deletion was not associated with stomach cancer but they observed evidence of a relatinship between null genotype of GSTM1 and risk of stomach cancer in China. Gao et al. (2002) showed no association between GSTT1 and stomach cancer in China. But they found an association between GSTM1 and stomach cancer. Tamer et al. (2005) showed an association between GSTM1 and stomach cancer, but not GSTT1. Palli et al. (2005) showed no association between GSTM1 and stomach cancer in Italian population. Boccia et al. (2007) found that GSTT1 polymorphisms appeared to modulate individuals susceptibility to stomach cancer in Italian population, particularly when combined

with cigarette smoke. In our study, we found no association for GSTT1, but slightly increased risk for GSTM1 althought not statistically significant in Turkish population.

In our study, risk factors for stomach cancer were investigated in different categories. For demographic factors, lots of cases were males (56.7%), whereas lots of controls were females (63.4%). Case group was older than control group. 60.6% of the case group was old and 75.2% of the control group was young. 66.9% of the cases were living in İzmir and 75.2% of the controls were living in İzmir. The proportion of "no education" status was higher among cases (17.3%) than those of the controls (5.0%) as expected. Individuals who had "no education" were 10.06 times more likely to have cancer than individuals who graduated from university. There were no statistically significant difference in birth place, residence place, marital status, insurance and income variables between two groups. For medical factors, the proportion of weight loss variable was significantly higher among cases (75.6%) than those of the controls (37.6%) as expected. There was no difference in family cancer history variable between two groups (p-value = 0.451). Smoking prevalence were 55.9% and 43.6% in cases and controls, respectively. The difference was not statistically significant in smoking variable between two groups (p-value = 0.064). However when this variable was investigated as pack-years of smoking, there was significant difference in pack-years variable between two groups (p-value = 0.001). Using cigarette more than 20 pack-years was increasing the risk 3.95 times for stomach cancer compared to less than 20 pack-years category. For drinking habits, there were significant differences in Turkish coffee intake and soft drink intake variables between two groups (p-value = 0.0001). This difference was probably due to digestive problems of the cases thus they were consuming these drinks less than the controls. For nutritional factors, there were significant differences in salt, margarine, butter, sugar, spicy food and poultry variables between two groups. For pathological factors, there was no association between GSTM1&GSTT1 genotypes and histological types of stomach cancer cases. In addition, there were no differences in GSTM1 and GSTT1 genotype frequencies in respect to the location of stomach cancer.

There was no association for GSTT1 genotype in cases and controls, but there was an increased risk in GSTM1 null patients although a significant association was not found (OR = 1.61).

According to univariate and stratified analyses, this study showed that smoking is a significant risk factor for stomach cancer in this Turkish population. The significant relationship found in the univariate analysis for Turkish coffee and soft drinks should be regarded with caution. There is no reported association in the literature for these drinks. It is probable that patients with cancer did not consume these drinks after they have been diagnosed, but reported this as their usual behaviour. To minimize the possible confounding effects caused by difference in age and gender among cases and controls, they are controlled in the multivariate analysis.

According to stratified analysis, we found that gender and age were confounder. There was no interaction between GSTT1 genotype and age in stratified analysis although there was an interaction for a biological approach. This situation was also the same for GSTM1 genotype and smoking in stratified analysis. On the other hand, there was an interaction between GSTT1 genotype and gender in stratified analysis although there was no interaction for a biological approach. In addition, there were no interactions in all multivariate analysis.

In conclusion, the effect of smoking on stomach cancer was present after adjusting for age, gender, GSTM1&GSTT1. There were 1.67 and 1.69 fold increase of cancer in smokers although a significant association was not found.

In summary, this study revealed that GSTM1 polymorphism in stomach cancer has a potential role for interaction between this polymorphism and smoking from a biological approach. In addition, our data suggested an increased risk for GSTM1 genotype although a significant association was not found (OR = 1.61, 95% CI: 0.88-2.92). On the other hand, there was no association for GSTT1 genotype in cases and controls.

REFERENCES

Ahlbom, A., & Alfredsson, L. (2005). Interaction: A word with two meanings creates confusion. *Eur J Epidemiol,* 20, 563-564.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: John Wiley and Sons.

Assmann, S. F., Hosmer, D., & Lemeshow, S. (1996). Confidence intervals for measures of interaction. *Epidemiology,* 7, 286-290.

Bell, D. A., Taylor, J. A., Paulson, D. F., Robertson, C. N., Mohler, J. L., & Lucier, G. W. (1993). Genetic risk and carcinogen exposure: A common inherited defect of the carcinogen-metabolism gene Glutathione S-transferase M1 (GSTM1) that increases susceptibility to bladder cancer. *J. Natl. Cancer Inst.,* 85, 1159–1164.

Bewich, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. *Critical Care*, 9, 112-118.

Bhopal R. (2007). *Variation: role of error, bias and confounding. Lecture notes.* Division of Community Health Sciences, University of Edinburgh, Retrieved January 20, 2007.

Boccia, S., Tabatabaei, F, Persiani, R. Gianfagna, F., Rausei, S., Arzani, D., Greca, A., D'Ugo, D., Torre, G., Duijn, C., & Ricciardi, G. (2007). Polymorphisms in metabolic genes, their combination and interaction with tobacco smoke and alcohol consumption and risk of gastric cancer: A case control study in an Italian population. *BMC Cancer*, 7, 206-213.

Boivin, J. F., & Wacholder, S. (1985). Conditions for confounding of the risk ratio and the odds ratio. *Am J Epidemiol*, 121, 152–158.

Cai, L., Yu, S. Z., & Zhang, Z. F. (2001). Glutathione S-transferases M1, T1 genotypes and the risk gastric cancer: A case-control study. *World Journal of Gastroenterology,* 7 (4), 506-509.

Chen W. J. (2007). *Principles of epidemiology, case-control studies (II). Lecture Notes.* Institute of Epidemiology, College of Public Health, National Taiwan University, Retrieved March 2, 2007.

Chenevix, T. G., Young, J., Coggan, M., & Board, P. (1995). Glutathione S-transferase M1, and T1 polymorphisms: Susceptibility to colon cancer and age of onset. *Carcinogenesis*, 16, 1655–1657.

Chow, S. C., & Liu, J. (2004), *Design and Analysis of Clinical Trials: Concepts and Methodologies* (2nd ed.). John Wiley and Sons.

Cristensen, R. (1997). *Log-linear models and logistic regression* (2nd ed.). Springer-Verlag.

Deakin, M., Elder, J., Hendrickse, C., Peckham, D., Baldwin, D., Pantin, C., Wild, N., Leopard, P., Bell, D. A., Jones, P., Duncan, H., Brannigan, K., Alldersea, J., Fryer, A., & Strange, R. C. (1996). Glutathione S-transferase GSTT1 genotypes and susceptibility to cancer: studies of interactions with GSTM1 in lung, oral, gastric and colorectal cancers. *Carcinogenesis*, 17, 881–884.

Dorak, M. T. (2006). *Bias, confounding and fallacies in epidemiology. Lecture Notes*. Newcastle University, U.K., Retrieved December 11, 2006.

Dyke, G. W., Craven, J. L., Hall, R., & Garner, R. C. (1992). Smoking related DNA adducts in human gastric cancers. *Int. J. Cancer,* 52, 847–850.

Figueiras, A., Domenech, J. M., Cadarso, C. (1998). Regression models: Calculating the confidence interval of effects in the presence of interactions. *Statistics in Medicine*, 17, 2099-2105.

Gao, C., Takezaki, T., Wu, J., Li, Z., Liu, Y., Li, S., Ding, J., Su, P., Hu, X., Xu, T., Sugimura, H., & Tajima, K. (2002). GSTM1 and GSTT1 genotype, smoking, consumption of alcohol and tea and risk of esophageal and stomach cancers: A case control study of a high-incidence area in Jiangsu province, Chine. *Cancer Letters*, 188, 95-102.

Garcia, C. M., & Lubin, J. H. (1999). Power and sample size calculations in case-control studies of gene-environment interactions: comments on different approaches. *Am. J. Epidemiol,* 689–692.

Goldstein, A., Falk, R. T., Korczak, J. F., & Lubin, J. H. (1997). Detecting gene-environment interactions using a case-control design. *Genetic Epidemiol,* 14, 1085–1089.

Gonzalez, C. A., Sala, N, & Capella, G. (2002). Genetic suspectibility and gastric cancer risk. *Int. J. Cancer,* 100 249-260.

Gordis, L. (2004). *Epidemiology* (2nd ed.). Philadelphia: Elsevier Saunders.

Grayson, D. A. (1987). Confounding confounding. *Am J Epidemiol*, 126, 546–563.

Greenland, S. (1983). Tests for interaction in epidemiologic studies: A review and a study of power. *Statistics in Medicine*, 2, 243-251.

Greenland, S. (1993). Basic problems in interaction assessment. *Environmental Health Perspectives Supplements*, 101, 59-66.

Greenland, S., & Robins, J. (1985). Confounding and misclassification. *American Journal of Epidemiology*, 122 (3) 495-506.

Hansson, L. E., Baron, J., Nyren, O., Bergstorm, R., Wolk, A., & Adami, H. (1994). Tobacco, alcohol and the risk of gastric cancer: A population-based case control study in Sweden. *Int. J. Cancer*, 57, 26–31.

Harada, S., Misawa, S., Nakamura, T., Tanaka, N., Ueno, E., & Nyren, O. (1992). Detection of GST1 gene deletion by the polymerase chain reaction and its possible correlation with stomach cancer in Japanese. *Hum. Genet.,* 90, 62–64.

Hennekens, C. & Buring, J. (1987). *Epidemiology in medicine.* Boston: Little, Brown and Company.

Hirayama T. (1984). Epidemiology of stomach cancer in Japan with special reference to the strategy for the primary prevention. *Jpn J Clinical Oncology*, 14, 159-168.

Hosmer, D., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, USA: John Wiley and Sons.

Hu, J., Zhang, S., Ermin, J. et al. (1988). Diet and cancer of the stomach: A case control study in China. *Int J. Cancer*, 41, 331-335.

Huang, W., Chow, W., Rothman, N., Lissowska, J., Llaca, V., Yeager, M., Zatonski, W, & Hayes, R. (2005). Selected DNA repair polymorphisms and gastric cancer in Poland. *Carcinogenesis*, 28 (8), 1354-1359.

Hurtig, A. K. (2006). *Confounding*. *Lecture Notes*. Retrieved November 15, 2006.

Jaccard, J. (2001). *Interaction effects in logistic regression*. CA: Sage Publications.

Jedrychowski, W., Wahrendorf, J., Popiela, T., & Rachtan, J. (1986). A case control study of dietary factors and stomach cancer risk in Poland. *Int J Cancer,* 3 (7), 837-842.

Jepsen, P., Johnsen, S. P., Gillman, M. W., & Sørensen, H. T. (2004). Interpretation of observational studies. *Heart BMJ,* 90, 956-960.

Katoh, T., Nagata, N., Kuroda, Y., Itoh, H., Kawahara, A., Kuroki, N., Ookuma, R., & Bell, D. A. (1996). Glutathione S-transferase M1 (GSTM1), and T1 (GSTT1) genetic polymorphism and susceptibility to gastric and colorectal adenocarcinoma. *Carcinogenesis,* 17, 1855–1859.

Katoh, T., Inatomi, H., Nagaoka, A., & Sugita, A. (1995). Cytochrome P4501A1 gene polymorphism and homozygous deletion of the glutathione S-transferase M1 gene in urothelial cancer patients. *Carcinogenesis,* 16, 655–657.

Kalilani, L, & Atashili, J. (2006). Measuring additive interaction using odds ratios. *Epidemiologic Perspectives & Innovations*, 3.

Kleinbaum, D. G. (1994). Modeling strategy for assessing interaction and confounding. In *Logistic Regression* (191-266). New York: Spring-Verlag.

Kleinbaum, D. G., Kupper, L. L., Muller, K. E. (1998). Confounding and interaction in regression. In *Applied Regression Analysis and Other Multivariable Methods* (2nd ed) (163-180). CA: Duxbury Pres.

Kopman, J. S., (1981). Interaction between discrete causes. *Am J Epidemiol,* 113, 716-724.

Lan, Q., Chow, W. H., Lissowska, J., Hein, D. W., Buetow, K., Engel, L. S., Ji, B., Zatonski, W., Rothman, N. (2001). Glutathione S-transferase genotypes and stomach cancer in a population-based case-control study in Warsaw, Poland. *Pharmacogenetics,* 11, 655–661.

Le, C. T. (2003). *Introductory biostatistics.* New Jersey, USA: John Wiley and Sons.

Lee, J. K., Park, B. J., Yoo, K. Y., & Ahn, Y. O. (1995). Dietary factors and stomach cancer: A case-control study in Korea. *Int. J. Epidemiol*, 24, 33–41.

Menard S. (1996). *Applied logistic regression analysis.* CA: Sage.

McNamee R. (2003). Confounding and confounders. *Occup. Environ. Med*, 60, 227-234.

McNamee R. (2005). Regression modeling and other methods to control confounding. *Occup. Environ. Med*, 62, 500-506.

Miettinen, O. S, & Cook, E. F. (1981). Confounding: Essence and detection. *Am J Epidemiol*, 114, 593–603.

Nazario, C. M., Szklo, M., Diamond, E., Roman-Franco, A., Climent, C., Suarez, E., & Conde, J. (1993). Salt and gastric cancer: A case-control study in Puerto Rico. *Int. J. Epidemiol*, 22, 790–797.

Nelson, H. H., Wiencke, J. K., Christiani, D. C., Cheng, T. J., Zuo, Z. F., Schwartz, B. S., Lee, B. K., Spitz, M. R., Wang, M., Xu, X. P., & Kelsey, K. T. (1995). Ethnic differences in the prevalence of the homozygous deleted genotype of Glutathione S-transferase M. *Carcinogenesis,* 16, 1243–1245.

NCSS Inc. (2009). *Number Cruncher Statistical Systems*. Retrieved 2004. http://www.ncss.com

Neter, Kutner, Nachtsheim, & Wasserman (1996). *Applied linear regression models* (3th ed.). USA: Irwin.

Palli, D., Saieva, C., Gemma, S., Masala, G., Miguel, M. J. G., Luzzi, I., D'Errico, M., Matullo, G., Ozzola, G., Manetti, R., Nesi, G., Sera, F., Zanna, I., Dogliotti, E., & Testai, E. (2005). GSTT1 and GSTM1 gene polymorphisms and gastric cancer in a high-risk Italian population. *Int. J. Cancer,* 115, 284–289.

Pampel, F. C., (2000). *Logistic regression: A primer*. CA: Sage Publications.

Pinarbasi, H., Silig, Y., Cetinkaya, O., Seyfikli, Z, & Pinarbasi E. (2003). Strong association between the GSTM1-null genotype and lung cancer in a Turkish population. *Cancer Genetics and Cytogenetics,* 146, 125–129.

Preacher, K., (2004) *A Primer on Interaction Effects in Multiple Linear Regression*. North Carolina: Chapel Hill.

Rodriguez M. D., & Llorca J. (2004). Bias. *Epidemiol. Community Health*, 58, 635-641.

Rothman, K. J. (1974). Synergy and antagonism in cause-effect relationships. *Am J Epidemiol,* 99, 385-388.

Rothman, K. J. (1976). The estimation of synergy or antagonism. *Am J Epidemiol*, 103, 506-511.

Rothman, K. J., Greenland, S., & Walker, A. M. (1998). Concepts of interaction. *Am J Epidemiol*, 112, 467-470.

Rothman K. J., & Greenland S. (1998). *Modern Epidemilogy* (2nd ed.). Philadelphia: Lippincott Williams & Wilkins.

Rousseeuw, P. J., & Christmann, A. (2003). Robustness against separation and outliers in logistic regression. *Computational Statistics & Data Analysis*, 43, 315–332.

Royston, P., & Saurbrei, W. (2004). A new approach to modeling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Statistics in Medicine*, 23, 2509-2525.

Saadat, I, & Saadat, M. (2001). Glutathione S-transferase M1 and T1 null genotypes and the risk of gastric and colorectal cancers. Cancer Letters,169, 21–26.

Schneider, D. (2007). Bias, confounding and the role of chance: Principles of epidemiology. *Lecture Notes*. 2 March 2007.

Schneider, J., Bernges, U., Philip, M., & Woitowitz, H. J. (2006). GSTM1, GSTT1, and GSTP1 polymorphism and lung cancer risk in relation to tobacco smoking. *Cancer Letters,* 208, 65–74.

Setiawan, V., Zhang, Z., Yu, G., Li, Y., Lu, M., Tsai, C., Cordova, D., Wang, M., Guo, C, Yu, S., & Kurtz, R. (2000). GSTT1 and GSTM1 genotypes and the risk of gastric cancer: A case control study in Chinese population. *Cancer Epidemiology*, 9, 73-80.

Skrondal, A. (2003). Interaction as departure from additivity in case-control studies. *Am J Epidemiol*, 158, 251-258.

Smith, P. G., & Day, N. E. (1984). The design of case-control studies: the infuence of confounding and interaction effects. *International of Journal of Epidemiology*, 13, 356-365.

Solis, J. (1998). A closer look at confounding. *Family Medicine,* 38, 584-588.

SPSS Inc. (2009). *Statistical package for the* social *sciences.* Retrieved 2004. http://www.spss.com

Tamer, L., Ateş, N., Ateş, C., Ercan, B., Eliipek, T., Yıldırım, H., Çamdeviren, H., Atik, U., & Aydin, S. (2005). GSTM1, T1 and P1 genetic polymorphisms, cigarette smoking and gastric cancer risk. *Cell Biochemistry and Function*, 23, 267-272.

Terry, P., Nyren, O., & Yuen, J. (1998). Protective effect of fruits and vegetables on stomach cancer: A cohort of Swedish twins. *Int. J. Cancer*, 76, 35–37.

Thompson, W. D. (1991). Effect modification and the limits of biological inference from epidemiologic data. *J Clin Epidemiol,* 44, 221-232.

Tredaniel, J., Boffetta, P., Buiatti, E., Saracci, R., & Hirsch, A. (1997). Tobacco smoking and gastric cancer: Review and meta analysis. *Int. J. Cancer*, 72, 565–573.

Yalcin, B., Aydin, F., Zengin, N., Ilhan, M., Isikdogan, A., Aykan, F., Demir, G., Celik, I., Turhal, S., Icli F., & Akbulut. H. (2006). The clinicopathological and socioeconomic features of patients with gastric cancer in Turkey. *Journal of Clinical Oncology*, 24.

Ylöstalo, P. V., & Knuuttila, M. L. (2006), Confounding and effect modification: possible explanation for variation in the results on the association between oral and systemic diseases. *J. Clin. Periodontology,* 33, 104-108.

World Cancer Research Fund / American Institute for Cancer Research. Food, Nutritution, Physical Activity, and the Prevention of Cancer: a Global Perspective, Washington DC:AICR, 2007.

Association for International Cancer Research. (2009). Stomach Cancer FAQs. Retrieved November 15, 2008, from http://www.aicr.org.uk/StomachCancerFAQs.stm.

Helicobacter Foundation. (2009). Disease Stomach. Retrieved December 10, 2008, from http://www.helico.com/disease_stomach.html.

Appendix 1: Sample Size Calculation According to Some Characteristics

Confidence Interval	Power	Control/Case Proportion	Prevelance (%)	Odds Ratio	Control	Case	Total
95	80	01:01	50	3.71	49	49	98
95	90	01:01	50	3.71	63	63	126
95	80	02:01	50	3.71	74	37	111
95	**80**	**01:01**	**50**	**2.35**	**101**	**101**	**202**
95	90	01:01	50	2.35	132	132	264
95	80	02:01	50	2.35	152	76	228
95	80	01:01	40	3.71	45	45	90
95	90	01:01	40	3.71	57	57	114
95	80	02:01	40	3.71	68	34	102
95	**80**	**01:01**	**40**	**2.35**	**97**	**97**	**194**
95	90	01:01	40	2.35	126	126	252
95	80	02:01	40	2.35	144	72	216
95	80	01:01	30	3.71	44	44	88
95	90	01:01	30	3.71	57	57	114
95	80	02:01	30	3.71	66	33	99
95	**80**	**01:01**	**30**	**2.35**	**101**	**101**	**202**
95	90	01:01	30	2.35	132	132	264
95	80	02:01	30	2.35	150	75	225
95	80	01:01	20	3.71	50	50	100
95	90	01:01	20	3.71	65	65	130
95	80	02:01	20	3.71	74	37	111
95	**80**	**01:01**	**20**	**2.35**	**121**	**121**	**242**
95	90	01:01	20	2.35	158	158	316
95	80	02:01	20	2.35	176	88	264

Appendix 2: The Epidemiological Questionnaire About Determining the Association Between Life Styles and Personal Habits

This epidemiological questionnaire is performed by Dokuz Eylül University Arts and Sciences Faculty Department of Statistics and Dokuz Eylül University Faculty of Medicine Department of Public Health. Finding the association between life styles and personal habits of participants is the aim of this questionnaire. Your records will be kept confidential. Information obtained from the questionnaire will be only used for scientific purposes.

Participants ID Number: …… Date: ……

Family Name/Surname ……

Phone Number (Home/Mobile/Job): …….

Address: ……..

Date of Birth (Month/Day/Year): ……...

Place of Birth (Country/Town): ……..

Living Place(Country/Town): ……

1. In total, how many days/months/years have you lived in İzmir?: ……..

2. Gender: □ Male □ Female

3. Your current height: ……cms

4. Your current weight: ……kilos

5. How many kilos have you lost in last six month?: ……. kilos

6. What is the highest level or year of school you have completed?

□ No Education

□ Education

□ Primary School

□ Middle School

□ High School

□ University

7. What is your marital status?

□ Single

□ Married

□ Divorced

□ Widowed

□ Separated

8. What is your job status?: ……

(If you are retired, what was your job?)

9. What is your health insurance?

□ Absent

□ Yeşil Card

□ SSK

□ BAĞKUR

□ EMEKLİ SANDIĞI

□ Private Insurance

10. In total, how much Money is your family income (TL)?

□ 0-500 □ 500-1000 □ 1000-2000 □ 2000-3000 □ 3000 and above

11. Do you have an important health problem?

□ Yes: ……..

□ No

12. Do you have cancer in your family ?

□ Yes: ……..

□ Mother/Father/Brother&Sister/Anyone else: …….

□ No

13. Have you ever smoked cigarette?

□ Yes (Go to 16. question)

□ Quit (Go to 14. question)

□ No (Go to 18. question)

14. How old were you when you quit?: ……..

15. How many did you smoke?: ……./per day/per week/per month

(Go to 18. question)

16. How old were you when you started?: ……..

17. How many did you smoke?: ……./per day/per week/per month

18. Nowadays, Have you ever used alcohol?

□ Yes: ……..… (Beer, Wine, Rakı, Others…) ……./per day/per week/per month

□ Quit

□ No

19. Do you regularly consume the following drinks?

Tea	□ Yes	□ No
Herbal Tea	□ Yes	□ No
Instant Coffee	□ Yes	□ No
Turkish Coffee	□ Yes	□ No
Soft Drinks	□ Yes	□ No

20. Do you avoid eating any of the food items below?

Salt	□ Yes	□ No
Margarine	□ Yes	□ No
Oil	□ Yes	□ No
Butter	□ Yes	□ No
Sugar	□ Yes	□ No
Hot Food	□ Yes	□ No
Spicy Food	□ Yes	□ No
Red Meat	□ Yes	□ No
Poultry	□ Yes	□ No

Thank you for your co-operation.

Prof. Dr. Gül ERGÖR

Research Ass. Özgül VUPA