SURVIVAL MODELLING APPROACH TO TIME TO FIRST CLAIM AND ACTUARIAL PREMIUM CALCULATION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

DERYA AKBULUT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF ACTUARIAL SCIENCES

MARCH 2011

Approval of the thesis:

SURVIVAL MODELLING APPROACH TO TIME TO FIRST CLAIM AND ACTUARIAL PREMIUM CALCULATION

Submitted by **DERYA AKBULUT** in partial fulfillment of the requirements for the degree of Master of Science **in Department of Actuarial Sciences**, **Middle East Technical University** by,

Prof.Dr.Ersan Akyıldız
Director, Graduate School of **Applied Mathematics** ____________

Assoc. Prof. Dr. A.Sevtap Kestel
Head of Department, **Actuarial Sciences** ____________

Assoc. Prof. Dr. A.Sevtap Kestel
Supervisor, **Department of Actuarial Sciences, METU** ____________

Examining Committee Members:

Prof. Dr. Ömer L. Gebizlioğlu
Department of Statistics, Ankara University ____________

Assoc. Prof. Dr. A. Sevtap Kestel
Department of Actuarial Sciences, METU ____________

Assoc. Prof. Dr. Sarp Üner
Institute of Public Health, Hacettepe University ____________

Assist. Prof. Dr. Berna Burçak Başbuğ Erkan
Department of Statistics, METU ____________

Assist. Prof. Dr. Zeynep Kalaylıoğlu
Department of Statistics, METU ____________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : DERYA AKBULUT

Signature :

ABSTRACT

SURVIVAL MODELLING APPROACH TO TIME TO FIRST CLAIM AND ACTUARIAL PREMIUM CALCULATION

Akbulut, Derya
M.Sc., Department of Actuarial Sciences
Supervisor : Assoc. Prof. Dr. A. Sevtap Kestel

March 2011, 105 pages

Health problems of the human beings in a society are one of the main components of the social security systems due to the dimension of the financial burden it might bring on individuals, employers, insurance companies and governments. Morbidity measures, such as incidence and prevalence of a specific disease in a certain population enable researchers to estimate for individuals the probability of being diagnosed or being prone to the diseases. This information is usually not tractable because of the non-availability of the convenient data or recordings for many countries as well as Turkey. Even if it is available, it is commonly limited with largely varying characteristics about the type and coverage of the diseases. In this regard, the pattern that a population follows for an acute disease may not be the same for chronic diseases. Having those indicators determined for a group of insureds will enable underwriters to have more profitable and economical premium calculation and precision on required reserve estimation.

Based on their characteristics such as acute or chronic behaviour, the gender, and the location of residency of people, the diseases show different behaviour on their occurrences. From the insurer's point of view, it is important to get to know how often the insureds visit a health institute and as a result of those visits, how large the provisions provided by insurance companies will be. Having a term like 12 months of validity on the health insurance, the occurrence of a claim due to any morbidity in the specified time span is uncertain to insurance company. Therefore, the time to first claim will be an important determinant on how many among a portfolio of N health insurance policies will cause a total claim amount of S during a year.

The aim of this thesis is to investigate and determine the effect of duration elapsed to first failure as the time to first claim on the actuarial evaluation of certain morbidities. For this purpose, two different morbidities have been taken into account in order to model the time to first failure within a policy year by using survival models and find the time to first claim for determined periods in a year and its impact on actuarial premium calculation.

For analyses and application purposes rheumatic heart disease (RHD) and poisoning morbidities are selected and the data of these morbidities based on some specifications, like gender, location and age are studied. Parametric and non-parametric methods are used to model the distribution of the time to first claim through distribution fitting techniques. Parametric techniques applied to the data do not give appropriate results. Therefore, non-parametric analysis namely Kaplan-Meier and life table methods are employed. The resulting survival tables for the periods of a week within a policy year are used to calculate the actuarial premium and its variance under the assumptions of individual and collective risk models.

A simulation work on claim amount and frequency distributions is performed to illustrate the application of the proposed study. The results are compared with respect to the gender, location, age, and type (acute or chronic) of the morbidities for a year.

Keywords: Survival models, morbidity, rheumatic heart disease, acute RHD, chronic RHD, poisoning, time to first failure, life table method, Kaplan-Meier method, actuarial premium

ÖZ

İLK HASARIN GERÇEKLEŞMESİNE KADAR GEÇEN SÜREYE YAŞAM MODELİ UYGULAMASI VE AKTÜERYAL PRİM HESAPLAMASI

Akbulut, Derya
Yüksek Lisans , Aktüerya Bilimleri Bölümü
Tez Yöneticisi : Doç. Dr. A. Sevtap Kestel

Mart 2011, 105 Sayfa

Bir toplumda yaşayan bireylerin sağlık problemleri, bireyler, çalışanlar, sigorta şirketleri ve devletler üzerinde oluşturabileceği finansal yük dolayısıyla sosyal güvenlik sistemlerinin en temel bileşenlerindendir. Bir hastalığa ilişkin insidans ve prevalans gibi morbidite ölçüleri araştırmacıların belirli bir toplumdaki hastalığa yakalanmış ya da yakalanmaya meyilli kişilerin oranlarını tahmin edebilmelerini sağlar. Türkiye de dahil olmak üzere birçok ülkede kayıtların elverişli ve kullanılabilir olmamasından dolayı bu bilgi kolay elde edilememektedir. Söz konusu olan bilgi işlenebilir olsa dahi, hastalığın türü ve kapsamı hakkında sınırlı bilgi verir. Bu sebeple, örneğin, akut hastalıklar görülen bir toplumun yapısı kronik hastalıklar gözlenen toplumların yapısıyla aynı değildir. Bir grup sigortalanan kişi için bu göstergelerin bilinmesi, aktüerlerin karlı ve iktisadi prim değerlemelerini ve gerekli rezerv hesabını doğru bir şekilde yapabilmelerini sağlar.

Akut ya da kronik hastalık olarak ortaya çıkabilen hastalıklar yaşa, cinsiyete ya da yaşanılan yere göre ortaya çıkma frekanslarında farklılık gösterir. Sigortacı gözüyle bakıldığında, sigortalanan bir kişinin ne sıklıkta bir sağlık kurumuna başvurduğu ve bu başvuruların sigorta şirketine maliyetinin ne olacağı önem kazanır. 12 aylık zaman periyodu için düzenlenen bir sağlık sigortası poliçesi satın aldığımızda, herhangi bir morbiditeye bağlı hasarın ortaya çıkma olasılığı sigorta şirketi için belirsizdir. Bu sebeple, bir yıl olarak belirlenen zaman dilimi içerisinde, N adet sağlık sigortasından oluşan bir portföydeki oluşacak toplam hasar miktarı olan S'yi belirlemede ilk hasar oluşuncaya kadar geçen süre belirleyici olacaktır.

Bu tezin amacı ilk hasar oluşana kadar geçen sürenin belirlenen hastalıkların aktüeryal değerlemesine etkisinin araştırılmasıdır. Bu amaçla, iki farklı morbidite seçilerek bir poliçe yılı içerisinde ilk hasar oluşana kadar geçen sürenin modellemesi yaşam modelleri kullanılarak yapılmıştır.

Analizlerin yapılması amacıyla romatizmal kalp hastalıkları (RHD) ve zehirlenmeler morbiditeleri seçilmiş ve cinsiyetin etkisi temel alınarak çalışmalar yapılmıştır. İlk hasar oluşana kadar geçen sürenin modellemesinde parametrik ve parametik olmayan yöntemler kullanılmıştır. Kaplan-Meier ve yaşam tablosu yöntemleri bu çalışmada kullanılan parametrik olmayan yöntemlerdir. Bir poliçe yılı içerisinde belirlenen haftalık zaman periyodlarına göre oluşturulan yaşam tabloları, bireysel ve toplu risk modelleri varsayımları altında aktüeryal prim ve varyans hesaplamasında kullanılmıştır.

Önerilen çalışmanın uygulamasını göstermek adına frekans ve hasar miktarı değişkenleri için benzetişim ile elde edilmiş veri seti kullanılmıştır. Sonuçlar cinsiyet, yaş, yaşanılan yer ve hastalık türüne (akut ya da kronik) göre bir poliçe yılı baz alınarak karşılaştırılmıştır.

Anahtar Kelimeler: Yaşam modelleri, romatizmal kalp hastalığı, akut RHD, kronik RHD, zehirlenmeler, ilk hasar oluşana kadar geçen süre, yaşam tablosu metodu, Kaplan-Meier metodu, aktüeryal prim

To my family

ACKNOWLEDGMENTS

I would like to thank everyone who has supported me in many aspects in the preparation of this thesis. First of all, I am deeply thankful to Assoc. Prof. Dr. Sevtap Kestel who proposed me the subject of this thesis, for her friendship, guidance, motivation, endless support and encouragemalet during my thesis and in the whole of my graduate education so far. I would like thank Prof. Dr. Ömer L. Gebizlioglu whose lectures gave me the opportunity to learn more about the Actuarial Sciences. I especially thank Assoc. Prof. Dr. Fatih Tank for his guidance and valuable advises during my research. Also, I thank Mr. İbrahim Erkan for his help and guidance on software use. And, I would like to express my gratefulness to my family, especially my husband, for their endless support, patience and help. Lastly, I thank my friends and the members of the Institute of Applied Mathematics a lot for their help and kindness during my studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1. Motivation and Literature Review

Insurance companies should take into account the risks related to morbidity and mortality in product development, premium and reserve calculations in life and health insurances. It is important to identify the recurrence of illnesses in a policy year that the insured will face to and the risks that the insured prone to related to the specified morbidities. The risk factors like age, gender, genetics, socio-economic conditions, occupation, life-style and habits play important role whether this risk is acceptable or not. It's essential to know the probability of an insured applying to a health institute due to one of the morbidities covered by the insurance policy per year to evaluate the economic burden of this insured person to the insurance company.

It's the fact that morbidity tables and studies producing information on morbidity in Turkey do not include the same level of information compared to the ones internationally accepted and commonly used. A recent project introducing the morbidity tables for the selected 32 morbidities and for the years 2008, 2009 for Turkey has been recently completed (METU, IAM, 2010).

Theoretical methodologies for studying morbidity depend strongly on the availability of the data. The fundamental measures of morbidity are incidence and prevalence and morbidity tables are constructed with respect to these measures. Prevalence is the measure of commonness and existence of different morbidity cases in a specified period of time and place in a population, whereas incidence is about the measure of having a morbidity at a first time in a population. With respect to measurement of morbidity risks, prevalence is used to identify the burden of the morbidity to health services and incidence is used to identify the risk of falling in morbidity. On the other hand, in survival analysis, there exists models and quantitative analysis to estimate the probability of one person's falling in a morbidity in a specified period.

Survival models, which are mostly used in life sciences enable researchers to model the lifetime of a living in which the age and the lifetime of the object/individual plays an important role. There are three types of survival analysis approach which are parametric, non-parametric and semi-parametric named by Cox-Proportional Regression model (Lee, 1992).

Reliability analysis, which is mostly about failure of systems such as machines or equipments, is a helpful tool in risk assessment when determining the insurance premiums for risks (Natvig, 2002). The failure of a unit in a system means the termination of the unit's ability to perform the required function. The relevance of reliability theory to insurance was pointed by Straub (1971) and in his paper, he applies results and techniques from reliability theory to establish bounds for unknown loss probabilities.

Epidemiological and biostatistical studies on cancer and any other chronic morbidities have expanded markedly since the 1950s. Therefore, there has been a need to develop more sophisticated approaches to identify the potential etiological factors in populations living in a wide variety of environments and under very different socio-economic conditions. With this respect, log-linear and logistic models are created to analyse categorical data, and the related proportional hazards model for survival time studies (Breslow and Day, 1980). There are not many studies on survival analysis applied to health insurance. However, there are plenty of researches about investigating the lifetimes of patients suffering from certain morbidities. In the study made by Yetkin (2006) 481 patients having heart diseases have been examined for five years after they had surgery. The lifetimes of these patients and the factors that affect their lifetimes are analyzed by Cox Regression Analysis. Another study held by Topçu (2007) introduces Kaplan-Meier estimate for the survival function and Greenwood formula is used to obtain the variance of the estimates on a data set of breast cancer patients.

Lu, et al (2008) mention the difficulty of sampling distribution of a premium rate without access to the original data. From two studies of adult polycystic kidney disease, they obtain, not the original data, but the cases and exposures used for Kaplan-Meier estimates of the survival probability. By this method, they estimate the premium rates for insurance contracts. And the study held by Czado and Rudolph (2002), shows the application of survival analysis methods to long-term care insurance.

The report held by Kısa and Younis (2006) discusses how the Turkish people in poverty gain access to medical care with the green card system and presents implications for policy changes and suggestions toward increasing service utilization. Also, the health policy brief (Bovbjerg, 2007) highlights that having health coverage is associated with better health-related outcomes.

The study done by Macdonald et al develops a model that can be used to assess the impact on insurance underwriting of genetic information relevant to coronary heart disease and stroke with the major risk factors of interest. This study also extends the model to include other critical illnesses and describes some applications of the model.

1.2. Aim of the study

Insurance experts are mostly focused on the impact of annual aggregate loss within a policy year. Health insurance expenses are getting bigger due to environmental, economical, and longevity risks. Its financial burden on insurance companies are unpredictable and can also be catastrophical. For this reason, the distribution of the claims per year will give an insight to the experts to get prepared for claim payments and required reserve estimations. A health insurance portfolio of individual insureds may arise a claim frequency distribution fitting to the statistical distributions listed in actuarial theory. However, such modelling require complete information collected over years. Based on the information on registrations from a specific morbidity to hospitals under the coverage of the largest insurance system in Turkey, Social Security Institute (SSI), the required information to accomplish this modeling is not complete. Therefore, under such restricted information, the number of claims per policy per year does not yield a tractable statistical distribution. It gains importance how long it takes for an insured to apply to a health institute because of a certain morbidity as the first time within the policy year. Thus, time to first claim will be an important indicator.

The aim of this thesis is to determine the time to first claim distribution for selected morbidities based on real data. The recordings on selected morbidities for the year 2008 on insured base is taken from Turkish Social Security System.

The thesis follows the steps stated below:

1. Stochastic and statistical modelling of time to first claim data for two morbidities.
2. Assessing parametric and non-parametric models on time to first claim.
3. The comparison of results obtained by non- parametric Kaplan-Meier and life table methods.
4. Determining hazard rates by using survival functions.
5. Determination of actuarial expected loss under individual and collective risk models for the selected morbidity based on the simulation study under certain assumptions.

1. 3. Organization of the thesis

The thesis starts with introduction of the main idea, which is the application of non-parametric survival models to the time to first claim data. In the second chapter, fundamental concepts in health sciences and morbidity measures are presented. Then, the general characteristics of rheumatic heart disease (RHD) and poisoning morbidities, their status in the world and in Turkey are presented. In the third chapter, important characteristics of survival models, Kaplan-Meier and life table methods are explained in detail. Estimation of survival and hazard functions, Greenwood's formula to estimate the variance of survival estimates, and the log-rank test statistic to test the difference between survival functions belonging to different groups are shortly explained. Chapter four is devoted to actuarial premium calculation. Individual and collective risk models, the importance of expected loss in premium calculations and the distribution assumptions for claim frequency and claim amount are explained. Chapter five presents the application of the methodology. Two morbidities mentioned above are taken to perform the application of the proposed approach. The data set, variables, modelling, analyses and the results are presented. Chapter six contains conclusion, comments and future studies.

CHAPTER 2

MAIN CONCEPTS AND DEFINITIONS

In epidemiology, risk refers to the likelihood of an individual in a defined population developing a disease or other adverse health problems. The association between risk of disease and both individual and social characteristics (risk factors) is often the starting point for causal analysis in epidemiology. Epidemiological studies measure and interpret disease frequency, usually by comparing the patterns in one population relative to another where measures of disease frequency include incidence and prevalence rates (Gordis, 2000).

Morbidity is the condition of being diseased and states information about the ratio of sick to well persons in the society. Morbidity tables gives information about the occurrence of a certain disease, in a determined time interval and predefined conditions in a society. There are several areas we use morbidity tables such as risk rating, premium calculation in insurance, product development, comorbidity (two or more coexisting medical conditions or disease processes that are additional to an initial diagnosis) and calculation of risk reserves (METU, IAM, 2010). A data set is needed that have long enough period of construction and it should include personel information like age, gender and any other important characteristics to construct morbidity tables. Because Turkey’s health insurance recording and data systems are inefficient and decentralized, the exact prevalence and incidence rates for various diseases and causes of death cannot be determined.

Morbidity and mortality measures are routinely collected in the developed countries but are lacking in the developing countries. The lack of routinely collected reliable and accurate morbidity and mortality measures disables effective eradication of infectious diseases and control of chronic morbidities in these developing countries.

To explain the frequency of illness producing and death causing diseases we need the number of people in the population, therefore we explain disease frequency in terms of rates. A disease rate is defined as the number of persons with a disease per unit of the population of the place of interest at a given time period. The occurrence of disease can be measured using rates or proportions. Rates tell us how fast the disease is occurring in a population and proportions tell us what fraction of the population is affected (Gordis, 2000).

The **incidence** of a disease is defined as the number of new cases of a disease that occur during a specified period of time in a population at risk for developing the disease as follows (Gordis, 2000):

$$\text{Incidence} / 1{,}000 = \frac{\substack{no\ of\ cases\ of\ a\ disease\ occurring\ in\ the\ population \\ during\ a\ specified\ period\ of\ time}}{\substack{no\ of\ persons\ exposed\ to\ the\ risk\ of\ developing\ the\ disease \\ during\ the\ same\ period}} * 1{,}000 \quad (2.1)$$

The important element in the definition of incidence is *new cases* of disease. Incidence is a measure of events- the disease develops in a person who do not have the disease previously. As incidence is a measure of events, incidence is a measure of risk. This risk can be looked at in any population group, such as a particular age group, males or females, or an occupational group. To measure incidence, a period of time must be specified, and it must be known that all of the individuals in the group represented by the denominator have been followed up for that entire period. The choice of time period is arbitrary: it can be a week, a month or a year.

Prevalence is defined as the number of affected persons present in the population at a specific time divided by the number of persons in the population at that time.

$$\text{Prevalence} / 1{,}000 = \frac{\substack{total\ cases(old\ and\ new)\ of\ a\ disease\ present\ in\ the\ population \\ at\ a\ specified\ time}}{\substack{no\ of\ persons\ in\ the\ population\ at\ that \\ specified\ time}} * 1{,}000 \quad (2.2)$$

Prevalence can be viewed as a slice through the population at a point in time at which it is determined who has the disease and who does not. But in so doing, when the disease developed is not determined. Thus when we survey a community to estimate the prevalence of a disease, the duration of the disease is not generally taken into account. Consequently, the numerator of prevalence includes a mix of people with different durations of disease, and as a result we do not have a measure of risk. Therefore, if we wish to measure risk, we must use incidence as stated before.

Distribution of diseases by personal attributes depends on age, gender, marital status, race or nationality, occupation and social class (Gordis, 2000). Some of those in detail are:

Age: Knowledge of the distribution of diseases and deaths by age is useful to the clinicians. Certain diseases are more common in the younger age groups than the older groups. Chronic diseases such as cancer, heart diseases, diabetes, hypertension are more common in older people, whereas many infectious diseases like measles, mumps, and rubella considered as childhood diseases. Age is also related to the severity of certain infections.

Gender: In most societies, more males (M) are born alive than females (F). Many of the chronic diseases such as diabetes, obesity, and certain benign tumors are more common in female than male while lung cancers, peptic ulcers, and coronary hearth diseases tend to concentrate in male.

Marital Status: Certain factors affect the distribution of diseases by marital status such as consanguinity, natural selection and environmaletal factors. In the case of consanguinity, people tend to marry their relatives thus predisposing both husband and wife to genetically acquired diseases. Environmental factors such as place of living or job, may predispose both husband and wife to the same environmentally related diseases.

Occupation: It has been known that certain occupations carry certain disease risks. Epidemiologic studies have revealed the concentration of lung cancers in asbestos workers. Dry cleaners who are exposed to aniline dyes and other chemicals have been found to develop more bladder cancer. Drug addition has been established to be more frequent among physicians and nurses as compared to other professions and trauma is very highly associated with people in the mining, construction and agricultural occupations.

Social Class: Social class is another acquired attribute of persons, which shows a correlation with the distribution of diseases. Socio-economic status is affected by certain variables such as education, income, and place of residence. For instance, bronchitis tend to be more common in the lower social class, while coronary heart diseases and leukemia tend to be more common in the upper social class.

2.1. Selected Morbidities

Two different morbidities are considered in the content of this study due to their characteristics. Rheumatic heart disease and poisoning morbidities are taken in the study for the aplication of the methodology proposed. Both of the morbidities have sub classifications and they are classified with respect to ICD 10.

The ICD (International Classification of Diseases) is the international standard classification for all general epidemiological, many health management purposes and clinical use. These include the analysis of the general health situation of population groups and monitoring of the incidence and prevalence of diseases and other health problems in relation to other variables such as the characteristics and circumstances of the individuals affected, reimbursement, resource allocation, quality and guidelines (WHO).

It is used to classify diseases and other health problems recorded on many types of health and vital records including death certificates and health records. In addition to enabling the storage and retrieval of diagnostic information for clinical, epidemiological and quality purposes, these records also provide the basis for the compilation of national mortality and morbidity statistics by WHO Member States (WHO).

2.1.1 Poisoning

A poison is any substance that is harmful to a person's body when eaten, breathed, injected, or absorbed through the skin. Any substance can be poisonous if enough is taken. Poisoning is exposure to a poison on one occasion or during a short period of time. Symptoms of poisoning are nausea, vomiting, pain, trouble breathing, seizure, confusion, or abnormal skin color. The ICD-10 codes are T36-T50 and age and gender are the basic risk factors. Poisoning may also be a deliberate attempt to commit murder or suicide. The damage caused by poisoning depends on the poison, the amount taken, and the age and underlying health of the person who takes it.

Drugs are the most common source of serious toxications and related deaths. Other common toxications are gases, household products, agricultural products, plants, industrial chemicals, vitamins, and foods (particularly certain species of mushrooms and fish).

Poisonings are either intentional or unintentional. With an 80% increase from 2001 to 2006, poisonings are one of the fastest-rising causes of unintentional death in the United States (www.eoearth.org). In 2005, 23,618 (72%) of the 32,691 poisoning deaths in the United States were unintentional, and 3,240 (10%) were of undetermined intent. Unintentional poisoning death rates have been rising steadily since 1992 (CDC, 2008). In 2006, unintentional poisoning caused about 703,702 emergency department visits (CDC, 2008). In 2006, poison control centers reported about two million unintentional poisoning or poison exposure cases (Bronstein et al, 2007). Almost 25% of these unintentional emergency department visits resulted in hospitalization or transfer to another facility (CDC, 2008).

In 2000, approximately a million people died as a result of suicide, and possibly as many as a quarter of these deaths resulted from ingestion of chemicals. In that year, unintentional poisoning was the 9th most common cause of death globally in young adults (15-29 years), and in this age group it was the 6th most common cause of death in India and the 9th most common in China. More than 94% of fatal poisonings occurred in low and middle-income countries. According to WHO data, in 2002 an estimated 350,000 people died worldwide from unintentional poisoning. More than 2 million people suffer some type of poisoning each year in the United states (WHO).

Most fatal poisonings result from unintentional drug overdoses. While children rarely die today from unintentional poisonings, non-fatal poisonings remain a childhood concern. An estimated 40,000 children under the age of 4 are injured by unintentional poisonings every year (www.eoearth.org).

Drug overdose rates have increased among males and females. In 2006, there were 17,740 drug overdose deaths among males and 8,660 among females. While males are more likely to die from a drug overdose, female rates have nearly tripled since 1999, according to the Centers for Disease Control and Prevention (www.eoearth.org**).**

In 2000, poisonings led to $26 billion in medical expenses and made up 6% of the economic costs of all injuries in the United States. Males accounted for 75% of the total costs of poisoning injuries ($19 billion). Females accounted for 25% of the total costs of poisoning injuries (almost $7 billion) (Finkelstein et al. 2006).

Within the population of Social Security Institution in Turkey, the number of people who applied to health institutes because of poisoning in 2008, is 19,916. According to the data, the highest application numbers are seen in 0-6 age group. Figure 2.1 shows the graphical representation of application numbers (METU, IAM, 2010).

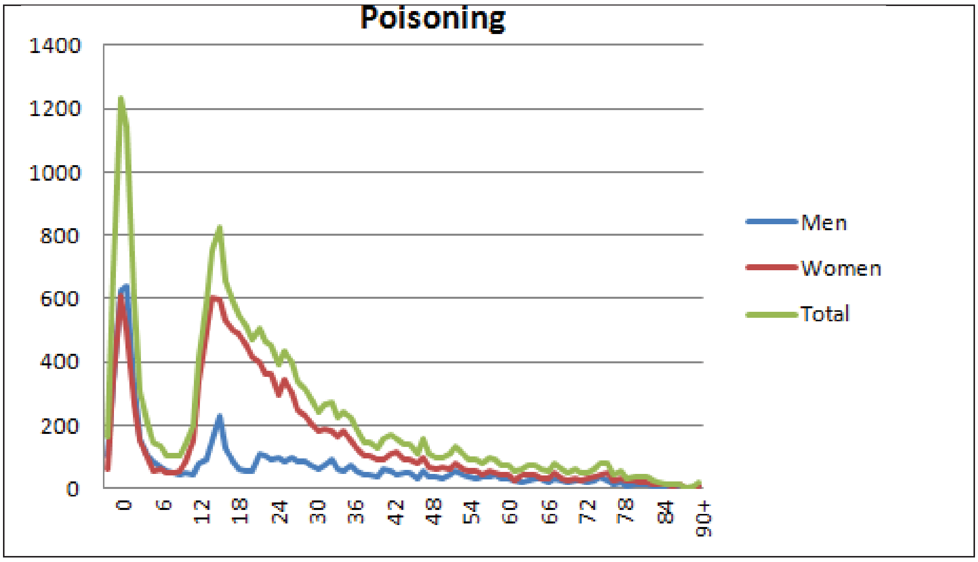


Figure 2.1: The number of applications from poisoning in Turkey (2008). (METU, IAM, 2010)

2.1.2 Rheumatic Heart Disease

Rheumatic heart disease (RHD) is a condition in which the heart valves are damaged by rheumatic fever (RF) which begins with a strep throat and caused by Group A Streptococcus bacteria. It can affect many of the body's connective tissues-especially those of the heart, joints, brain or skin. Fever, painful, tender, red swollen joints, heart palpitations, chest pain, shortness of breath, skin rashes, fatigue, small painless nodules under the skin are the symptoms of RF. Main risk factors of RHD are age, genetics, diabetes, nutrition, alcohol, and smoking. ICD-10 codes of RHD are I02, I05-I09, where I02 refers to acute RHD and I05-I09 refer to chronic RHD. The type of disease is important because chronic and acute morbidities differ in terms of number of applications to health services. Therefore, they show different patterns in the distribution among the population.

An acute morbidity starts suddenly in an unexpected time. Most of the acute morbidities are in the microbic morbidities group. Chronic morbidities tend to persist for a very long time and cause more deaths than other morbidities and a majority of them are terminal. Therefore, they have long, latent incubation periods of many months or years. There is considerable need for economic sources for diagnosis, treatment and rehabilitation for the chronic morbidity. As a result, its cost for the society is higher. As countries become developed and industrialized, they experience an epidemiological transition in which communicable or infectious morbidities give way to non-communicable or chronic morbidities as major causes of morbidity and mortality. Chronic morbidities, such as heart disease, stroke, cancer, are by far the leading cause of mortality in the world, representing 60% of all deaths (TÜİK, Turkey's Statistical Yearbook, 2009).

Although there is a decrease in the incidence of RF and a similar decrease in the prevalence of RHD, they remain significant causes of cardiovascular morbidities in the world today in both industrialized and industrializing countries. According to studies there are 200,000 new cases of RHD each year in the developing world, with 180,000 deaths are related to the disease. In 2005, the overall prevalence of RHD worldwide was estimated at 15.6 million, with 80% of these cases occurring in the developing to occur amoung the school-aged population (WHO).

Reliable data on the incidence of RF are scarce in the world. In some countries, however, local data obtained from RF registers of school-children provide useful information on trends. The annual incidence of RF in developed countries began to decrease in the 20^{th} century, with a marked decrease after the 1950; it is now below

1.0 per 100,000. A few studies conducted in developing countries report incidence rates ranging from 1.0 per 100,000 school-aged children in Costa Rica, 72.2 per 100,000 in French Polynesia, 100 per 100,000 in Sudan, to 150 per 100,000 in China (WHO, 2001).

The prevalence of RHD has also been estimated in surveys, mainly of school-aged children. The survey results showed there was wide variation between countries as shown in the Table 2.1.

Table 2.1: Examples of reported prevalence of RHD in schoolchildren (WHO, 2001)

WHO Region (country, city)	Year	Rate (per 1000 population)
Africa		
Kenya (Nairobi)	1994	2.7
Zambia (Lusaka)	1986	12.5
Ethiopia (Addis Ababa)	1999	6.4
Conakry (Republic of Guinea)	1992	3.9
DR Congo (Kinshasa)	1998	14.3
Americas		
Cuba (Havana, Santiago, P. del Rio)	1987	0.2–2.9
Bolivia (La Paz)	1986–1990	7.9
Eastern Mediterranean		
Morocco	1989	3.3–10.5
Egypt (Cairo)	1986–1990	5.1
Sudan (Khartoum)	1986–1990	10.2
Saudi Arabia	1990	2.8
Tunisia	1990	3.0–6.0
South-East Asia		
Northern India	1992–1993	1.9–4.8
India	1984–1995	1.0–5.4
Nepal (Kathmandu)	1997	1.2
Sri Lanka	1998	6
Western Pacific		
Cook Islands	1982	18.6
French Polynesia	1985	8.0
New Zealand (Hamilton)	1983	6.5 (Maoris) 0.9 (non-Maoris)
Samoa	1999	77.8
Australia (Northern Territory)	1989–1993	9.6

Based on the data set taken from Social Security Institution, the number of people who applied to health institutes from RHD in 2008, is 151,563. RHD is efficient in all age groups, however it shows an increase from age 13s and reaches the highest levels in ages 50s and 60s (METU, IAM, 2010). The average number of application age does not show definite differences in regions and cities in Turkey. However, there is a significant difference on the application frequencies between male and female. The graphical representation of the data shows that in each age group, number of applications of female is higher than male (METU, IAM, 2010).

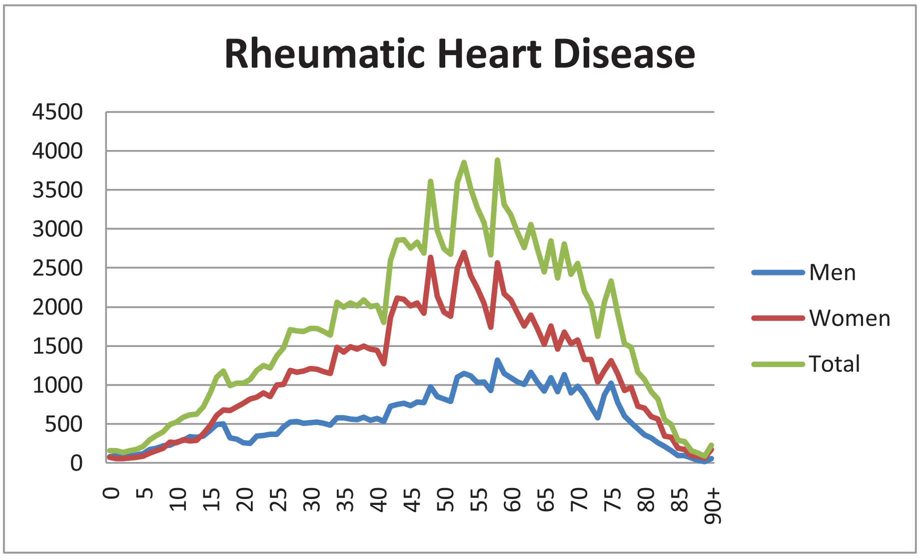


Figure 2.2: The number of applications from RHD in Turkey (2008). (METU, IAM, 2010).

It is well known that socio-economic and environmental factors play an indirect, but important role in the magnitude and severity of RF and RHD. Factors such as a shortage of health-care resources for providing quality health care, inadequate expertise of health-care providers, poverty, overcrowding, malnutrition, and a low level of awareness of the disease in the community can all impact the expression of the disease in populations (Report of a WHO Expert Organization, 2001).

Although RF and its most important sequel, RHD are worldwide problems, they are most prevalent in developing countries. In these countries, RF accounts for up to 60% of all cardiovascular disease in children and young adults and it has the potential to undermine national productivity, since young adults are the most productive segment of the population in these countries. In addition, 67% of school-aged patients drop out of school due to RF, which stifles their ability to realize their full potential (Report of a WHO Expert Organization, 2001).

Moreover, the burden of managing RHD puts additional pressure on the economies of these countries, which are often characterized by a low gross domestic product and gross national product. In countries of African region, for example, the direct medical cost of managing one patient with RHD for six years was estimated to be US $17,375 in 1987, increasing to US $31,661 with surgical procedures. The economic effects of the disability and premature death caused by these morbidities are felt at both the individual and national levels through higher direct and indirect health-care costs (Report of a WHO Expert Organization, 2001).

CHAPTER 3

METHODOLOGY

3.1. Survival Analysis

Survival analysis is one of the statistical methods applied for studying the occurrence and timing of events. Survival analysis is designed for longitudinal data on the occurrence of events. An event is a qualitative change that can be situated in time. A qualitative change refers to a transition from one discrete state to another. The initial purpose of a survival analysis is to model and analyze time-to-event data. Such events are generally called “failures”.

The basic goals of survival analysis are to estimate, interpret and compare survival and hazard functions and to assess the relationship of explanatory variables to survival time.

Examples of survival data are, the lifetime of an individual (mortality studies), the lifetime of electronic devices or systems (reliability engineering), duration of first marriage (sociology), length of newspaper or magazine subscription (marketing), and worker’s compensation claims (insurance) and their various influencing risk or prognostic factors.

For survival analysis, the best observation plan is prospective. A set of individuals are observed at some well-defined point in time, and the times at which the events of interest occur are recorded. It’s not necessary that every individual experience the event.

On the other hand, the reliability of a component or system is defined as the probability that the component or system remains operating from time zero to time t, given that it was operating at time zero.

Failure rate is the probability per unit time that the component or system experiences a failure at time t, given that the component or system was operating at time zero and has survived to time t.

Mean time to first failure describes time to first failure for non-repairable components. It is the mean time expected until the first failure of a piece of equipment. For constant failure rate systems, mean time to failure (MTTF) is the inverse of the failure rate. However, for repairable components the concept of mean time between failures is used. Mean time to first failure should not be confused with MTTF which is the average time between failures for a repairable system with non-zero repair times. The time to first failure is important not only for non-repairable systems but also for repairable systems because if someone knows time until the first failure, then he knows that there will be failures in the future.

All of the standard approaches to survival analysis are probabilistic or stochastic. That is, the times at which events occur are assumed to be realizations of some random process. It follows that T, the failure time of an individual from a homogeneous population, is a nonnegative random variable having a probability distribution. There are many different models for survival data, and what often distinguishes one model from another is the probability distribution for T.

The distribution of lifetimes is usually described or characterized by three functions.

1. Probability density function
2. Hazard function
3. Survival function

In survival analysis, it is more common to work with a closely related function called the survival function, defined as (Allison, 1995):

S (t) = P (Surviving beyond time t) = Pr (T>t) (3.1)

=1-P (an individual experience an event before time t) = 1-F (t)

$0 \le S(t) \le 1$

where F(t) is the cumulative distribution function. It can be estimated from a data set by calculating the frequency of cases in which the event had not occurred by time t. The survival function is fundamental to survival analysis, because obtaining survival probabilities for different values of *t* provides crucial information from survival data.

For a nonnegative random variable T, the mean survival time is:

$$E(T) = \int_0^{\infty} tf(t)dt = \int_0^{\infty} S(t)dt \,. \qquad (3.2)$$

For a given time t, the greater the risk, the smaller S(t), and hence the shorter mean survival time E(T).

As t gets larger, S never increases (and usually decreases). Within these restrictions, S(t) can have a wide variety of shapes. A steep survival curve represents low survival rate or short survival time, whereas a gradual or flat survival curve represents high survival rate or longer survival. The survival function or the survival curve is used to find the 50th percentile (median) and other percentiles (e.g.25th and 75th) of survival time and to compare survival distributions of two or more groups (Allison,1995).

Median survival time is the length of time that half of the study population survives. The time that corresponds to S(t)=0.5 is the time at which 50% of the population is estimated to have been experienced the event. Median survival offers two advantages over mean survival. First, it is less affected by extremes, whereas the mean is significantly affected by even a single outlier. Second, if we use mean survival, we would have to observe all the events in the population before the mean could be calculated. However, to calculate median survival, we would only have to observe the events of half the group. Lastly, the mean is generally used to describe the central tendency of a distribution, but in survival distributions the median is often better because a small number of individuals with exceptionally long or short lifetimes will cause the mean survival time to be disproportionately large or small (Allison, 1995).

Although different survival functions can have the same basic shape, their hazard functions can differ dramatically. The hazard function is also called as the risk or mortality rate as a measure of the potential of failure at time t. The hazard function is usually more informative about the underlying mechanism of failure than the survival function. For this reason, modeling the hazard function is an important method for summarizing survival data.

In contrast to the survival function, which focuses on *not* failing, the hazard function focuses on failing and models the risk. The hazard function is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr\{t \le T < t + \Delta t \mid T \ge t\}}{\Delta t} \tag{3.3}$$

of the interval [t,t+Δt). In practice, when there are no censored observations the hazard function is estimated as the proportion of individuals having failures in an interval per unit time, given that they have survived to the beginning of the interval:

$$\hat{h}(\mathrm{t}) = \frac{\text{no of patients dying in the interval beginning at time t}}{\text{no of patients surviving at t}} \bullet (\text{interval width}) \tag{3.4}$$

$$= \frac{\text{no of patients dying per unit time in the interval}}{\text{no of patients surviving at t}}$$

Actuaries usually use the average hazard rate of the interval in which the number of patients having failures per unit time in the interval is divided by the average number of survivors at the midpoint of the interval:

$$\hat{h}(\mathrm{t}) = \frac{\text{no of patients dying per unit time in the interval}}{\frac{(\text{no of patients surviving at t}) - (\text{no of deaths in the interval})}{2}} \ . \tag{3.5}$$

The estimate in the Equation (3.4) gives a higher hazard rate than the Equation (3.5) and thus a more conservative estimate.

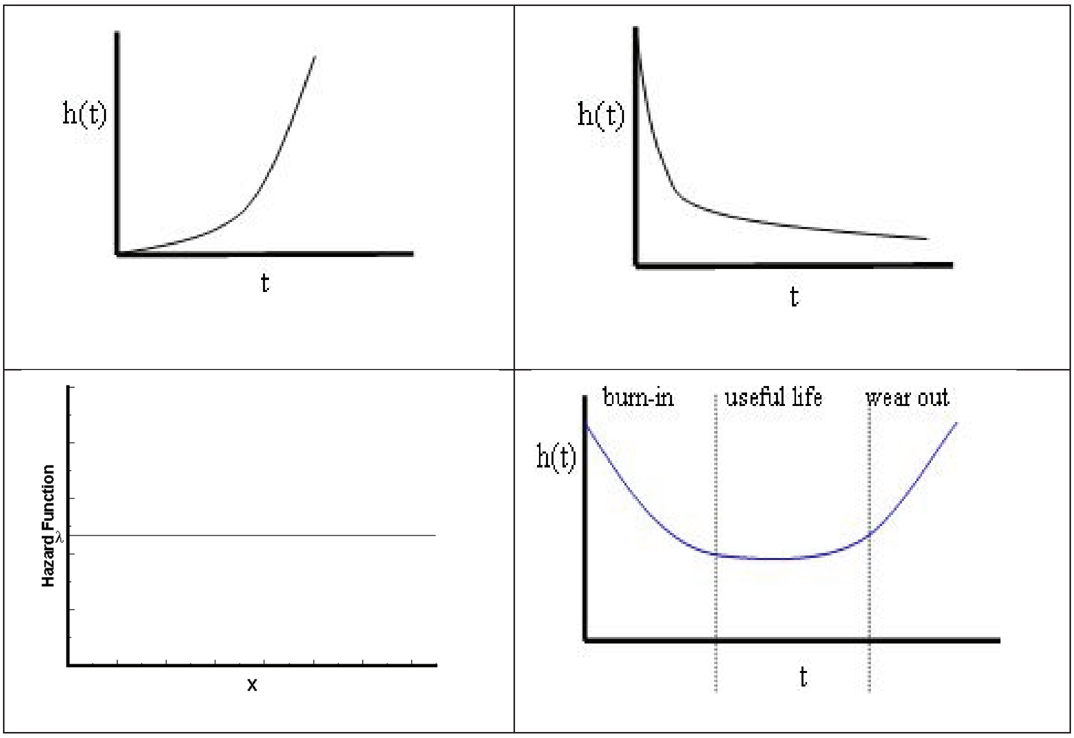


Figure 3.1: Different types of hazard functions

The hazard function may increase, decrease, remain constant or indicate a more complicated process. The first situation in the Figure 3.1, the increasing hazard function is probably the most likely situation. In this case, items are more likely to fail as time passes. For example, patients with acute leukemia who do not respond to treatment have an increasing hazard rate.

The second situation, the decreasing hazard function, is less common. In this case, the item is less likely to fail as time passes. Items with this type of hazard function improve with time. For example, some metals work harden through use and thus have increased strength as time passes.

The third situation, constant hazard function, is the risk of healthy persons between 18 and 40 years of age whose main risks of deaths are accidents.

The forth situation is the bathtub shape hazard function. Human mortality study and reliability situations are characterized by failure rate functions having "bathtub shape". The failure rate decreases initially. A good example of this is seen in the standard mortality tables for humans. The risk of death is large for infants but decreases as age advances. Next phase is known as “useful life” phase, in which the failure rate is more or less constant. Finally, in the third phase, known as “wear out phase”, the failure rate increases. Again in humans, after the age of 30 an increasing proportion of the alive persons die as age advances.

Relations between survival and hazard functions

As mentioned before, hazard function is instanteneous failure rate at T=t given that the individual survived up to time t. The relation of hazard and survival functions can be seen in the following equation:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr\{t \le T < t + \Delta t | T \ge t\}}{\Delta t} = \frac{\frac{dS(t)}{dt}}{-S(t)} = -\frac{d}{dt} \log S(t) \tag{3.6}$$

Integrating both sides of the Equation (3.6) gives an expression for the survival function in terms of the hazard function:

$$S(t) = \exp\{-\int_0^t h(u)du\}\,. \tag{3.7}$$

The relationship betwen the pdf and the survival function is given by the following equation:

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}\,. \tag{3.8}$$

Lastly, to interpret the hazard function by using the survival function and its pdf, we use the following expression:

$$h(t) = \frac{f(t)}{S(t)}\,. \tag{3.9}$$

These formulas are extremely useful in any mathematical treatment of models for survival analysis because it is often necessary to move from one representation to another.

3.2. ESTIMATION ANALYSIS

3.2.1. Kaplan- Meier Method

Kaplan-Meier (KM) (Kaplan and Meier, 1958) method is a non-parametric method to estimate the survival function from lifetime data and also known as the product limit estimator. KM method cannot evaluate covariates like Cox Regression Model and there are no sensible interpretation for competing risks. Exact times of failure are required in KM analysis. In medical research, it is used to measure the fraction of patients living for a certain amount of time after treatment.

Kaplan-Meier Estimates of Survival Function

Consider the case where all patients are observed to occurrence of the event so that the survival times are exact and known. Let $t_1, t_2, ..., t_n$ be the exact survival times of n individuals under study. Conceptually, consider a group of patients as a random sample from a much larger population of similar patients. Setting survival times $t_1, t_2, ..., t_n$ in ascending order such that $t_1 \le t_2 ... \le t_n$.

Let,

n_i = the number of subjects at risk at the beginning of time period t_i

d_i= the number of subjects who die during time period t_i, the probability becomes

$$p_i = P\left(\text{surviving through } t_i \,\middle|\, \text{alive at beginning } t_i\right) \quad (3.10)$$

$$= P\left(T > t_i \,\middle|\, T > t_{i-1}\right)$$

$$q_i = 1 - p_i = P\left(\text{die in } t_i \,\middle|\, \text{alive at beginning at time } t_i\right) \quad (3.11)$$

From repeated applications the survival function can be expressed as (Lee, 1992):

$$S(t) = P(T > t) = \prod_{t_i \le t} p_i \ , \qquad (3.12)$$

where the estimates of p_i and q_i are :

$$\hat{q}_i = \frac{d_i}{n_i} \text{ and } \hat{p}_i = 1 - \hat{q}_i = 1 - \frac{d_i}{n_i} = \frac{n_i - d_i}{n_i} \qquad (3.13)$$

Therefore, the KM estimator of the survival function is

$$\hat{S}(t) = \prod_{t_i \le t} \hat{p}_i = \prod_{t_i \le t} \frac{n_i - d_i}{n_i} = \prod_{i=1}^{k} \frac{n_i - d_i}{n_i} \ , \qquad (3.14)$$

where S(t) is the estimated survival probability for any particular one of the t time periods.

If two or more t_i are equal (tied) observations, the largest i value is used. This gives a conservative estimate for the tied observations.

Since every person is assumed to be alive at the beginning of the study and no one survives longer than t_n , $\hat{S}(t_0) = 1$ and $\hat{S}(t_n) = 0$.

The probability of surviving k (≥ 2) or more years from the beginning of the study is a product of k observed survival rates:

$$\hat{S}(\mathrm{k}) = \mathrm{p}_1 * \mathrm{p}_2 * \ldots * \mathrm{p}_k \ , \qquad (3.15)$$

where p_i denotes the proportion of patients surviving at least i year.

Therefore, the KM estimate of the probability of surviving any particular number of years from the beginning of the study.

$$\hat{S}(t) = \hat{S}(t-1) * p_t \qquad (3.16)$$

KM estimator is defined for any time between 0 and the largest event time. The KM curve is a right continuous step function which steps down only at an observed event time.

It is important to know how certain our estimate of S(t) is. Within the frequentist framework, this is found via, (Kleinbaum, 1996)

$$V\{\hat{S}(t)\} = V\{\prod_{t_i \le t} \frac{n_i - d_i}{n_i}\} = V\{\prod_{t_i \le t} \hat{p}_i\}. \tag{3.17}$$

For simplicity, under the assumption that failures arise independently among the population the variance of log-transformed $\hat{p}_i$'s are taken:

$$V\{\log \hat{S}(t)\} = V\{\prod_{t_i \le t} \log \hat{p}_i\} = \prod_{t_i \le t} V\{\log \hat{p}_i\}. \tag{3.18}$$

After employing Delta Method to estimate the variance and Taylor expansion around mean value and some more mathematical steps, Greenwood's estimator of the variance of S(t) (Greenwood, 1926) is given as:

$$\hat{V}\{\hat{S}(t)\} = \hat{S}(t)^2 \sum_{t_i \le t} \frac{d_i}{n_i(n_i - d_i)}. \tag{3.19}$$

The confidence intervals for the survival probability is found as given below:

The square root of $\hat{V}[\hat{S}(t)]$ is the standard error of the KM estimate and $\hat{S}(t) \pm 1.96\sqrt{\hat{V}(\hat{S}(t))}$ gives a pointwise 95% confidence interval. However, Kalbfleisch and Prentice (2002) suggests to use

$$\hat{V}\{\log(-\log \hat{S}(t))\} = \frac{1}{(\log \hat{S}(t))^2} \sum_{t_i \le t} \frac{d_i}{n_i(n_i - d_i)}, \tag{3.20}$$

to get a confidence interval of log (- log$\hat{S}$(t)), and then to transform that back into the original scale such as,

$$c_1 = \log(-\log \hat{S}(t)) + z_{1-\frac{\alpha}{2}} \sqrt{V\{\log(-\log \hat{S}(t))\}}$$
$$c_2 = \log(-\log \hat{S}(t)) - z_{1-\frac{\alpha}{2}} \sqrt{V\{\log(-\log \hat{S}(t))\}} \quad (3.21)$$

So that the 1-α confidence interval for S(t) becomes $(\exp\{-e^{c_2}\}, \exp\{-e^{c_1}\})$.

Estimates of hazard and cumulative hazard functions based on KM method can also be found. Letting the estimate at an observed death time be t_i and $\hat{h}(t) = \frac{d_i}{n_i}$ and the estimate of hazard in the interval $t_i < t < t_{i+1}$ is $\hat{h}(t) = \frac{d_i}{n_i}(t_{i+1} - t_i)$.

Also, the estimates of cumulative hazard to time t and its variance are expresses as:

$$\hat{H}(t) = -\log \hat{S}(t) = -\log \prod_{t_i \le t} \frac{(n_i - d_i)}{n_i} \quad . \quad (3.22)$$

$$\hat{V}(\hat{H}(t)) = \sum_{t_i \le t} \frac{d_i}{n_i (n_i - d_i)} \quad . \quad (3.23)$$

3.2.2. Log-Rank Test

Log-rank test is used to compare KM curves to evaluate whether or not two or more groups are statistically equivalent. The log-rank test is a large-sample chi-square test that uses as its test criterion a statistic that provides an overall comparison of the KM curves being compared. This statistic, like many other statistics used in other kinds of chi-square tests, makes use of observed versus expected cell counts over categories of outcomes. The categories for the log-rank test statistic are defined by each of the ordered failure times for the entire set of data being analyzed.

The main objective is to compare two survival curves belonging to different treatment groups. The null hypothesis is:

$H_0: S_1(t) = S_2(t)$ against one of the following alternatives

$H_1: S_1(t) > S_2(t)$ or $H_2: S_1(t) < S_2(t)$ or $H_3: S_1(t) \neq S_2(t)$

Suppose that, Group I and Group II are groups of survival data and let t_i represent the *ith* order event time in the two groups combined. The contingency table (Table 3.1) is used to represent two sets of survival data.

Table 3.1: Contingency table for two groups

	Died	Survived	At risk
Group I	d_{1i}	$n_{1i} - d_{1i}$	n_{1i}
Group II	d_{2i}	$n_{2i} - d_{2i}$	n_{2i}
Total	$d_i = d_{1i} + d_{2i}$	$n_i - d_i$	$n_i = n_{1i} + n_{2i}$

where, d_{ji} is the number of people experiencing the event at time t_i in group j, n_{ji} is the number of people at risk in group j at time t_i, n_i is the total number at risk and d_i is the total number experiencing the event in both groups.

Then, the log-rank test statistics is:

$$Q = \frac{(\sum_{i=1}^{m} d_{1i} - \sum_{i=1}^{m} \hat{e}_{1i})^2}{\sum_{i=1}^{m} \hat{V}(\hat{e}_{1i})}, \qquad (3.24)$$

where $\hat{e}_{ji}$ is the expected number of individuals experiencing the event at time t_i in group j. This test statistic is distributed as χ_1^2 under the null hypothesis that the survival functions for the two groups are the same. Based on these statistics, the chi-square test statistic is calculated and the corresponding p-value is obtained. The generalization to this test for more than two groups also follow chi-square distribution with r-1 degrees of freedom, where r standing for the number of groups to be compared (Fox, 2006).

3.2.3. LIFE-TABLE METHOD

Life tables have historically been used by actuaries for estimating the survival distribution of humans, but applied well to reliability and biostatistical situations for which grouped data (rather than raw data). Grouped data displays the combined survival experience of a cohort of individuals who fall into natural groupings by age or calender time interval.

As clinical and epidemiologic research become more common, the life table method has been applied to patients with a given disease who have been followed for a period of time. Life tables constructed for patients are called clinical life tables.

If the number of observations is large and if event times are precisely measured, there will be many unique event times. The KM method then produces long tables that may be unpractical for presentation and interpretation. This problem can be avoided with the life table method because event times are grouped into intervals that can be as long or short as desired. In addition, the life table method (also known as the actuarial method) can produce estimates and plots of the hazard function. The drawback to the life table method is that the choice of intervals is usually somewhat arbitrary, leading to arbitrariness in the results and possible uncertainty about how to choose the intervals.

The survival estimate by using life table method is calculated from the conditional probabilities of failure in the following way:

For interval i, let t_i be the start time and q_i be the conditional probability of failure. The probability of surviving to t_i or above becomes (Allison, 1995):

$$\hat{S}(t_i) = \prod_{j=1}^{i-1} (1 - q_j) \,. \qquad (3.25)$$

It should be noted that the similarity between the Equation (3.25) and KM estimator. Both are the estimates of the probability of failure in an interval. The major difference is: The interval boundaries for the KM estimator are determined by the event times themselves. Thus, each interval for KM estimation extends from one unique event time up to, but not including, the next unique event time (Allison, 1995).

CHAPTER 4

ACTUARIAL RISK MODELLING

4.1.Health Insurance

Life and health insurers sell individual and group health insurance plans that cover medical expenses from sickness or injury. Individual health insurance policies can generally be classified as: hospital-surgical insurance, major medical insurance, long-term care insurance, and disability-income insurance. Group health insurance is preferable to individual coverage for several reasons. First, employers frequently make available a number of group health insurance plans to their employees, ranging from traditional group indemnity plans to managed care plans, such as those involving health maintenance organizations. Second, health insurance is typically broader in coverage (Rejda, 2002).

In Turkey, all health care and related social activities are controlled and coordinated by the Ministry of Health. The ministry is responsible of providing health care for the people and organizing preventive health services, building and operating state hospitals, supervising private hospitals, training medical personel, regulating the price of medical drugs nationwide, controlling drug production and the pharmacies. The Turkish health system is a combination of national health insurance and private health insurance. A general health insurance system has been implemented since the early 1990s under the umbrella of Social Security Institution. The suggested system is financed from the premiums based on actual contributions. In the transition period to the general health insurance system, the “green-card” system was started in 1992 as a step toward ensuring equity in the distribution of state subsidies to needy citizens for health services.

Today, three main mechanisms for financing healthcare can be seen worldwide. These are taxation, insurance (social and private) and non-insurance funding systems. Social health insurance systems involve mandatory and earnings related contributions to institutions independent from general government revenue. The Turkish healthcare system is financed by taxes (41%), insurance premiums (31%), and out-of–pocket payments (28%). Besides social security system, private health insurance is developed in Turkey. Employers pay insurance fees to cover work-related injuries, professional job diseases or maternity. Both employers and employees contribute specified proportions to cover fees for illness, disability, retirement and death benefits.

In recent years both health expenses and chronic diseases have been increasing in Turkey (Türk Sağlık Sen Report). The basic health indicators shows that Turkey is far beyond the OECD countries, where the protective health services neglected, instead treatment oriented health services is implemented. As a result of this fact, in 2006 over two million operations made in the hospitals affiliated to the Ministry of Health.

In 2002, total health expenditures cost 3,594 billion TL, and this has increased to 11,6 billion TL in 2006. And the drug expenditures cost 1,878 billion in 2002, whereas this number has increased to 5,265 billion TL in 2006. The number of surgeries done in government hospitals in 2002 was 836,518, whereas in 2005 this number has increased to 2,288,489. This huge increase on the number of surgeries is the indicator of treatment oriented health care system (Türk Sağlık Sen Report).

Therefore, the financial burden of health system on government and insurance companies is getting an important issue. The control over those expenses are manyfold. One of the important control over the unexpected losses lies on the precise estimation of the premiums to be charged from insurers.

The aim of the insurer is to know exactly how much the company will pay out in benefits and then charges premiums to match this amount but, it is hard to predict future claims. The stochastic nature and the correct estimation of both the frequency and severity of claims are fundamental components of constructing a realistic model.

4.2. Individual and Collective Risk Models

There exist two fundamental actuarial approaches to estimate actuarial premium in insurance. These are individual and collective risk models, which are explained in the following sections.

4.2.1 Individual Risk Model (IRM)

The insurer wants to know the total benefits paid on an entire portfolio of policies. And the IRM is used to obtain the present value of the total amount paid on all policies in the portfolio, as the sum of the individual random variables. The IRM considers individual policies and the claims produced by each policy.

Let the random loss of a segment of company's risks be denoted by S and let $S = X_1 + X_2 + ... + X_n$ denote the aggregate loss. Here X_i is the loss on insured unit i and n is the number of risk units insured. Under the assumption that X_i's to be independent random variables, the distribution of their sums can be calculated by convolution method (Bowers, 1997 and Kaas, 2001).

Defining $S = \sum_{i=1}^{n} X_i$ and $X_i = I_i B_i$ where I_i is an indicator function having Bernoulli(q_i) distribution, q_i being probability of claim. The random variable B_i can have an arbitrary distribution and represents the amount of the payment in respect of the i^{th} policy given that the payment was made. Letting, $q_k = P(I_k = 1)$, $1 - q_k = P(I_k = 0)$ and B_i , I_i's are independent (Kaas, 2001). Expected value and variance of S are given as follows:

$$E[S] = \sum_{i=1}^{n} E[X_i] \quad (4.1)$$

$$E[X_i] = E_I[E_{X|I}[X_i | I_i]] = q_i E[B_i] \quad (4.2)$$

$$Var[S] = \sum_{i=1}^{n} Var[X_i]$$

(4.3)

4.2.2 Collective Risk Model (CRM)

In CRM, a random process that generates claims for a portfolio of policies is assumed. This process is characterized in terms of the individual policies comprising the portfolio. The CRM is particularly useful for casualty insurance such as automobile, home or health policies. Because, in a given period there can be several claims under a single policy. This can be idealized as many visits to practitioner from the same policy owner. The main advantage of a CRM is that it is a computationally efficient model, which is also rather close to reality.

Let $S = X_1 + X_2 + .. + X_N$, where N denotes the number of claims and X_i is the *ith* claim. Number of claims N is a random variable, and that the individual claims X_i are independent identically distributed. Also, the claim number N and the claim amounts X_i are independent. The CRM views total claims as a compound distribution. Even though the insurer is ultimately interested in the total payout, it has been found advantageous to first model frequency and severity separately and to then combine the results to determine total claims.

The expected value of S by using the conditional distribution of S, given N is: (Bowers, 1997).

$$E[S] = E[E[S|N]] = \sum_{n=0}^{\infty} E[X_1 + X_2 .. + X_N | N = n] \Pr[N = n] \tag{4.4}$$

$$= \sum_{n=0}^{\infty} E[X_1 + X_2 .. + X_n | N = n] \Pr[N = n]$$

$$= \sum_{n=0}^{\infty} E[X_1 + X_2 .. + X_n] \Pr[N = n]$$

$$= \sum_{n=0}^{\infty} n\mu_1 \Pr[N = n] = \mu_1 E[N]$$

where $\mu_1 = E[X]$.

The variance of the process is:

$$Var(S|N = n) = Var(\sum_{i=1}^{N} X_i | N = n) \tag{4.5}$$

$$= Var(\sum_{i=1}^{n} X_i | N = n)$$

$$= Var(\sum_{i=1}^{n} X_i) = nVar(X)$$

$$E(S|N=n)=E(\sum_{i=1}^{N}X_i|N=n)$$

$$=E(\sum_{i=1}^{n}X_i|N=n)$$

$$=E(\sum_{i=1}^{n}X_i|N=n)=nE(X)$$

$$Var(S|N)=NVar(X), E(S|N)=NE(X)$$

$$Var[S]=E[Var[S|N]]+Var[E[S|N]]$$

$$=E[NVar[X]]+Var[NE(X)]$$

$$=E[N]Var[X]+E(X)^2 Var[N]$$

Variance represents uncertainty, and this decomposes the uncertainty in the value of S into two parts. The first term gives the uncertainty resulting from the severity, and second term gives the uncertainty arising from the frequency.

The most commonly used severity distributions are: Gamma (α, β) which is used if the tail of the cumulative distribution function is not too 'heavy', such as in motor insurance for damage to the own vehicle; Lognormal (μ, σ^2) distribution is usually for branches with somewhat heavier tails, like fire insurance; Pareto (α, x_0) distribution is used for branches with a considerable probability of large claims, notably liability insurance.

Besides these distributions there are a lot more possibilities, including the inverse Gaussian and mixtures/combinations of exponential distributions.

Appropriate distributions for claim frequency N are the Poisson, Binomial and the Negative Binomial distribution. The Poisson distribution is one of the most common distributions used for modelling claim frequency. It arises in a natural way from our independence and time-homogeneous assumptions by taking the limit of binomials and due to of its nice properties.

Difference between models can be analyzed through the mean and the variance of some insured profiles. For example, the Poisson distribution is equidispersed since its mean and variance are both equal to λ. In such case S is told to follow a Compound Poisson distribution with parameter λ.

Given the fact that sum of Compound Poisson is also distributed as Poisson. Sum of claim amounts S_i, for S_i's being mutually independent random variables, collective risk models can be expressed as follows (Bowers, 1997):

Theorem:

Let $S_i \sim CompoundPoisson(\lambda_i)$ and $P_i(x)$ claim distribution for i=1,..,m. Then, $S=\sum_{i=1}^{m} S_i \sim CompoundPoisson(\lambda)$ where $\lambda=\sum_{i=1}^{m}\lambda_i$ and $P(x)=\sum_{i=1}^{m}\frac{\lambda_i}{\lambda}P_i(x)$.

Proof :

Let $M_i(t)$ denote the moment generating function (m.g.f.) of $P_i(x)$ and we obtain the m.g.f. of the compound Poisson distribution,

$$M_S(t)=e^{\lambda[M_X(t)-1]} \quad . \tag{4.6}$$

According to the Equation (4.6), the m.g.f. of S_i is $M_{S_i}(t)=\exp\{\lambda_i[M_i(t)-1]\}$.

By the assumed independence of $S_1, S_2, ..., S_m$, the m.g.f. of their sum is

$$M_S(t)=\prod_{i=1}^{m} M_{S_i}(t)=\exp\{\sum_{i=1}^{m}\lambda_i[M_i(t)-1]\}. \tag{4.7}$$

Finally, we rewrite the exponent to obtain

$$M_S(t)=\exp\{\lambda[\sum_{i=1}^{m}\frac{\lambda_i}{\lambda}M_i(t)-1]\}.$$

(4.8)

Since this is the m.g.f. of the Compound Poisson distribution, specified by $\lambda=\sum_{i=1}^{m}\lambda_i$ and $P(x)=\sum_{i=1}^{m}\frac{\lambda_i}{\lambda}P_i(x)$, the theorem follows.

This result has two important consequences: First, if we combine m insurance portfolios, where the aggregate claims of each of the portfolios have Compound Poisson distributions and are mutually independent, then the aggregate claims for the combined portfolio will also have a Compound Poisson distribution. Second, we can consider a single insurance portfolio for a period of m years. Here we assume independence among the annual aggregate claims for the m years and that the aggregate claims for each year have a Compound Poisson distribution.

A special case is when the S_i have fixed claims X_i , hence $S_i = x_i N_i$ with $N_i \sim$ Poisson (λ_i). The random variable $S = x_1 N_1 + x_2 N_2 + ... + x_m N_m$ is compound Poisson with parameters, assuming the X_i to be all different: $\lambda = \lambda_1 + \lambda_2 ... + \lambda_m$ and $p(x_i) = \frac{\lambda_i}{\lambda}$, i=1,2,...,m.

4.3 Actuarial expectation in life and health insurance:

The present value of an insurance contract, which guarantees to pay an amount of $c_{j,k+1}$ at the end of the year k+1 to an insured person who develops a morbidity depending on j risk factor is: $A = c_{j,k+1} v^{k+1}$, where v is the discount factor with a determined interest rate.

The net premium calculated by the actuarial expected value principle is as follows (METU, IAM, 2010):

$$E(A) = \sum_{j=1}^{m} \sum_{k=0}^{\infty} c_{j,k+1} v^{k+1} {}_k p_x q_{j,x+k} \tag{4.9}$$

where for a given index k, ${}_k p_x$ is the probability that a person aged x do not have a morbidity j in k years, and $q_{j,x+k}$ is the probability that a person aged x fall in a morbidity j between the ages of $x+k$ and $x+k+1$. The net premium could be taken from the insured as monthly or yearly paymalets.

The above premium calculation formula yields the average value of a claim amount c_{ij}, caused from an insured person, aged x, who has lived a one-year policy period i and yields claims depending on the morbidity j.

However, in this study, the individual characteristics of individuals are neglected. Therefore, the age of the insured, mortality and morbidity rates associated to the age will not be associated to the premium calculation and the main focus is given to the claim load of the company which may result from the very first visit of the policy holder who survived up to the time of first claim and caused a certain claim amount. This approach allows to investigate the aggregate claim amount and its distribution for any specified period within a policy year. And the first claim occurrence for selected morbidity within one year period is taken into account for the individuals covered by Social Security System in Turkey. In this context, individual and collective risk models are used to determine the average claim amounts and its financial burden on the insurance company.

CHAPTER 5

CASE STUDY APPLICATIONS

Application of survival models by using Kaplan-Meier and life table methods is performed on two morbidities mentioned in detail in previous chapters. The real life data taken for RHD and poisoning is used to implement the methodology. The survival probabilities, hazard rates for both morbidities are calculated. Gender, age, location (three big metropolitans in Turkey) based comparisons are done.

5.1.Data set

The data set has been taken from Social Security Institution (SSI) for the year 2008 which contains the variables based on the registrations to the health institutes for morbidities RHD and poisoning as follows:

- Decoded ID number
- Age
- Gender
- Date of birth
- Application date
- ICD-10 code
- Place of treatment
- Place of birth
- Place of living
- Frequency of visits to health institutes

Tha data is processed by using SPSS and Minitab softwares for analyses. Origin time is taken as December 31, 2007. To avoid zero counts in tha data set for each individual, the time to first application to a health institute, starting from December 31, 2007 to December 31, 2008 has been retrived in terms of days.

It has to be mentioned that the data set do not represent the number of individuals in Turkey or the number of whole insureds in the Social Security Institution (SSI). The analyses are done based on the number of people who applied to health institutes within the population of SSI. The participants in the SSI constitute 80% of the whole population in Turkey. Even though this percentage is high enough to represent Turkey, the outcomes of this study is not generalized to whole population, instead kept limited to SSI coverage.

5.2.Application

Assume T_i represent time to first claims obtained from i^{th} policy holder for the selected year and refers to the time (duration) elapsed untill the first visit to health institute from prescribed morbidity. Here $T_i \in (0,366]$. As the data set is limited to one year, there exists no information if an individual for a selected morbidity has history on this disease or not. For this reason and for suitability of the proposed model, the followings are set:

(i) The history of the disease in individual base is ignored as the past information about having the disease is not known and the data do not give us the incidence of these diseases. The data gives us only the number of people, who apply to the health institutes from proposed diseases in the specified year 2008.

(ii) Homogeneity among individuals: It is assumed that population, which is SSI is homogeneous.

(iii) Independency of individuals: In the analysis, only the first failure times are taken into account for each person. Therefore, first failure times originating from individuals in the population are independent.

(iv) Sub-refractions are homogeneous in ICD-10 coding: There are so many sub-refractions for both RHD and poisoning morbidities because different factors that causes RHD or poisoning coded under a different name. However, these are taken in the general definition of RHD and poisoning.

(v) The period of observation is limited to one year (12 months) as all health insurance policy requires.

The first step taken on the time to first claim data is to fit a statistical distribution. The data has been evaluated under the assumption of distributions such as: normal, gamma, 3-parameter gamma, weibull, 3-parameter weibull, exponential, 2-parameter exponential, log-normal, 3-parameter log-normal, logistic, log-logistic, 3-parameter loglogistic, smallest extreme value, largest extreme value by using Minitab software. The decision criteria on selecting the best fitting distribution on the data set has shown that none of the distributions yield an appropriate fit to the time to first claims (Appendix A).

As the parametric modelling does not yield an appropriate distribution, non-parametric methods explained in detail in chapter III are employed. By using KM and life table methods the estimates of the survival function, and the standard errors of the estimates are produced. In life table method the investigator decides the time intervals considering the time length of the data which is done on weekly basis in this study.

The life table method should be preferred to avoid long tables as KM is more efficient for the data having small length. In KM method the survival probability according to event times is calculated, however, in the life table method the estimations are made with respect to the middle class values. In the life table method, we see the changes at the end of each interval, however, with the KM method changes happen whenever there is one event. Life table method assumes the patients uniformly distributed in each time interval. According to this assumption, the population at risk exposed to the potential risk is in the half interval.

The following comparisons in terms of survival functions are aimed:

1. For the selected morbidities, comparison on the basis of gender.
2. For the selected morbidities, comparison on the basis of three metropolitans in Turkey, namely, Ankara, İstanbul and İzmir.
3. For the selected morbidities, comparison on the basis of ages, ranging from zero to 90 above.
4. Comparison on the basis of the type of disease, such as acute and chronic RHD.

For comparisons of different survival curves both Gehan-Wilcoxon and log-rank tests are used. Gehan-Wilcoxon test gives more weight to failures at early time points, in contrast, the log-rank test gives equal weight to all time points as previously mentioned. In log-rank test statistics, the hazard rates for different groups are assumed to be constant for all time intervals. The calculations in log-rank test are based on sequential chi-square testing of various time-points, comparing expected values and observed values. The null hypothesis tested is the risk of an event is the same in all different groups.

Thereafter, the actuarial life tables performed by using the results obtained from survival analysis. Age, gender, health status, lifestyle, geographical location are major factors in determining future morbidity and mortality. In practice, the effects of some of these factors are handled by producing several different life tables. Within this consideration we performed life tables for male and female seperately. The methodology explained above will be shown for RHD and poisoning morbidities in the following section.

As final step of the application, the results of survival analysis explained above are incorparated in modelling the actuarial premium calculation. Actuarial calculation requires the knowledge of some important inputs such as claim amount distribution, frequency distributions and their parameters. Under the scarcity of data, a simulated data set is used to illustrate the impact of estimated survival function on actuarial premium calculation.

5.3.Analyses

The statistical analysis shown in this section is presented in two seperate parts for both morbidities. In the first part, the descriptive statistics and survival curves are done and explained for both of the morbidities without making any sub classifications to identify the general characteristics of the data. In the second part, survival and hazard curves are analyzed with respect to age, gender, and location for each morbidity. After those steps, the application in the aspects of actuarial premium calculation is performed.

5.3.1.Poisoning

The time to first application to the health institute for both genders is presented in the Figure 5.1. It can be seen that 33% of the applications are male and 67% of the applications are female.

The mean time to first claim is 192.77 days and the median is 194 days for male. The difference between mean and median is not large. The mean and median values are close to each other and the distribution seems to be symmetric also with respect to box-plot representation. However, the lifetime data does not follow a suitable distribution by goodness of fit results shown in the same table.

The mean time to first claim is 193.94 days and the median is 193 days for female. Again, the mean and median values are close to each other and the distribution seems to be symmetric. However, the data does not follow a suitable distribution.

It is important to mention that, for poisoning, the means and standard deviations for both genders do not show major differences.

The survival and hazard curves for poisoning are in the Figure 5.2. For both genders survival curves are decreasing and correspondingly hazard curves are increasing. Whenever a person applies to the health institutes among SSI population, the surviving probability for the rest of the group decreases. That means as time passes, the probability of survival is decreasing and respectively the risk of hazard is increasing for both genders. Time to first claim around 200 days has a significant decay, concluding to many of the applications occur at that time. The downward step at about 200 days can also be seen for female, however, the decrease for male has higher rate than for female. This effect can also be seen in hazard functions.

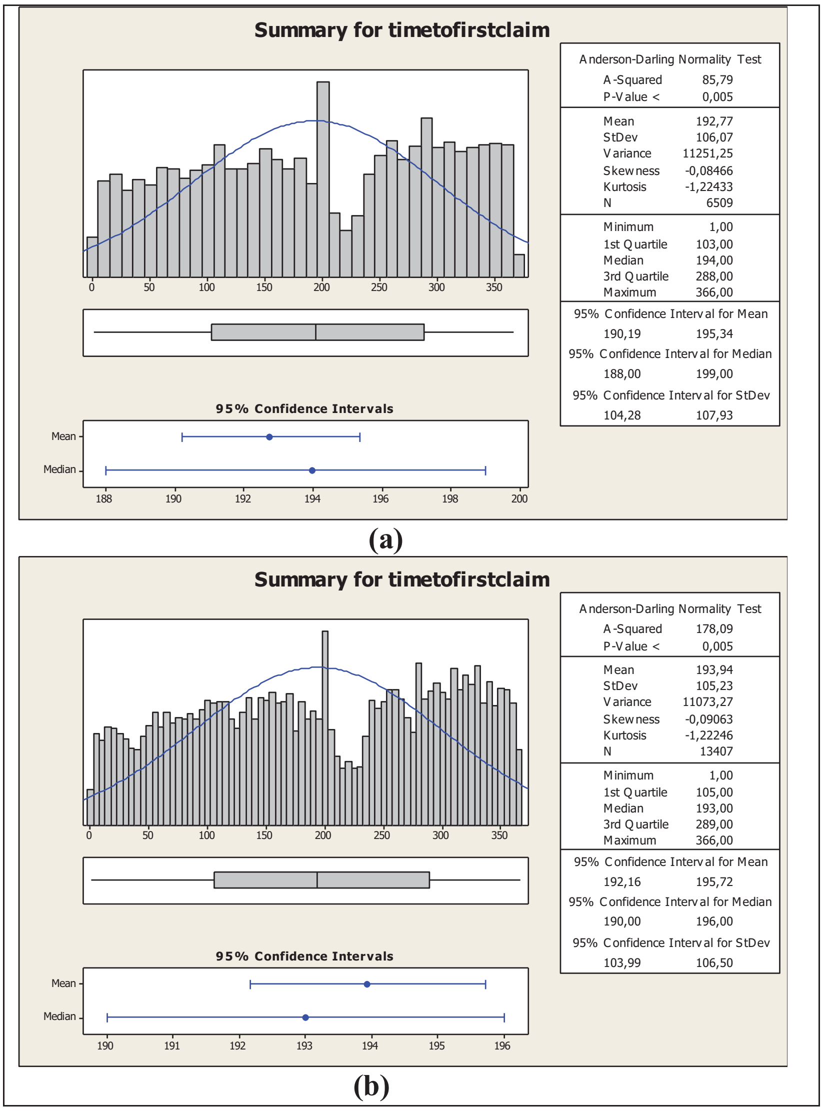


Figure 5.1: Descriptive statistics and frequency histograms for poisoning (2008) : (a) Male (b) Female

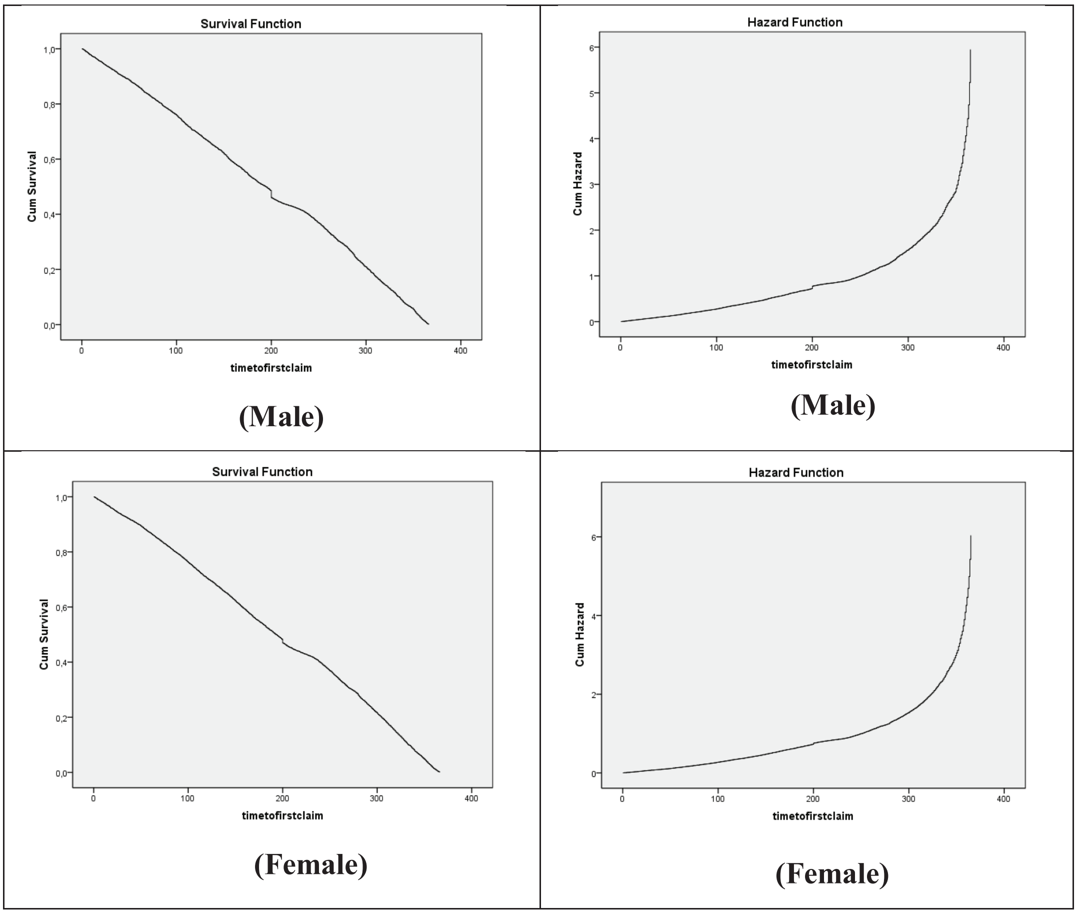


Figure 5.2: Plots of survival probabilities and hazard rates for male and female for poisoning

The next step is the construction of life tables for both genders. The life tables for total (male and female) are also performed and presented in Appendix (B1-B3). For each gender three tables are constructed. The first one represents the number of people at risk and the number of people who failed in each specified period and the conditional probability of failure for each interval (Table 5.1). Second table represents the survival probabilities in each interval (Table 5.2). And the third table represents the hazard and density estimates for each interval (Table 5.3). It should be noted that in Table 5.3 the intervals for the last table is the half of the intervals taken for the other two life tables because the hazard estimates are obtained with respect to the half of the interval.

Table 5.1: Life table of poisoning for both genders (2008)

Interval		MALE				FEMALE			
lower	upper	number entering	number failed	conditional probability of failure	standard error	number entering	number failed	conditional probability of failure	standard error
0	6	6509	91	0,01398	0,0014553	13407	170	0,01268	0,0009663
7	13	6418	104	0,0162	0,001576	13237	214	0,01617	0,0010962
14	20	6314	106	0,01679	0,0016169	13023	205	0,01574	0,0010907
21	27	6208	113	0,0182	0,0016967	12818	216	0,01685	0,0011369
28	34	6095	96	0,01575	0,0015948	12602	180	0,01428	0,001057
35	41	5999	115	0,01917	0,0017704	12422	179	0,01441	0,0010693
42	48	5884	83	0,01411	0,0015374	12243	194	0,01585	0,0011286
49	55	5801	109	0,01879	0,0017828	12049	241	0,02	0,0012755
56	62	5692	115	0,0202	0,0018649	11808	237	0,02007	0,0012906
63	69	5577	129	0,02313	0,0020129	11571	233	0,02014	0,0013058
70	76	5448	111	0,02037	0,0019141	11338	251	0,02214	0,0013818
77	83	5337	108	0,02024	0,0019274	11087	247	0,02228	0,0014017
84	90	5229	127	0,02429	0,0021288	10840	236	0,02177	0,0014017
91	97	5102	111	0,02176	0,0020424	10604	262	0,02471	0,0015075
98	104	4991	129	0,02585	0,0022461	10342	277	0,02678	0,0015876
105	111	4862	156	0,03209	0,0025274	10065	277	0,02752	0,0016307
112	118	4706	122	0,02592	0,0023165	9788	266	0,02718	0,0016435
119	125	4584	114	0,02487	0,0023001	9522	234	0,02457	0,0015866
126	132	4470	120	0,02685	0,0024175	9288	252	0,02713	0,0016858
133	139	4350	125	0,02874	0,002533	9036	272	0,0301	0,0017975
140	146	4225	105	0,02485	0,002395	8764	262	0,0299	0,0018191
147	153	4120	154	0,03738	0,0029552	8502	286	0,03364	0,0019554
154	160	3966	140	0,0353	0,0029303	8216	282	0,03432	0,0020085
161	167	3826	110	0,02875	0,0027016	7934	289	0,03643	0,0021033
168	174	3716	131	0,03525	0,0030253	7645	265	0,03466	0,0020921
175	181	3585	140	0,03905	0,0032354	7380	239	0,03238	0,0020606
182	188	3445	113	0,0328	0,0030346	7141	250	0,03501	0,0021751
189	195	3332	88	0,02641	0,0027779	6891	261	0,03788	0,0022996
196	202	3244	271	0,08354	0,004858	6630	387	0,05837	0,0028793
203	209	2973	81	0,02725	0,0029857	6243	186	0,02979	0,0021518
210	216	2892	61	0,02109	0,002672	6057	156	0,02576	0,0020353
217	223	2831	53	0,01872	0,0025474	5901	134	0,02271	0,0019393
224	230	2778	59	0,02124	0,0027355	5767	126	0,02185	0,001925
231	237	2719	88	0,03236	0,0033938	5641	189	0,0335	0,0023959
238	244	2631	116	0,04409	0,0040024	5452	272	0,04989	0,0029486
245	251	2515	131	0,05209	0,0044308	5180	262	0,05058	0,0030447
252	258	2384	143	0,05998	0,0048633	4918	308	0,06263	0,003455
259	265	2241	139	0,06203	0,0050952	4610	290	0,06291	0,0035759
266	272	2102	147	0,06993	0,0055627	4320	268	0,06204	0,0036701
273	279	1955	110	0,05627	0,0052116	4052	224	0,05528	0,0035901
280	286	1845	166	0,08997	0,0066617	3828	335	0,08751	0,0045673
287	293	1679	178	0,10602	0,0075132	3493	307	0,08789	0,0047906
294	300	1501	143	0,09527	0,0075779	3186	308	0,09667	0,0052354
301	307	1358	147	0,10825	0,008431	2878	302	0,10493	0,0057127
308	314	1211	152	0,12552	0,0095204	2576	321	0,12461	0,0065074
315	321	1059	137	0,12937	0,0103129	2255	340	0,15078	0,0075354
322	328	922	128	0,13883	0,0113872	1915	330	0,17232	0,0086302
329	335	794	149	0,18766	0,0138561	1585	298	0,18801	0,0098142
336	342	645	164	0,25426	0,0171457	1287	336	0,26107	0,0122431
343	349	481	98	0,20374	0,0183652	951	273	0,28707	0,0146698
350	356	383	179	0,46736	0,0254943	678	312	0,46018	0,0191414
357	363	204	147	0,72059	0,031416	366	275	0,75137	0,0225926
364	371	57	57	1	0	91	91	1	0

For example, in Table 5.1, it can be seen that 113 male applications to the health institutes between 21 and 27th days of 2008 has been observed. This information yields a conditional probability of 0.0182, which tells us 18.2 out of 1000 will apply to the health institutes in the specified time interval.

Table 5.2: Survival probabilities for poisoning per week per year (M, F) (2008)

	MALE				FEMALE			
time	survival probability	standard error	95% confidence Interval lower	upper	survival probability	standard error	95% confidence Interval lower	upper
7	0,986019	0,0014553	0,983167	0,988872	0,98732	0,0009663	0,985426	0,989214
14	0,970041	0,002113	0,9659	0,974183	0,971358	0,0014405	0,968535	0,974182
21	0,953756	0,0026031	0,948654	0,958858	0,956068	0,00177	0,952599	0,959537
28	0,936396	0,0030249	0,930467	0,942325	0,939957	0,0020517	0,935935	0,943978
35	0,921647	0,0033308	0,915119	0,928175	0,926531	0,0022533	0,922115	0,930947
42	0,903979	0,0036518	0,896822	0,911136	0,91318	0,0024318	0,908413	0,917946
49	0,891228	0,0038592	0,883664	0,898791	0,89871	0,0026057	0,893603	0,903817
56	0,874481	0,0041065	0,866433	0,88253	0,880734	0,0027991	0,875248	0,88622
63	0,856814	0,0043415	0,848305	0,865323	0,863057	0,0029691	0,857237	0,868876
70	0,836995	0,0045783	0,828022	0,845968	0,845678	0,00312	0,839563	0,851793
77	0,819942	0,0047626	0,810607	0,829276	0,826956	0,003267	0,820553	0,833359
84	0,803349	0,0049266	0,793693	0,813005	0,808533	0,0033981	0,801873	0,815193
91	0,783838	0,0051021	0,773838	0,793838	0,79093	0,003512	0,784047	0,797813
98	0,766784	0,0052415	0,756511	0,777058	0,771388	0,0036268	0,76428	0,778496
105	0,746966	0,0053887	0,736404	0,757527	0,750727	0,0037361	0,743405	0,75805
112	0,722999	0,0055469	0,712127	0,733871	0,730066	0,0038339	0,722552	0,737581
119	0,704256	0,0056567	0,693169	0,715343	0,710226	0,003918	0,702547	0,717905
126	0,686741	0,005749	0,675474	0,698009	0,692772	0,0039844	0,684963	0,700582
133	0,668305	0,0058358	0,656867	0,679743	0,673976	0,0040484	0,666042	0,681911
140	0,649101	0,0059155	0,637507	0,660695	0,653688	0,0041092	0,645635	0,661742
147	0,63297	0,0059743	0,62126	0,644679	0,634146	0,0041599	0,625993	0,6423
154	0,60931	0,0060475	0,597457	0,621163	0,612814	0,0042069	0,604569	0,62106
161	0,587802	0,0061011	0,575843	0,59976	0,59178	0,0042448	0,583461	0,6001
168	0,570902	0,0061348	0,558878	0,582926	0,570225	0,0042754	0,561845	0,578604
175	0,550776	0,0061654	0,538692	0,56286	0,550459	0,0042962	0,542038	0,558879
182	0,529267	0,0061868	0,517141	0,541393	0,532632	0,004309	0,524187	0,541078
189	0,511907	0,0061957	0,499763	0,52405	0,513985	0,0043165	0,505525	0,522445
196	0,498387	0,0061974	0,48624	0,510534	0,494518	0,004318	0,486055	0,502981
203	0,456752	0,0061742	0,444651	0,468853	0,465652	0,004308	0,457209	0,474096
210	0,444308	0,0061589	0,432237	0,456379	0,451779	0,0042981	0,443355	0,460203
217	0,434936	0,0061448	0,422893	0,44698	0,440143	0,0042872	0,431741	0,448546
224	0,426794	0,0061307	0,414778	0,43881	0,430148	0,0042759	0,421768	0,438529
231	0,417729	0,006113	0,405748	0,429711	0,42075	0,0042636	0,412394	0,429107
238	0,40421	0,0060827	0,392288	0,416131	0,406653	0,0042423	0,398339	0,414968
245	0,386388	0,0060353	0,374559	0,398217	0,386365	0,0042052	0,378123	0,394607
252	0,366262	0,0059716	0,354558	0,377966	0,366823	0,0041622	0,358665	0,374981
259	0,344293	0,0058893	0,33275	0,355835	0,34385	0,0041022	0,33581	0,35189
266	0,322937	0,0057958	0,311578	0,334297	0,32222	0,004036	0,314309	0,33013
273	0,300353	0,005682	0,289217	0,31149	0,30223	0,0039661	0,294457	0,310004
280	0,283454	0,0055861	0,272505	0,294402	0,285522	0,0039008	0,277877	0,293168
287	0,257951	0,0054228	0,247322	0,268579	0,260536	0,0037908	0,253106	0,267965
294	0,230604	0,005221	0,220371	0,240837	0,237637	0,003676	0,230432	0,244842
301	0,208634	0,0050364	0,198763	0,218505	0,214664	0,003546	0,207714	0,221614
308	0,18605	0,0048234	0,176596	0,195504	0,192138	0,0034026	0,185469	0,198807
315	0,162698	0,0045748	0,153731	0,171664	0,168196	0,0032304	0,161864	0,174527
322	0,14165	0,004322	0,133179	0,150121	0,142836	0,0030219	0,136913	0,148759
329	0,121985	0,0040565	0,114034	0,129935	0,118222	0,0027884	0,112757	0,123687
336	0,099094	0,0037034	0,091835	0,106352	0,095995	0,0025442	0,091008	0,100981
343	0,073898	0,0032426	0,067542	0,080253	0,070933	0,0022171	0,066588	0,075278
350	0,058842	0,0029169	0,053125	0,064559	0,050571	0,0018924	0,046862	0,05428
357	0,031341	0,0021597	0,027108	0,035574	0,027299	0,0014073	0,024541	0,030058
364	0,008757	0,0011548	0,006494	0,011021	0,006787	0,0007091	0,005398	0,008177

This table (5.2) illustrates that on the 28th day of 2008, 93.6% of the group will survive with given confidence interval of 95%.

Table 5.3: Hazard rates for poisoning per week per year (M, F) (2008)

	MALE				FEMALE			
time	hazard estimates	Standard error	density estimates	standard error	hazard estimates	Standard error	density estimates	standard error
3,5	0,002011	0,0002108	0,0019972	0,0002079	0,001823	0,0001398	0,0018114	0,000138
10,5	0,002334	0,0002288	0,0022826	0,000222	0,002328	0,0001592	0,0022803	0,0001546
17,5	0,002419	0,0002349	0,0023264	0,0002241	0,002267	0,0001583	0,0021844	0,0001514
24,5	0,002624	0,0002469	0,0024801	0,0002313	0,002428	0,0001652	0,0023016	0,0001553
31,5	0,002268	0,0002315	0,002107	0,0002135	0,002055	0,0001532	0,001918	0,000142
38,5	0,002765	0,0002578	0,002524	0,0002333	0,002073	0,000155	0,0019073	0,0001416
45,5	0,002029	0,0002228	0,0018217	0,0001987	0,002282	0,0001638	0,0020672	0,0001473
52,5	0,00271	0,0002595	0,0023923	0,0002272	0,002886	0,0001859	0,002568	0,0001639
59,5	0,002916	0,0002719	0,002524	0,0002333	0,002896	0,0001881	0,0025253	0,0001626
66,5	0,003343	0,0002943	0,0028312	0,0002468	0,002906	0,0001904	0,0024827	0,0001612
73,5	0,002941	0,0002791	0,0024362	0,0002293	0,003198	0,0002018	0,0026745	0,0001672
80,5	0,00292	0,000281	0,0023703	0,0002262	0,003218	0,0002048	0,0026319	0,0001659
87,5	0,003512	0,0003116	0,0027873	0,0002449	0,003144	0,0002047	0,0025147	0,0001622
94,5	0,003142	0,0002982	0,0024362	0,0002293	0,003574	0,0002208	0,0027917	0,0001708
101,5	0,003741	0,0003293	0,0028312	0,0002468	0,003878	0,000233	0,0029515	0,0001755
108,5	0,004658	0,0003729	0,0034238	0,0002708	0,003986	0,0002395	0,0029515	0,0001755
115,5	0,003752	0,0003397	0,0026776	0,0002401	0,003936	0,0002413	0,0028343	0,0001721
122,5	0,003597	0,0003369	0,002502	0,0002323	0,003554	0,0002323	0,0024934	0,0001616
129,5	0,003887	0,0003548	0,0026337	0,0002382	0,003929	0,0002475	0,0026852	0,0001676
136,5	0,004165	0,0003725	0,0027435	0,000243	0,004366	0,0002647	0,0028983	0,0001739
143,5	0,003595	0,0003508	0,0023045	0,0002231	0,004336	0,0002678	0,0027917	0,0001708
150,5	0,005442	0,0004384	0,0033799	0,0002691	0,004888	0,000289	0,0030474	0,0001783
157,5	0,005133	0,0004338	0,0030727	0,0002569	0,004989	0,000297	0,0030048	0,000177
164,5	0,004167	0,0003973	0,0024142	0,0002282	0,0053	0,0003117	0,0030794	0,0001792
171,5	0,005126	0,0004478	0,0028751	0,0002487	0,005039	0,0003095	0,0028237	0,0001717
178,5	0,00569	0,0004808	0,0030727	0,0002569	0,004703	0,0003041	0,0025466	0,0001633
185,5	0,004764	0,0004481	0,0024801	0,0002313	0,00509	0,0003219	0,0026639	0,0001669
192,5	0,003823	0,0004075	0,0019314	0,0002045	0,005515	0,0003413	0,0027811	0,0001705
199,5	0,012454	0,0007558	0,0059478	0,0003537	0,008589	0,0004364	0,0041236	0,0002066
206,5	0,003946	0,0004384	0,0017778	0,0001963	0,004321	0,0003168	0,0019819	0,0001443
213,5	0,003045	0,0003899	0,0013388	0,0001706	0,003727	0,0002984	0,0016622	0,0001323
220,5	0,0027	0,0003708	0,0011632	0,0001591	0,003281	0,0002834	0,0014278	0,0001227
227,5	0,003067	0,0003992	0,0012949	0,0001678	0,003156	0,0002811	0,0013426	0,000119
234,5	0,0047	0,0005009	0,0019314	0,0002045	0,004868	0,000354	0,0020139	0,0001455
241,5	0,006441	0,0005978	0,0025459	0,0002343	0,007309	0,0004431	0,0028983	0,0001739
248,5	0,00764	0,0006673	0,0028751	0,0002487	0,007413	0,0004578	0,0027917	0,0001708
255,5	0,008834	0,0007384	0,0031385	0,0002596	0,009236	0,000526	0,0032819	0,0001848
262,5	0,009144	0,0007752	0,0030507	0,000256	0,009279	0,0005446	0,0030901	0,0001795
269,5	0,010352	0,0008533	0,0032263	0,0002631	0,009146	0,0005584	0,0028557	0,0001727
276,5	0,008271	0,0007882	0,0024142	0,0002282	0,008122	0,0005424	0,0023868	0,0001581
283,5	0,013459	0,0010434	0,0036433	0,0002791	0,013074	0,0007136	0,0035696	0,0001926
290,5	0,015993	0,0011968	0,0039067	0,0002888	0,013133	0,0007487	0,0032712	0,0001845
297,5	0,014291	0,0011936	0,0031385	0,0002596	0,014512	0,0008258	0,0032819	0,0001848
304,5	0,016349	0,0013462	0,0032263	0,0002631	0,015821	0,000909	0,0032179	0,0001831
311,5	0,019132	0,0015483	0,003336	0,0002674	0,018985	0,0010573	0,0034204	0,0001886
318,5	0,019759	0,0016841	0,0030068	0,0002542	0,023296	0,0012592	0,0036228	0,000194
325,5	0,021312	0,0018785	0,0028093	0,0002459	0,026939	0,0014763	0,0035163	0,0001912
332,5	0,029584	0,0024106	0,0032702	0,0002648	0,029646	0,0017081	0,0031753	0,0001819
339,5	0,041614	0,0032148	0,0035994	0,0002775	0,042895	0,0023136	0,0035802	0,0001929
346,5	0,032407	0,0032525	0,0021509	0,0002156	0,047882	0,002857	0,0029089	0,0001743
353,5	0,087126	0,0062019	0,0039286	0,0002896	0,085386	0,0046131	0,0033245	0,000186
360,5	0,16092	0,0109671	0,0032263	0,0002631	0,171929	0,0082805	0,0029302	0,0001749
367,5	0,285714	0	0,001251	0,000165	0,285714	0	0,0009696	0,0001013

Additionally, the rate of failure for the days of 21 and 28th is found to be 0.0026. This illustrates the 26 in 10,000 rate of failure is expected for male applicants amoung SSI group.

In the next part, for investigation and calculation of the effects of sub classifications to the survival analysis the data set is divided into different sub categories according to the gender, age and location. And time to first application to health institutes data is compared in respect of specified characteristics.

According to the gender, the value of Gehan-Wilcoxon test statistics is 0,497 and the log-rank test statistics is 0,800 which concludes that there exists no significant difference in survival times between genders for poisoning. The results are presented in Table 5.4 and Figure 5.3.

Table 5.4: Wilcoxon (Gehan) and Log-rank test results for poisoning

Overall Comparisons[a]

Wilcoxon (Gehan) Statistic	df	Sig.
,462	1	,497

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	,064	1	,800

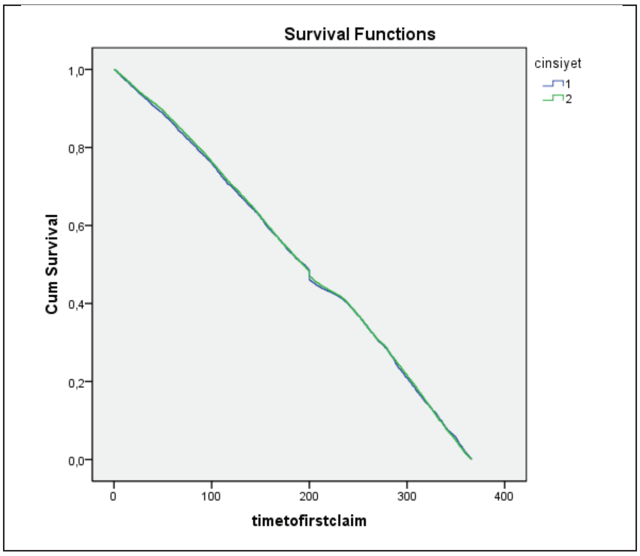


Figure 5.3: Comparison of two survival curves for poisoning

To illustrate the impact of location in Turkey, the comparison is done with respect to the three biggest cities; İstanbul, Ankara, and İzmir. According to the data set, 23% of the applications appears from Ankara, 70% of applications from İstanbul and 7% of applications from İzmir. It is reasonable to have İstanbul leading as it has the highest population among the other two. Table 5.5 shows means and medians for survival times for the three cities. The median time is the lowest for Ankara which means the time until half of the population applies to the hospitals is 171 days.

Table 5.5: Means and medians for survival time for the selected cities for poisoning

Means and Medians for Survival Time

ikamet_ili	Mean[a]				Median			
	Estimate	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound	Estimate	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
6	182,848	2,366	178,210	187,486	171,000	3,363	164,408	177,592
34	198,723	1,352	196,073	201,372	200,000	1,286	197,479	202,521
35	200,866	4,531	191,985	209,748	205,000	11,608	182,248	227,752
Overall	195,142	1,139	192,910	197,374	197,000	1,336	194,382	199,618

a. Estimation is limited to the largest survival time if it is censored.

The log-rank test shows that the result of comparison of survival times for the three cities are significantly different as shown in Table 5.6.

Table 5.6: Log-rank test results for the selected cities for poisoning

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	13,551	1	,000

This result can also be seen from the survival and hazard curves. Survival curves show the proportion of surviving, e.g. the proportion of people who has not applied to hospitals from poisoning. For example, in Figure 5.4, in the 160th day, the proportion of people surviving is the smallest in Ankara than the other cities. That means, the people in Ankara are more exposed to this morbidity than the other two.

The decrease in survival proportions is the slowest for İzmir. The hazard rates which focuses on the risk, on the contrary to survival curves, show increasing risk for the investigated populations.

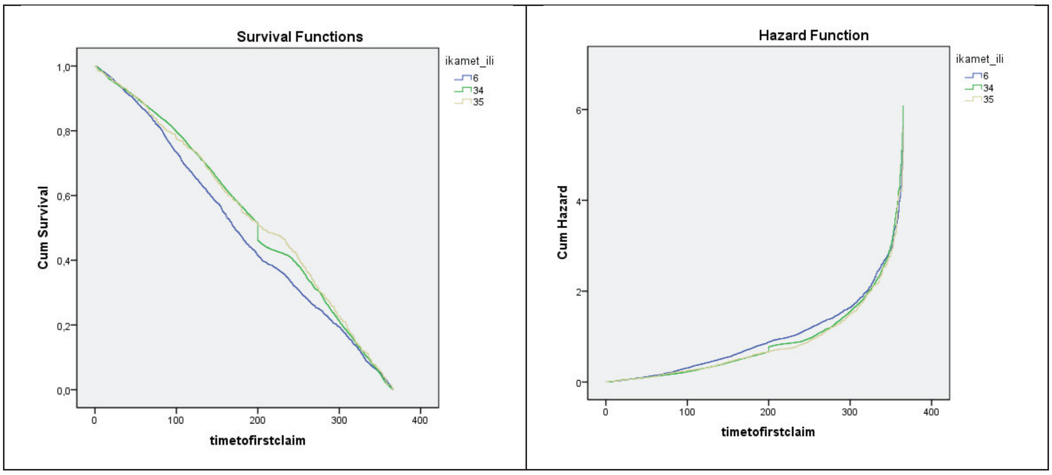


Figure 5.4: Survival and hazard curves of three cities for poisoning

Lastly, the comparison of survival curves with respect to age is done. Table 5.7 shows median survival times and the number of cases for each age with respect to both genders. The last column in the table shows the p-values obtained from the log-rank test with respect to each age. According to this test statistics only for the ages 18 and 23, there exists statistically significant difference between genders. For these ages survival and hazard curves also represented to see this difference graphically. For the ages 88, 90 and 91 among female applicants and for the ages 94 and 99 amoung male applicants, there exists not complete information for comparison.

Table 5.7: Number of events, median survival times and the test results for poisoning (1: Male, 2: Female)

age	gender	Number of Events	median survival time	p*value	age	gender	Number of Events	median survival time	p*value	age	gender	Number of Events	median survival time	p*value
0	1	104	275,000	0,277	32	1	61	192,000	,425	64	1	29	187,000	,613
	2	62	293,000			2	179	163,000			2	43	222,000	
1	1	337	242,000	0,987	33	1	76	200,000	,638	65	1	32	153,000	,785
	2	347	248,000			2	188	195,000			2	42	164,000	
2	1	624	214,000	0,431	34	1	90	176,000	,223	66	1	32	153,000	,670
	2	606	200,000			2	181	190,000			2	33	183,000	
3	1	640	173,000	0,34	35	1	62	163,000	,114	67	1	23	256,000	,400
	2	501	160,000			2	163	200,000			2	33	171,000	
4	1	324	168,000	0,946	36	1	59	225,000	,987	68	1	30	237,000	,173
	2	269	174,000			2	182	197,000			2	50	206,000	
5	1	157	159,000	0,835	37	1	74	200,000	,472	69	1	28	149,000	,713
	2	150	154,000			2	152	200,000			2	34	155,000	
6	1	104	171,000	0,491	38	1	58	158,000	,074	70	1	22	106,000	,456
	2	100	156,000			2	128	183,000			2	26	190,000	
7	1	86	185,000	0,322	39	1	43	208,000	,762	71	1	29	200,000	,325
	2	58	147,000			2	102	196,000			2	31	132,000	
8	1	70	185,000	0,45	40	1	45	193,000	,815	72	1	25	246,000	1,000
	2	64	160,000			2	103	199,000			2	25	226,000	
9	1	54	176,000	0,78	41	1	37	200,000	,876	73	1	18	155,000	,907
	2	50	168,000			2	90	200,000			2	30	182,000	
10	1	52	170,000	0,644	42	1	62	200,000	,315	74	1	28	138,000	,365
	2	52	169,000			2	94	163,000			2	36	179,000	
11	1	45	191,000	,966	43	1	58	173,000	,773	75	1	37	166,000	,821
	2	59	177,000			2	110	170,000			2	42	157,000	
12	1	52	240,000	,294	44	1	42	190,000	,340	76	1	28	249,000	,988
	2	88	195,000			2	119	209,000			2	50	252,000	
13	1	46	199,000	,336	45	1	52	197,000	,659	77	1	16	142,000	,831
	2	152	249,000			2	91	200,000			2	27	168,000	
14	1	80	193,000	,512	46	1	49	191,000	,891	78	1	22	150,000	,062
	2	336	226,000			2	93	196,000			2	35	237,000	
15	1	93	224,000	,783	47	1	31	254,000	,490	79	1	11	182,000	,263
	2	492	211,000			2	80	234,000			2	24	260,000	
16	1	153	201,000	,287	48	1	59	169,000	,571	80	1	15	150,000	,933
	2	603	190,000			2	97	164,000			2	25	199,000	
17	1	232	180,000	,938	49	1	41	187,000	,849	81	1	15	185,000	,835
	2	593	184,000			2	68	190,000			2	21	189,000	
18	1	131	154,000	,010	50	1	39	184,000	,748	82	1	16	148,000	,452
	2	528	181,000			2	62	189,000			2	21	206,000	
19	1	87	216,000	,348	51	1	35	131,000	,126	83	1	8	158,000	,130
	2	501	186,000			2	66	210,000			2	17	146,000	
20	1	63	171,000	,300	52	1	46	217,000	,820	84	1	8	128,000	,633
	2	486	182,000			2	65	242,000			2	14	251,000	
21	1	57	181,000	,984	53	1	56	179,000	,405	85	1	4	138,000	,651
	2	455	196,000			2	78	220,000			2	13	198,000	
22	1	57	206,000	,104	54	1	46	182,000	,753	86	1	6	342,000	,059
	2	415	193,000			2	64	203,000			2	9	257,000	
23	1	109	205,000	,034	55	1	40	162,000	,854	87	1	4	63,000	,113
	2	399	184,000			2	54	194,000			2	13	312,000	
24	1	102	177,000	,140	56	1	35	200,000	,499	88	2	4	175,000	
	2	361	211,000			2	57	178,000		89	1	4	142,000	,773
25	1	93	195,000	,837	57	1	40	200,000	,417		2	4	151,000	
	2	360	200,000			2	42	158,000		90	2	1	266,000	
26	1	96	205,000	,323	58	1	41	181,000	,768	91	2	3	126,000	
	2	299	190,000			2	57	200,000		92	1	1	196,000	1,000
27	1	85	197,000	,902	59	1	42	226,000	,724		2	4	123,000	
	2	347	204,000			2	50	197,000		93	1	3	201,000	1,000
28	1	97	173,000	,641	60	1	32	162,000	,437		2	2	101,000	
	2	300	185,000			2	43	210,000		94	1	1	175,000	
29	1	87	198,000	,286	61	1	31	233,000	,447	95	1	1	358,000	,317
	2	251	200,000			2	45	259,000			2	1	134,000	
30	1	88	200,000	,296	62	1	28	142,000	,811	99	1	1	200,000	
	2	228	190,000			2	29	176,000						
31	1	76	200,000	,910	63	1	21	179,000	,989					
	2	206	185,000			2	44	194,000						

The following figure represents the survival curves of the ages 18 and 23. For the 18 years old people the survival probabilities are higher for female than male. And for the 23 years old people the survival probabilities for male are higher.

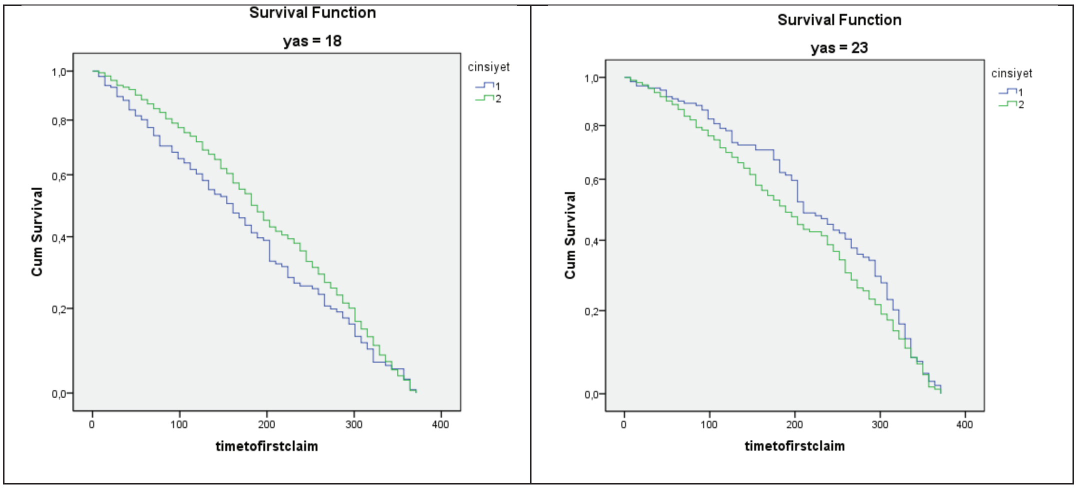


Figure 5.5: Survival curves for 18 and 23 ages of poisoning

5.3.2 RHD

The time to first application to the health institute for both genders is presented in the Figure 5.6, which shows 33% of the applications come from male and 67% of the applications come from female. The mean time to first claim is 190.87 days and the median is 189 days for male. The values between mean and median are close to each other indicating that the distribution to be symmetric. However, as in poisoning case, the lifetime data do not follow a suitable distribution.

The mean time to first claim is 185.69 days and the median is 178 days for female. The mean and median values are not close to each other and the distribution seems to be symmetric. However, the data do not follow a suitable distribution. As the mean is higher than median, the distribution is skewed to large values. Therefore, time to first claims for females accumulates at early times.

It is seen that the mean and median values for male take higher values than female for RHD, whereas these values do not show major differences between genders for poisoning.

The survival and hazard curves for RHD are seen in Figure 5.7. For both genders survival curves are decreasing and correspondingly hazard curves are increasing. There is a continuous decrease in survival curve till 200 days, thereafter, the decrease becomes slower for approximately 30 days. The same pattern also can be seen in the hazard curve. Female data follows the same structure in survival and hazard curves as male data in RHD.

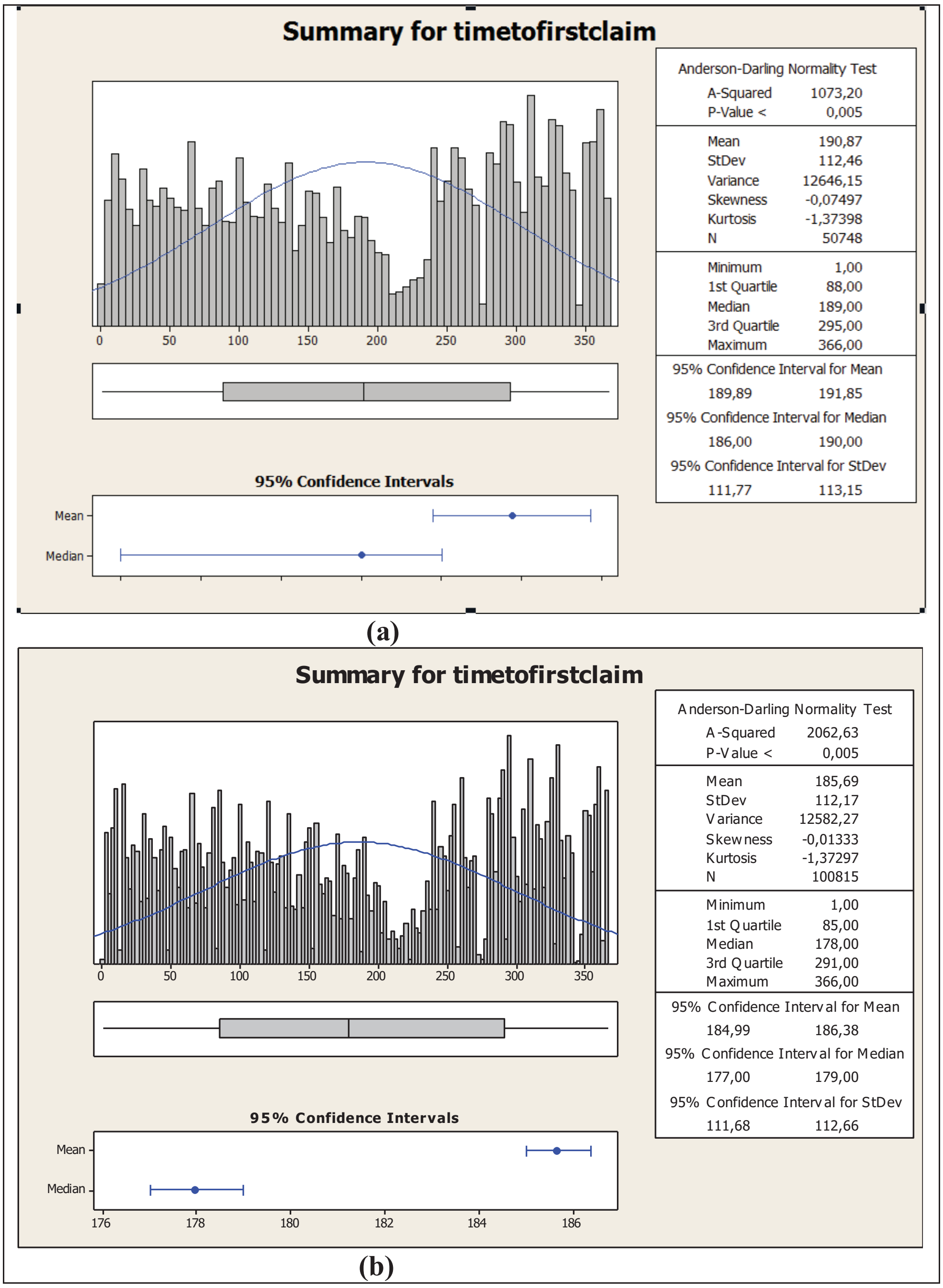


Figure 5.6: Descriptive statistics and frequency histograms for RHD (2008) : (a) Male (b) Female

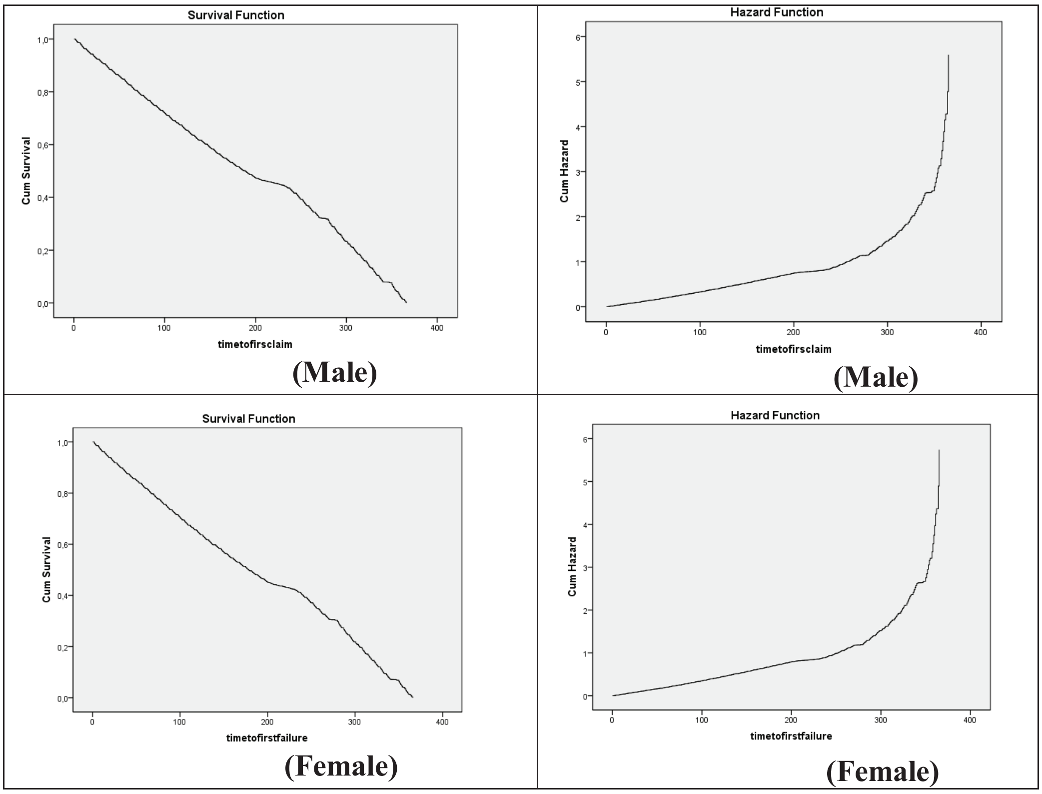


Figure 5.7: Survival probabilities and hazard rates for RHD (M, F)

The next step is the representation of life tables for both genders and for the overall data. For each gender three tables are constructed. The first one, represents the number of people at risk and the number of people who failed in each specified period and the conditional probability of failure for each interval (Table 5.8). Second table represents the survival probabilities in each interval (Table 5.9). And the third table represents the hazard and density estimates for each interval (Table 5.10). It should be noted again that the intervals for the last table is the half of the intervals taken for the other two life tables because the hazard estimates is obtained with respect to the half of the interval. The tables corresponding to the total are listed in Appendix (B4- B6).

Table 5.8: Life table of RHD for both genders (2008)

Interval		MALE				FEMALE			
lower	upper	number entering	number failed	conditional probability of failure	standard error	number entering	number failed	conditional probability of failure	standard error
0	6	50748	696	0,01371	0,0005163	100815	1506	0,01494	0,000382
7	13	50052	1180	0,02358	0,0006782	99309	2403	0,0242	0,0004876
14	20	48872	1011	0,02069	0,0006438	96906	2208	0,02278	0,0004793
21	27	47861	985	0,02058	0,000649	94698	2154	0,02275	0,0004845
28	34	46876	907	0,01935	0,0006362	92544	2057	0,02223	0,0004846
35	41	45969	1108	0,0241	0,0007153	90487	2240	0,02475	0,0005165
42	48	44861	960	0,0214	0,0006832	88247	1941	0,022	0,0004937
49	55	43901	874	0,01991	0,0006667	86306	1732	0,02007	0,0004773
56	62	43027	1109	0,02577	0,0007639	84574	2137	0,02527	0,0005396
63	69	41918	1061	0,02531	0,0007672	82437	2128	0,02581	0,0005523
70	76	40857	955	0,02337	0,0007475	80309	2001	0,02492	0,00055
77	83	39902	953	0,02388	0,0007644	78308	2078	0,02654	0,0005743
84	90	38949	999	0,02565	0,000801	76230	2136	0,02802	0,0005977
91	97	37950	942	0,02482	0,0007986	74094	1943	0,02622	0,0005871
98	104	37008	979	0,02645	0,0008342	72151	2091	0,02898	0,0006245
105	111	36029	1010	0,02803	0,0008696	70060	2092	0,02986	0,000643
112	118	35019	816	0,0233	0,0008062	67968	1751	0,02576	0,0006077
119	125	34203	1024	0,02994	0,0009215	66217	2043	0,03085	0,000672
126	132	33179	969	0,02921	0,0009244	64174	1975	0,03078	0,0006818
133	139	32210	941	0,02921	0,0009384	62199	1944	0,03125	0,0006977
140	146	31269	739	0,02363	0,000859	60255	1460	0,02423	0,0006264
147	153	30530	928	0,0304	0,0009825	58795	2006	0,03412	0,0007487
154	160	29602	957	0,03233	0,001028	56789	1738	0,0306	0,0007228
161	167	28645	826	0,02884	0,0009888	55051	1651	0,02999	0,0007269
168	174	27819	806	0,02897	0,0010056	53400	1656	0,03101	0,0007502
175	181	27013	807	0,02987	0,0010358	51744	1723	0,0333	0,0007887
182	188	26206	774	0,02954	0,0010458	50021	1559	0,03117	0,000777
189	195	25432	758	0,0298	0,0010663	48462	1525	0,03147	0,000793
196	202	24674	712	0,02886	0,0010657	46937	1482	0,03157	0,0008071
203	209	23962	405	0,0169	0,0008327	45455	889	0,01956	0,0006495
210	216	23557	291	0,01235	0,0007197	44566	537	0,01205	0,0005168
217	223	23266	276	0,01186	0,0007098	44029	550	0,01249	0,0005293
224	230	22990	333	0,01448	0,000788	43479	694	0,01596	0,000601
231	237	22657	554	0,02445	0,0010261	42785	1196	0,02795	0,0007969
238	244	22103	986	0,04461	0,0013886	41589	2005	0,04821	0,0010504
245	251	21117	1203	0,05697	0,001595	39584	2182	0,05512	0,0011471
252	258	19914	1254	0,06297	0,0017213	37402	2258	0,06037	0,0012315
259	265	18660	1211	0,0649	0,0018034	35144	2270	0,06459	0,0013112
266	272	17449	1132	0,06487	0,0018646	32874	2012	0,0612	0,001322
273	279	16317	179	0,01097	0,0008154	30862	283	0,00917	0,0005426
280	286	16138	1493	0,09251	0,0022809	30579	2995	0,09794	0,0016998
287	293	14645	1429	0,09758	0,0024521	27584	2837	0,10285	0,001829
294	300	13216	1394	0,10548	0,0026719	24747	2746	0,11096	0,0019966
301	307	11822	1119	0,09465	0,0026923	22001	2122	0,09645	0,0019902
308	314	10703	1359	0,12697	0,0032182	19879	2650	0,13331	0,0024108
315	321	9344	1254	0,1342	0,0035263	17229	2507	0,14551	0,0026864
322	328	8090	1386	0,17132	0,0041891	14722	2564	0,17416	0,0031256
329	335	6704	1421	0,21196	0,0049916	12158	2657	0,21854	0,0037479
336	342	5283	1259	0,23831	0,0058617	9501	2271	0,23903	0,0043755
343	349	4024	158	0,03926	0,0030618	7230	266	0,03679	0,0022139
350	356	3866	1650	0,4268	0,0079549	6964	2887	0,41456	0,0059034
357	363	2216	1517	0,68457	0,0098714	4077	2793	0,68506	0,0072746
364	371	699	699	1	0	1284	1284	1	0

It can be observed that the number of male applicants to the health institutes from RHD between 14^{th} and 20^{th} days of the year is 1011 with conditional probability of 0.02069. This means, 20 males in 1000 will apply to health institutes at specified time interval.

Table 5.9: Survival probabilities for RHD per week per year (M, F) (2008)

	MALE				FEMALE			
time	survival probability	standard error	95% confidence Interval lower	upper	survival probability	standard error	95% confidence Interval lower	upper
7	0,986285	0,0005163	0,985273	0,987297	0,985062	0,000382	0,984313	0,985811
14	0,963033	0,0008376	0,961391	0,964675	0,961226	0,000608	0,960034	0,962418
21	0,943111	0,0010282	0,941096	0,945126	0,939325	0,0007519	0,937851	0,940798
28	0,923701	0,0011785	0,921392	0,926011	0,917959	0,0008643	0,916265	0,919653
35	0,905829	0,0012965	0,903288	0,90837	0,897555	0,000955	0,895683	0,899427
42	0,883995	0,0014215	0,881209	0,886782	0,875336	0,0010404	0,873297	0,877375
49	0,865078	0,0015166	0,862106	0,868051	0,856083	0,0011055	0,853916	0,85825
56	0,847856	0,0015943	0,844731	0,850981	0,838903	0,0011578	0,836634	0,841172
63	0,826003	0,0016829	0,822705	0,829301	0,817706	0,001216	0,815322	0,820089
70	0,805096	0,0017584	0,801649	0,808542	0,796598	0,0012678	0,794113	0,799082
77	0,786277	0,0018197	0,782711	0,789844	0,776749	0,0013115	0,774179	0,77932
84	0,767498	0,0018752	0,763823	0,771174	0,756137	0,0013524	0,753487	0,758788
91	0,747813	0,0019277	0,744034	0,751591	0,73495	0,00139	0,732226	0,737675
98	0,72925	0,0019725	0,725384	0,733116	0,715677	0,0014207	0,712893	0,718462
105	0,709959	0,0020144	0,706011	0,713907	0,694936	0,0014501	0,692094	0,697778
112	0,690057	0,0020529	0,686033	0,69408	0,674185	0,0014761	0,671292	0,677078
119	0,673977	0,0020808	0,669899	0,678056	0,656817	0,0014953	0,653886	0,659748
126	0,653799	0,0021119	0,64966	0,657938	0,636552	0,0015149	0,633583	0,639521
133	0,634705	0,0021375	0,630515	0,638894	0,616962	0,001531	0,613961	0,619963
140	0,616162	0,0021588	0,611931	0,620393	0,597679	0,0015444	0,594652	0,600706
147	0,6016	0,0021732	0,597341	0,605859	0,583197	0,0015528	0,580154	0,58624
154	0,583314	0,0021885	0,579024	0,587603	0,563299	0,0015621	0,560238	0,566361
161	0,564456	0,002201	0,560142	0,56877	0,54606	0,001568	0,542986	0,549133
168	0,548179	0,0022092	0,543849	0,552509	0,529683	0,001572	0,526602	0,532764
175	0,532297	0,0022149	0,527956	0,536638	0,513257	0,0015742	0,510172	0,516342
182	0,516395	0,0022183	0,512047	0,520743	0,496166	0,0015747	0,49308	0,499253
189	0,501143	0,0022195	0,496793	0,505493	0,480702	0,0015736	0,477618	0,483786
196	0,486206	0,0022187	0,481858	0,490555	0,465576	0,001571	0,462496	0,468655
203	0,472176	0,0022161	0,467833	0,47652	0,450875	0,0015671	0,447804	0,453947
210	0,464196	0,0022138	0,459857	0,468535	0,442057	0,0015641	0,438992	0,445123
217	0,458461	0,0022119	0,454126	0,462797	0,436731	0,0015621	0,433669	0,439792
224	0,453023	0,0022097	0,448692	0,457354	0,431275	0,0015598	0,428218	0,434332
231	0,446461	0,0022068	0,442136	0,450786	0,424391	0,0015566	0,42134	0,427442
238	0,435544	0,002201	0,43123	0,439858	0,412528	0,0015504	0,409489	0,415567
245	0,416115	0,0021881	0,411826	0,420403	0,39264	0,001538	0,389626	0,395654
252	0,39241	0,0021675	0,388161	0,396658	0,370996	0,0015214	0,368014	0,373978
259	0,367699	0,0021404	0,363504	0,371894	0,348599	0,0015008	0,345657	0,35154
266	0,343836	0,0021085	0,339704	0,347969	0,326082	0,0014764	0,323189	0,328976
273	0,32153	0,0020733	0,317466	0,325594	0,306125	0,0014515	0,30328	0,30897
280	0,318003	0,0020673	0,313951	0,322054	0,303318	0,0014478	0,30048	0,306156
287	0,288583	0,0020114	0,284641	0,292525	0,27361	0,0014041	0,270858	0,276362
294	0,260424	0,0019482	0,256606	0,264242	0,245469	0,0013554	0,242813	0,248126
301	0,232955	0,0018764	0,229277	0,236633	0,218231	0,0013009	0,215682	0,220781
308	0,210905	0,0018109	0,207356	0,214454	0,197183	0,0012531	0,194727	0,199639
315	0,184125	0,0017205	0,180753	0,187498	0,170897	0,0011855	0,168574	0,173221
322	0,159415	0,001625	0,15623	0,1626	0,14603	0,0011122	0,14385	0,14821
329	0,132104	0,0015031	0,129158	0,13505	0,120597	0,0010257	0,118587	0,122607
336	0,104103	0,0013557	0,101446	0,10676	0,094242	0,0009202	0,092438	0,096045
343	0,079294	0,0011994	0,076943	0,081645	0,071716	0,0008126	0,070123	0,073308
350	0,07618	0,0011776	0,073872	0,078488	0,069077	0,0007987	0,067512	0,070642
357	0,043667	0,0009071	0,041889	0,045445	0,04044	0,0006204	0,039224	0,041656
364	0,013774	0,0005174	0,01276	0,014788	0,012736	0,0003532	0,012044	0,013428

For the same interval the survival probability is found to be 0.9431 with 95% confidence interval. It should be noted that the survival probabilities decrease with respect to increase in time period through the end of the year.

Table 5.10: Hazard rates for RHD per week per year (M, F) (2008)

	MALE				FEMALE			
time	hazard estimates	Standard error	density estimates	standard error	hazard estimates	Standard error	density estimates	standard error
3,5	0,001973	0,0000748	0,0019593	0,0000738	0,00215	0,0000554	0,002134	0,0000546
10,5	0,003408	0,0000992	0,0033217	0,0000956	0,003499	0,0000714	0,0034051	0,0000686
17,5	0,002986	0,0000939	0,002846	0,0000886	0,003293	0,0000701	0,0031288	0,0000659
24,5	0,002971	0,0000946	0,0027728	0,0000875	0,003287	0,0000708	0,0030523	0,0000651
31,5	0,002791	0,0000927	0,0025532	0,000084	0,003211	0,0000708	0,0029148	0,0000636
38,5	0,003485	0,0001047	0,0031191	0,0000927	0,003581	0,0000757	0,0031741	0,0000663
45,5	0,00309	0,0000997	0,0027024	0,0000864	0,003177	0,0000721	0,0027504	0,0000618
52,5	0,002873	0,0000972	0,0024603	0,0000825	0,002896	0,0000696	0,0024543	0,0000585
59,5	0,00373	0,000112	0,0031219	0,0000927	0,003656	0,0000791	0,0030282	0,0000648
66,5	0,003662	0,0001124	0,0029867	0,0000907	0,003736	0,000081	0,0030154	0,0000647
73,5	0,003379	0,0001093	0,0026884	0,0000862	0,003604	0,0000806	0,0028355	0,0000628
80,5	0,003453	0,0001119	0,0026827	0,0000861	0,003842	0,0000843	0,0029446	0,0000639
87,5	0,003712	0,0001174	0,0028122	0,0000881	0,00406	0,0000878	0,0030268	0,0000648
94,5	0,003591	0,000117	0,0026518	0,0000856	0,003796	0,0000861	0,0027533	0,0000619
101,5	0,00383	0,0001224	0,0027559	0,0000872	0,004201	0,0000919	0,002963	0,0000641
108,5	0,004062	0,0001278	0,0028432	0,0000886	0,00433	0,0000947	0,0029644	0,0000641
115,5	0,003368	0,0001179	0,0022971	0,0000798	0,003728	0,0000891	0,0024812	0,0000588
122,5	0,004342	0,0001357	0,0028826	0,0000892	0,004477	0,000099	0,002895	0,0000634
129,5	0,004234	0,000136	0,0027278	0,0000868	0,004465	0,0001005	0,0027986	0,0000624
136,5	0,004235	0,0001381	0,0026489	0,0000855	0,004536	0,0001029	0,0027547	0,0000619
143,5	0,003417	0,0001257	0,0020803	0,000076	0,003504	0,0000917	0,0020689	0,0000538
150,5	0,004409	0,0001447	0,0026123	0,000085	0,004959	0,0001107	0,0028425	0,0000628
157,5	0,004694	0,0001517	0,002694	0,0000863	0,00444	0,0001065	0,0024628	0,0000586
164,5	0,00418	0,0001454	0,0023252	0,0000802	0,00435	0,000107	0,0023395	0,0000571
171,5	0,0042	0,0001479	0,0022689	0,0000793	0,0045	0,0001106	0,0023466	0,0000572
178,5	0,004333	0,0001525	0,0022717	0,0000793	0,004837	0,0001165	0,0024415	0,0000583
185,5	0,004283	0,0001539	0,0021788	0,0000777	0,004523	0,0001145	0,0022091	0,0000555
192,5	0,004322	0,000157	0,0021338	0,0000769	0,004567	0,0001169	0,002161	0,0000549
199,5	0,004183	0,0001567	0,0020043	0,0000746	0,004583	0,000119	0,0021	0,0000541
206,5	0,002435	0,000121	0,0011401	0,0000564	0,002822	0,0000946	0,0012597	0,0000421
213,5	0,001776	0,0001041	0,0008192	0,0000479	0,001732	0,0000747	0,0007609	0,0000327
220,5	0,001705	0,0001026	0,0007769	0,0000466	0,001796	0,0000766	0,0007794	0,0000331
227,5	0,002084	0,0001142	0,0009374	0,0000512	0,002299	0,0000873	0,0009834	0,0000372
234,5	0,003536	0,0001502	0,0015595	0,0000659	0,00405	0,0001171	0,0016948	0,0000487
241,5	0,006518	0,0002075	0,0027756	0,0000875	0,007057	0,0001576	0,0028411	0,0000628
248,5	0,008377	0,0002414	0,0033865	0,0000965	0,008098	0,0001733	0,0030919	0,0000655
255,5	0,009288	0,0002622	0,00353	0,0000984	0,008893	0,0001871	0,0031996	0,0000666
262,5	0,009582	0,0002752	0,003409	0,0000968	0,009535	0,0002	0,0032166	0,0000667
269,5	0,009579	0,0002845	0,0031866	0,0000936	0,009019	0,000201	0,002851	0,0000629
276,5	0,001576	0,0001178	0,0005039	0,0000376	0,001316	0,0000782	0,000401	0,0000238
283,5	0,013857	0,0003582	0,0042028	0,0001072	0,014712	0,0002685	0,004244	0,0000764
290,5	0,014654	0,0003872	0,0040227	0,0001049	0,015489	0,0002904	0,0040201	0,0000744
297,5	0,015907	0,0004254	0,0039242	0,0001036	0,016783	0,0003197	0,0038911	0,0000732
304,5	0,014194	0,0004238	0,00315	0,0000931	0,014477	0,0003139	0,0030069	0,0000646
311,5	0,019369	0,0005242	0,0038256	0,0001024	0,020404	0,0003953	0,0037551	0,000072
318,5	0,020551	0,0005788	0,00353	0,0000984	0,022418	0,0004464	0,0035525	0,0000701
325,5	0,026768	0,0007158	0,0039016	0,0001034	0,027253	0,0005358	0,0036332	0,0000708
332,5	0,03387	0,0008922	0,0040002	0,0001046	0,03505	0,0006748	0,003765	0,0000721
339,5	0,03865	0,0010793	0,0035441	0,0000986	0,038782	0,0008063	0,0032181	0,0000668
346,5	0,005722	0,0004551	0,0004448	0,0000353	0,005354	0,0003282	0,0003769	0,0000231
353,5	0,077512	0,0018367	0,0046448	0,0001125	0,074709	0,001342	0,0040909	0,000075
360,5	0,148689	0,0032599	0,0042704	0,000108	0,148853	0,0024041	0,0039577	0,0000738
367,5	0,285714	0	0,0019677	0,0000739	0,285714	0	0,0018195	0,0000505

Additionally, the rate of failure for the day of 21- 28th is found to be 0.002971. This illustrates the 29 in 10,000 rate of failure is expected for male applicants amoung SSI group.

In the next part, for investigation and calculation of the effects of sub classifications to the survival analysis the data set is analyzed according to the gender, age and location. And time to first aplication to health institutes data is compared in respect of those specified characteristics.

According to the gender, the value of both the Gehan-Wilcoxon and the log-rank test statistics are equal to 0, which concludes that there exists significant difference in survival times between genders for RHD. These are presented in the Table 5.11 and Figure 5.8. The survival probabilities for male are larger than the survival probabilities for female, the conclusion is that, the lifetimes for male are higher than female. This shows the risk of RHD is higher for female when the time passes within a year.

Table 5.11: Gehan-Wilcoxon and log-rank test results for RHD

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	63,756	1	,000

Overall Comparisons

Wilcoxon (Gehan) Statistic	df	Sig.
71,745	1	,000

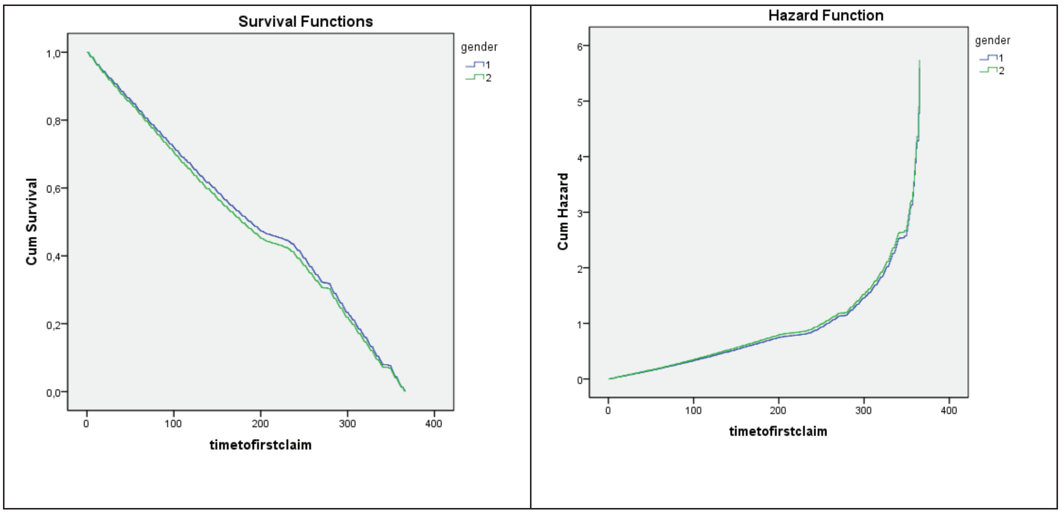


Figure 5.8: Comparison of survival and hazard curves for RHD

The comparison is done with respect to the three biggest cities in Turkey, İstanbul, Ankara, and İzmir. According to the application numbers, 27% of the applications appears from Ankara, 47% of applications from İstanbul and 26% of applications from İzmir. The highest application numbers to hospitals is seen in İstanbul. And following table shows means and medians for survival times for the three cities. The median time is the lowest for İstanbul, which means the time until half of the population applies to the hospitals is the smallest in İstanbul, and equals to 170 days (Table 5.12)

Table 5.12: Means and medians for survival time for the selected cities for RHD

Means and Medians for Survival Time

ikamet_ili	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
6	186,575	,905	184,802	188,349	172,000	1,642	168,781	175,219
34	181,226	,693	179,867	182,585	170,000	1,168	167,711	172,289
35	214,636	,950	212,774	216,499	252,000	1,009	250,022	253,978
Overall	191,330	,480	190,389	192,271	186,000	1,183	183,681	188,319

a. Estimation is limited to the largest survival time if it is censored.

The log-rank test shows the survival times for the three selected cities are significantly different as shown in Table 5.13.

Table 5.13: Log-rank test results for the selected cities for RHD

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	264,355	1	,000

This can be seen also from the survival and hazard curves respectively. Survival curves shows the proportion of surviving e.g. proportion of people who has not applied to hospitals from RHD. The survival patterns for Ankara and İstanbul seem identical and for İzmir there is a definite difference from the other cities. For example, in the 200^{th} day the proportion of people surviving is the smallest in İstanbul than the other cities. That means, people in İstanbul are more exposed to the morbidity than the other two. The decrease in survival proportions is the slowest for İzmir. The hazard rates which focuses on the risk, on the contrary to survival curves, there is an increasing risk for the investigated populations.

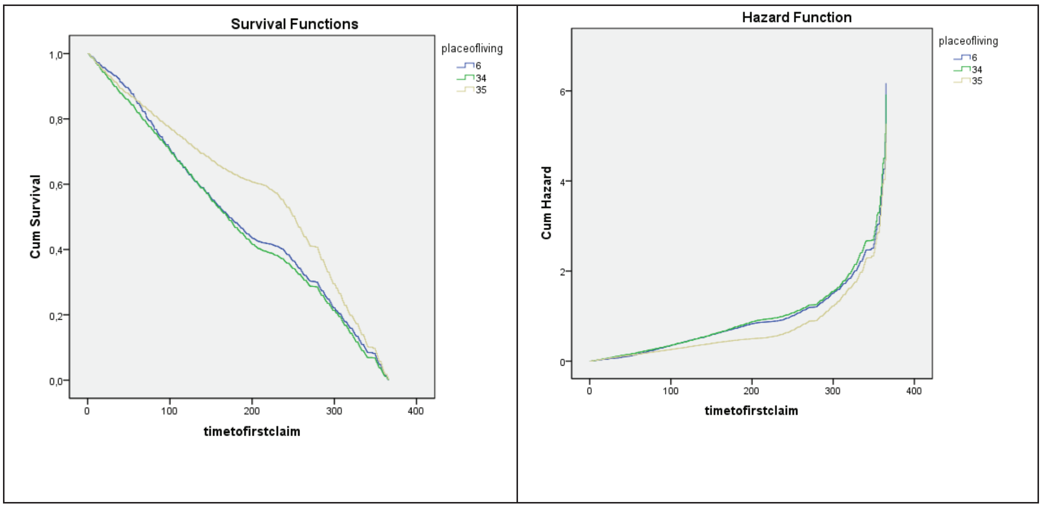


Figure 5.9: Survival and hazard curves of three cities for RHD

Lastly, the comparison of survival curves with respect to age is done. The Table 5.14 shows median survival times and the number of events for each age with respect to both genders. The last column in the table shows the p-values obtained from the log-rank tests with respect to each age. According to this test statistics for the ages 6, 9, 36, 38, 39, 49, 51, 52, 53, 54, 59, 62, 65, 67, and 75, there exists statistically significant difference between genders. For some of these ages survival and hazard curves also represented graphically to see the difference. For the age 99, there is only female applicant, therefore for this age the comparisons could not be done with respect to genders.

Table 5.14: Number of events, median survival times and the test results for RHD (1: Male, 2: Female)

age	gender	Number of Events	median survival time	p*value	age	gender	Number of Events	median survival time	p*value	age	gender	Number of Events	median survival time	p*value
0	1	82	294,000	,198	34	1	564	190,000	,378	68	1	1110	163,000	,713
	2	74	269,000			2	1458	182,000			2	1646	172,000	
1	1	96	129,000	,930	35	1	577	198,000	,141	69	1	879	185,000	,917
	2	59	137,000			2	1401	182,000			2	1495	185,000	
2	1	77	241,000	,257	36	1	555	190,000	,009	70	1	961	178,000	,689
	2	56	200,000			2	1468	175,000			2	1547	178,000	
3	1	91	186,000	,420	37	1	543	190,000	,113	71	1	860	189,000	,328
	2	64	245,000			2	1432	171,000			2	1303	186,000	
4	1	102	199,000	,464	38	1	582	192,000	,005	72	1	712	197,000	,228
	2	72	249,000			2	1475	170,000			2	1303	186,000	
5	1	120	207,000	,269	39	1	540	234,000	,000	73	1	574	175,000	,601
	2	91	170,000			2	1430	178,000			2	1026	180,000	
6	1	170	197,000	,018	40	1	560	170,000	,938	74	1	858	191,000	,763
	2	129	268,000			2	1414	179,000			2	1174	187,000	
7	1	193	179,000	,758	41	1	521	184,000	,501	75	1	1011	200,000	,012
	2	157	172,000			2	1243	171,000			2	1304	177,000	
8	1	217	179,000	,368	42	1	707	182,000	,295	76	1	764	179,000	,211
	2	186	193,000			2	1825	177,000			2	1120	192,000	
9	1	224	190,000	,030	43	1	733	176,000	,145	77	1	592	217,000	,551
	2	266	163,000			2	2064	170,000			2	919	197,000	
10	1	266	189,000	,714	44	1	746	186,000	,202	78	1	510	183,000	,308
	2	259	189,000			2	2049	179,000			2	966	186,000	
11	1	289	183,000	,205	45	1	722	190,000	,176	79	1	436	179,000	,452
	2	295	170,000			2	1969	176,000			2	712	190,000	
12	1	335	169,000	,213	46	1	769	186,000	,349	80	1	356	185,000	,399
	2	280	189,000			2	1999	175,000			2	701	175,000	
13	1	325	182,000	,923	47	1	755	194,000	,080	81	1	323	187,000	,713
	2	291	184,000			2	1870	176,000			2	585	189,000	
14	1	339	190,000	,795	48	1	952	176,000	,582	82	1	257	172,000	,607
	2	369	185,000			2	2562	172,000			2	561	171,000	
15	1	410	215,000	,836	49	1	827	199,000	,027	83	1	212	163,000	,603
	2	476	231,000			2	2066	175,000			2	340	171,000	
16	1	486	206,000	,793	50	1	793	181,000	,663	84	1	158	171,000	,157
	2	604	196,000			2	1898	179,000			2	331	213,000	
17	1	494	190,000	,841	51	1	777	198,000	,002	85	1	99	206,000	,739
	2	676	193,000			2	1835	175,000			2	189	176,000	
18	1	322	156,000	,006	52	1	1084	185,000	,031	86	1	97	236,000	,433
	2	661	193,000			2	2436	171,000			2	176	214,000	
19	1	302	242,000	,055	53	1	1120	191,000	,021	87	1	65	183,000	,771
	2	707	199,000			2	2617	172,000			2	91	205,000	
20	1	254	190,000	,972	54	1	1083	190,000	,003	88	1	34	226,000	,376
	2	751	190,000			2	2332	170,000			2	90	238,000	
21	1	248	197,000	,298	55	1	1001	197,000	,064	89	1	19	142,000	,145
	2	809	184,000			2	2160	176,000			2	68	178,000	
22	1	335	191,000	,319	56	1	1021	183,000	,357	90	1	11	162,000	,266
	2	832	214,000			2	1985	170,000			2	33	200,000	
23	1	351	199,000	,067	57	1	905	192,000	,581	91	1	10	220,000	,714
	2	881	189,000			2	1679	179,000			2	19	183,000	
24	1	363	207,000	,334	58	1	1298	175,000	,275	92	1	10	247,000	,550
	2	840	191,000			2	2484	172,000			2	37	239,000	
25	1	363	212,000	,167	59	1	1117	182,000	,002	93	1	11	121,000	,113
	2	989	196,000			2	2094	164,000			2	23	179,000	
26	1	460	178,000	,874	60	1	1072	178,000	,491	94	1	8	199,000	,930
	2	1000	183,000			2	2030	172,000			2	17	268,000	
27	1	513	185,000	,795	61	1	1020	168,000	,449	95	1	2	255,000	,130
	2	1176	189,000			2	1857	173,000			2	18	150,000	
28	1	529	197,000	,405	62	1	978	184,000	,001	96	1	2	122,000	,734
	2	1153	189,000			2	1686	164,000			2	13	169,000	
29	1	500	197,000	,518	63	1	1135	185,000	,518	97	1	1	23,000	,143
	2	1157	190,000			2	1855	178,000			2	5	144,000	
30	1	508	199,000	,512	64	1	1003	182,000	,344	98	1	1	50,000	,245
	2	1198	196,000			2	1678	168,000			2	8	240,000	
31	1	520	200,000	,068	65	1	911	193,000	,006	99	2	1	205,000	
	2	1179	181,000			2	1488	171,000						
32	1	501	182,000	,898	66	1	1073	184,000	,154					
	2	1155	183,000			2	1707	175,000						
33	1	483	191,000	,207	67	1	893	206,000	,014					
	2	1134	177,000			2	1422	182,000						

In the Figure 5.10 the survival curves of the ages 6, 36, 52 and 65 are illustrated. For the 36, 52 and 65 years old people the survival probabilities are higher for male than for female, whereas for the age 6 the survival probabilities are the other way around.

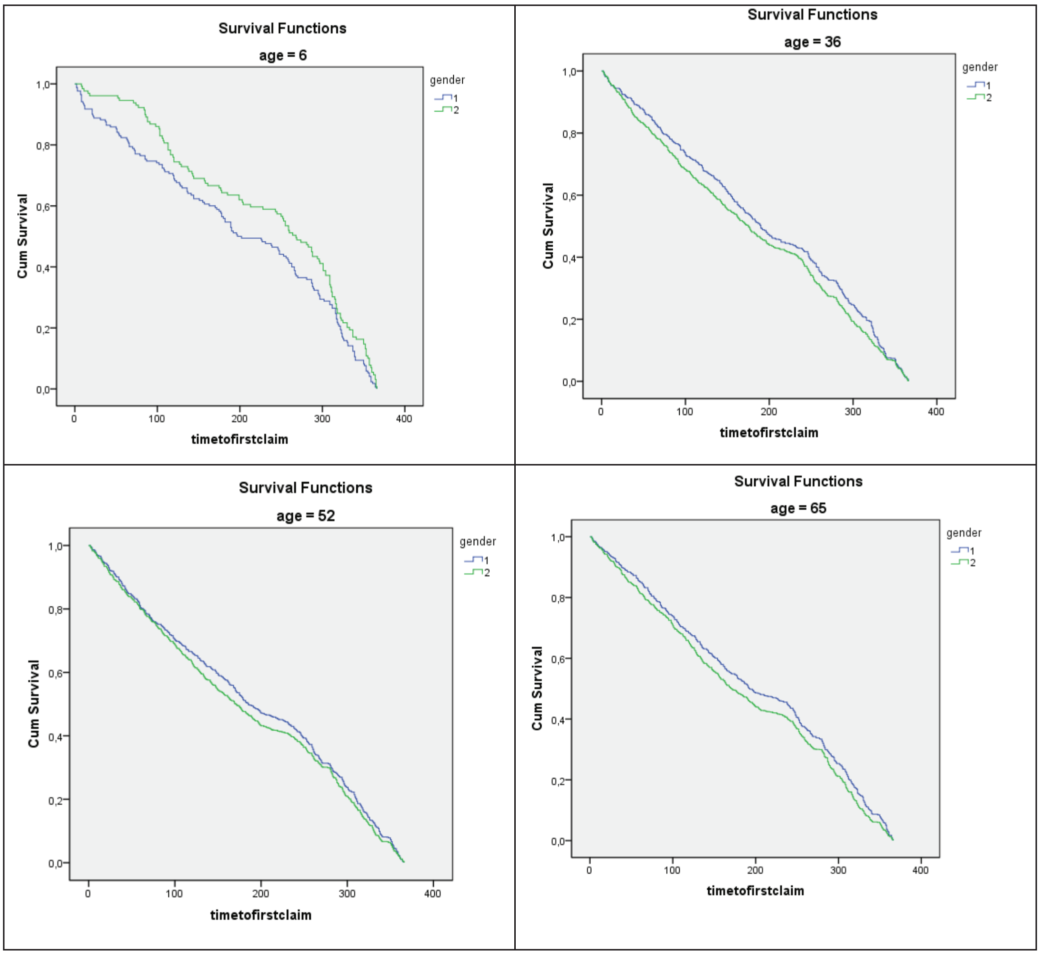


Figure 5.10: Survival curves for ages 6, 36, 52 and 65 for RHD

RHD morbidity has acute and chronic definitions in ICD-10 codings. In order to evaluate the impact of the type of the morbidity, RHD is re-analyzed with respect to this distinction.

5.3.2.1. Acute RHD

The survival and hazard curves for acute RHD are plotted in Figure 5.11. For both genders survival curves are decreasing and correspondingly hazard curves are increasing.

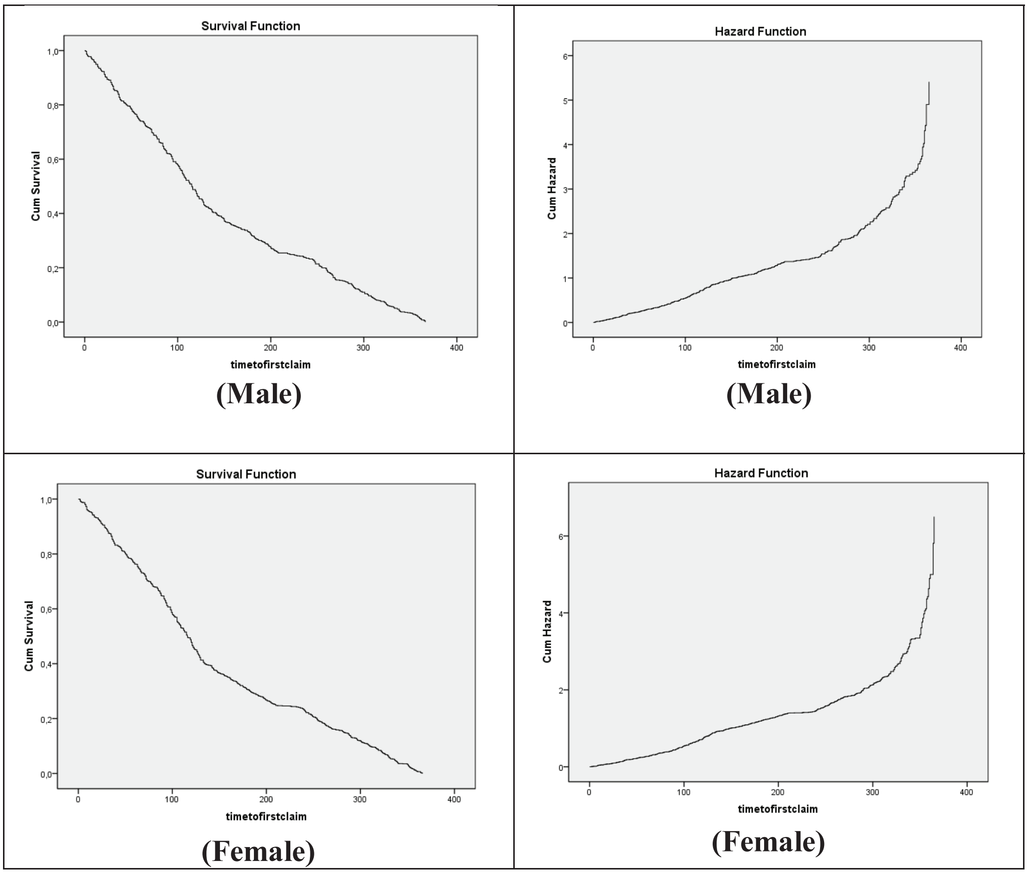


Figure 5.11: Survival probabilities and hazard rates for acute RHD (M, F)

The construction of life tables for both genders and for the total are performed. For each gender three tables are constructed. The first one, represents the number of people at risk and the number of people who failed in each specified period and the conditional probability of failure for each interval (Table 5.15). Second table represents the survival probabilities in each interval (Table 5.16). And the third table represents the hazard and density estimates for each interval (Table 5.17). It should be noted that the intervals for the last table is the half of the intervals taken for the other two life tables because the hazard estimates is obtained with respect to the half of the interval.

Table 5.15: Life table of acute RHD for both genders (2008)

Interval		MALE				FEMALE			
Lower	Upper	Number entering	Number failed	conditional probability of failure	standard error	Number entering	Number failed	conditional probability of failure	standard error
0	6	672	15	0,02232	0,005699	1333	17	0,01275	0,003073
7	13	657	19	0,02892	0,006538	1316	46	0,03495	0,005063
14	20	638	18	0,02821	0,006555	1270	28	0,02205	0,00412
21	27	620	21	0,03387	0,007265	1242	35	0,02818	0,004696
28	34	599	26	0,04341	0,008326	1207	43	0,03563	0,005335
35	41	573	26	0,04538	0,008695	1164	55	0,04725	0,006219
42	48	547	13	0,02377	0,006513	1109	29	0,02615	0,004792
49	55	534	20	0,03745	0,008216	1080	35	0,03241	0,005388
56	62	514	17	0,03307	0,007888	1045	29	0,02775	0,005081
63	69	497	16	0,03219	0,007918	1016	43	0,04232	0,006316
70	76	481	19	0,0395	0,008881	973	41	0,04214	0,006441
77	83	462	19	0,04113	0,009239	932	27	0,02897	0,005494
84	90	443	26	0,05869	0,011167	905	43	0,04751	0,007072
91	97	417	20	0,04796	0,010464	862	53	0,06148	0,008182
98	104	397	22	0,05542	0,011483	809	50	0,0618	0,008466
105	111	375	24	0,064	0,012639	759	54	0,07115	0,009331
112	118	351	23	0,06553	0,013208	705	46	0,06525	0,009301
119	125	328	23	0,07012	0,014099	659	58	0,08801	0,011036
126	132	305	20	0,06557	0,014174	601	50	0,08319	0,011265
133	139	285	14	0,04912	0,012802	551	25	0,04537	0,008866
140	146	271	9	0,03321	0,010885	526	25	0,04753	0,009277
147	153	262	15	0,05725	0,014353	501	18	0,03593	0,008315
154	160	247	8	0,03239	0,011264	483	16	0,03313	0,008143
161	167	239	6	0,0251	0,010119	467	19	0,04069	0,009142
168	174	233	6	0,02575	0,010377	448	21	0,04688	0,009986
175	181	227	14	0,06167	0,015967	427	21	0,04918	0,010465
182	188	213	11	0,05164	0,015164	406	18	0,04433	0,010216
189	195	202	8	0,0396	0,013722	388	15	0,03866	0,009787
196	202	194	13	0,06701	0,017952	373	21	0,0563	0,011935
203	209	181	10	0,05525	0,016982	352	17	0,0483	0,011427
210	216	171	0	0	0	335	6	0,01791	0,007246
217	223	171	4	0,02339	0,011558	329	1	0,00304	0,003035
224	230	167	3	0,01796	0,010278	328	3	0,00915	0,005256
231	237	164	4	0,02439	0,012045	325	7	0,02154	0,008053
238	244	160	4	0,025	0,012343	318	23	0,07233	0,014526
245	251	156	12	0,07692	0,021335	295	21	0,07119	0,014971
252	258	144	10	0,06944	0,021184	274	22	0,08029	0,016417
259	265	134	15	0,11194	0,027237	252	21	0,08333	0,017411
266	272	119	15	0,12605	0,030426	231	16	0,06926	0,016706
273	279	104	3	0,02885	0,016412	215	6	0,02791	0,011233
280	286	101	6	0,05941	0,023521	209	13	0,0622	0,016706
287	293	95	13	0,13684	0,035261	196	23	0,11735	0,022988
294	300	82	8	0,09756	0,032767	173	15	0,08671	0,021395
301	307	74	9	0,12162	0,037995	158	14	0,08861	0,022608
308	314	65	11	0,16923	0,046508	144	16	0,11111	0,026189
315	321	54	3	0,05556	0,031171	128	16	0,125	0,029232
322	328	51	12	0,23529	0,059397	112	21	0,1875	0,036881
329	335	39	5	0,12821	0,053534	91	21	0,23077	0,044167
336	342	34	9	0,26471	0,075661	70	22	0,31429	0,055486
343	349	25	2	0,08	0,054259	48	1	0,02083	0,020615
350	356	23	6	0,26087	0,091561	47	25	0,53191	0,072784
357	363	17	12	0,70588	0,11051	22	13	0,59091	0,104824
364	371	5	5	1	0	9	9	1	0

It can be noted that there exists no male applications to the health institutes between 210 and 216th days of the year. The conditional probability of having a male applicant in the period 63 and 69th days is found to be 0.03219.

Table 5.16: Survival probabilities for acute RHD per week per year (M, F) (2008)

	MALE				FEMALE			
Time	Survival probability	standard error	95% confidence interval lower	upper	Survival probability	standard error	95% confidence interval lower	upper
7	0,977679	0,0056987	0,966509	0,988848	0,987247	0,0030733	0,981223	0,99327
14	0,949405	0,0084547	0,932834	0,965976	0,952738	0,005812	0,941347	0,96413
21	0,922619	0,0103073	0,902417	0,942821	0,931733	0,0069077	0,918194	0,945272
28	0,891369	0,0120039	0,867842	0,914896	0,905476	0,008013	0,889771	0,921182
35	0,852679	0,0136723	0,825881	0,879476	0,873218	0,0091133	0,855357	0,89108
42	0,813988	0,0150105	0,784568	0,843408	0,831958	0,0102411	0,811886	0,85203
49	0,794643	0,0155832	0,7641	0,825185	0,810203	0,0107406	0,789151	0,831254
56	0,764881	0,016359	0,732818	0,796944	0,783946	0,0112722	0,761853	0,806039
63	0,739583	0,0169295	0,706402	0,772764	0,762191	0,0116609	0,739336	0,785045
70	0,715774	0,0173994	0,681672	0,749876	0,729932	0,0121608	0,706098	0,753767
77	0,6875	0,0178804	0,652455	0,722545	0,699175	0,0125613	0,674555	0,723795
84	0,659226	0,0182838	0,623391	0,695062	0,67892	0,012788	0,653856	0,703984
91	0,620536	0,0187191	0,583847	0,657224	0,646662	0,0130924	0,621001	0,672322
98	0,590774	0,0189674	0,553598	0,627949	0,606902	0,0133781	0,580681	0,633122
105	0,558036	0,0191575	0,520488	0,595584	0,569392	0,0135622	0,542811	0,595974
112	0,522321	0,0192687	0,484555	0,560087	0,528882	0,0136719	0,502086	0,555679
119	0,488095	0,0192825	0,450302	0,525888	0,494374	0,0136939	0,467534	0,521213
126	0,453869	0,0192057	0,416227	0,491511	0,450863	0,0136285	0,424151	0,477574
133	0,424107	0,0190644	0,386742	0,461473	0,413353	0,0134876	0,386918	0,439789
140	0,403274	0,0189236	0,366184	0,440363	0,394599	0,013387	0,368361	0,420837
147	0,389881	0,0188143	0,353006	0,426756	0,375844	0,0132659	0,349843	0,401845
154	0,36756	0,018599	0,331106	0,404013	0,362341	0,0131655	0,336537	0,388145
161	0,355655	0,0184667	0,319461	0,391849	0,350338	0,0130669	0,324727	0,375948
168	0,346726	0,0183593	0,310743	0,38271	0,336084	0,0129379	0,310726	0,361442
175	0,337798	0,0182448	0,302038	0,373557	0,32033	0,0127801	0,295282	0,345379
182	0,316964	0,0179491	0,281785	0,352144	0,304576	0,0126054	0,27987	0,329282
189	0,300595	0,0176877	0,265928	0,335262	0,291073	0,0124419	0,266687	0,315458
196	0,28869	0,0174808	0,254429	0,322952	0,27982	0,0122955	0,255721	0,303919
203	0,269345	0,017113	0,235804	0,302886	0,264066	0,0120743	0,240401	0,287731
210	0,254464	0,0168021	0,221533	0,287396	0,251313	0,0118807	0,228027	0,274599
217	0,254464	0,0168021	0,221533	0,287396	0,246812	0,0118092	0,223666	0,269957
224	0,248512	0,0166706	0,215838	0,281186	0,246062	0,0117971	0,22294	0,269183
231	0,244048	0,0165691	0,211573	0,276523	0,243811	0,0117605	0,220761	0,266861
238	0,238095	0,0164301	0,205893	0,270298	0,23856	0,0116735	0,21568	0,261439
245	0,232143	0,0162867	0,200222	0,264064	0,221305	0,0113701	0,19902	0,24359
252	0,214286	0,0158287	0,183262	0,245309	0,205551	0,0110682	0,183858	0,227245
259	0,199405	0,0154131	0,169196	0,229614	0,189047	0,0107243	0,168028	0,210066
266	0,177083	0,0147259	0,148221	0,205946	0,173293	0,010367	0,152974	0,193612
273	0,154762	0,013952	0,127416	0,182107	0,16129	0,0100738	0,141546	0,181035
280	0,150298	0,0137856	0,123278	0,177317	0,156789	0,0099589	0,13727	0,176308
287	0,141369	0,0134399	0,115027	0,167711	0,147037	0,0096998	0,128025	0,166048
294	0,122024	0,0126264	0,097277	0,146771	0,129782	0,0092046	0,111742	0,147823
301	0,110119	0,0120757	0,086451	0,133787	0,11853	0,0088532	0,101178	0,135882
308	0,096726	0,0114024	0,074378	0,119075	0,108027	0,0085021	0,091363	0,124691
315	0,080357	0,0104867	0,059804	0,100911	0,096024	0,0080696	0,080208	0,11184
322	0,075893	0,0102159	0,05587	0,095916	0,084021	0,0075984	0,069128	0,098914
329	0,058036	0,0090195	0,040358	0,075714	0,068267	0,0069077	0,054728	0,081806
336	0,050595	0,0084547	0,034024	0,067166	0,052513	0,0061095	0,040539	0,064488
343	0,037202	0,0073008	0,022893	0,051512	0,036009	0,005103	0,026007	0,046011
350	0,034226	0,0070135	0,02048	0,047972	0,035259	0,0050515	0,025358	0,04516
357	0,025298	0,0060575	0,013425	0,03717	0,016504	0,0034895	0,009665	0,023343
364	0,00744	0,0033151	0,000943	0,013938	0,006752	0,002243	0,002356	0,011148

When the life table method is used, the intervals of 7 days accumulates the survival probabilities within the specified time span. The day of 217 yields a survival probability of 0.254464 indicating that 25.44% of the applications appear first time within 217 days for males.

Table 5.17: Hazard rates for acute RHD per week per year (M, F) (2008)

	MALE				FEMALE			
time	hazard estimates	standard error	density estimates	standard error	hazard estimates	standard error	density estimates	standard error
3,5	0,003225	0,0008326	0,0031888	0,0008141	0,001834	0,0004447	0,0018219	0,000439
10,5	0,004192	0,0009616	0,0040391	0,0009134	0,005082	0,0007492	0,0049298	0,0007142
17,5	0,004088	0,0009635	0,0038265	0,0008898	0,003185	0,0006018	0,0030008	0,0005611
24,5	0,004922	0,0010739	0,0044643	0,0009588	0,004083	0,0006901	0,0037509	0,0006256
31,5	0,006338	0,0012428	0,0055272	0,0010628	0,005182	0,0007901	0,0046083	0,0006913
38,5	0,006633	0,0013004	0,0055272	0,0010628	0,006913	0,0009319	0,0058943	0,0007782
45,5	0,003436	0,0009529	0,0027636	0,000759	0,003785	0,0007028	0,0031079	0,0005708
52,5	0,005453	0,001219	0,0042517	0,0009365	0,004706	0,0007953	0,0037509	0,0006256
59,5	0,004804	0,001165	0,0036139	0,0008654	0,00402	0,0007465	0,0031079	0,0005708
66,5	0,004674	0,0011684	0,0034014	0,0008402	0,006177	0,0009417	0,0046083	0,0006913
73,5	0,005757	0,0013204	0,0040391	0,0009134	0,006149	0,0009601	0,004394	0,0006756
80,5	0,005998	0,0013758	0,0040391	0,0009134	0,004199	0,0008081	0,0028936	0,0005512
87,5	0,008638	0,0016933	0,0055272	0,0010628	0,006953	0,00106	0,0046083	0,0006913
94,5	0,00702	0,0015692	0,0042517	0,0009365	0,009062	0,0012442	0,00568	0,0007645
101,5	0,008142	0,0017352	0,0046769	0,0009807	0,009111	0,0012878	0,0053585	0,0007435
108,5	0,009445	0,0019269	0,005102	0,0010227	0,010539	0,0014332	0,0057872	0,0007714
115,5	0,009678	0,0020169	0,0048895	0,0010019	0,009636	0,0014199	0,0049298	0,0007142
122,5	0,010381	0,0021632	0,0048895	0,0010019	0,013152	0,0017251	0,0062158	0,0007982
129,5	0,009685	0,0021644	0,0042517	0,0009365	0,012401	0,0017521	0,0053585	0,0007435
136,5	0,007194	0,0019221	0,0029762	0,0007871	0,006632	0,0013261	0,0026792	0,0005308
143,5	0,004824	0,0016079	0,0019133	0,0006335	0,006955	0,0013906	0,0026792	0,0005308
150,5	0,00842	0,0021731	0,0031888	0,0008141	0,005226	0,0012317	0,0019291	0,0004516
157,5	0,004703	0,0016626	0,0017007	0,0005977	0,004812	0,0012028	0,0017147	0,0004261
164,5	0,003632	0,0014826	0,0012755	0,0005184	0,005933	0,0013608	0,0020362	0,0004638
171,5	0,003727	0,0015213	0,0012755	0,0005184	0,006857	0,0014959	0,0022506	0,0004872
178,5	0,009091	0,0024284	0,0029762	0,0007871	0,007203	0,0015713	0,0022506	0,0004872
185,5	0,007573	0,0022826	0,0023384	0,0006993	0,006477	0,0015263	0,0019291	0,0004516
192,5	0,005772	0,0020403	0,0017007	0,0005977	0,005632	0,0014538	0,0016075	0,0004127
199,5	0,009905	0,0027454	0,0027636	0,000759	0,008276	0,0018052	0,0022506	0,0004872
206,5	0,008117	0,0025657	0,0021259	0,0006672	0,00707	0,0017142	0,0018219	0,000439
213,5	0	*	0	*	0,002582	0,001054	0,000643	0,0002619
220,5	0,003381	0,0016905	0,0008503	0,0004239	0,000435	0,0004349	0,0001072	0,0001071
227,5	0,00259	0,001495	0,0006378	0,0003674	0,001313	0,0007578	0,0003215	0,0001854
234,5	0,003527	0,0017635	0,0008503	0,0004239	0,00311	0,0011756	0,0007502	0,0002828
241,5	0,003617	0,0018082	0,0008503	0,0004239	0,01072	0,0022337	0,0024649	0,0005095
248,5	0,011429	0,0032965	0,002551	0,0007298	0,010545	0,0022995	0,0022506	0,0004872
255,5	0,010277	0,0032479	0,0021259	0,0006672	0,01195	0,0025455	0,0023577	0,0004985
262,5	0,01694	0,0043661	0,0031888	0,0008141	0,012422	0,0027082	0,0022506	0,0004872
269,5	0,019218	0,0049509	0,0031888	0,0008141	0,01025	0,0025608	0,0017147	0,0004261
276,5	0,004181	0,0024137	0,0006378	0,0003674	0,004043	0,0016504	0,000643	0,0002619
283,5	0,008746	0,003569	0,0012755	0,0005184	0,009171	0,0025423	0,0013932	0,0003845
290,5	0,020985	0,0058044	0,0027636	0,000759	0,017809	0,0037062	0,0024649	0,0005095
297,5	0,014652	0,0051735	0,0017007	0,0005977	0,012948	0,0033397	0,0016075	0,0004127
304,5	0,018499	0,0061536	0,0019133	0,0006335	0,013245	0,0035361	0,0015004	0,0003989
311,5	0,026411	0,007929	0,0023384	0,0006993	0,016807	0,0041944	0,0017147	0,0004261
318,5	0,008163	0,0047111	0,0006378	0,0003674	0,019048	0,0047513	0,0017147	0,0004261
325,5	0,038095	0,010899	0,002551	0,0007298	0,029557	0,0064152	0,0022506	0,0004872
332,5	0,019569	0,0087312	0,0010629	0,0004736	0,037267	0,0080629	0,0022506	0,0004872
339,5	0,043584	0,0143578	0,0019133	0,0006335	0,053269	0,0111578	0,0023577	0,0004985
346,5	0,011905	0,0084106	0,0004252	0,0003002	0,003008	0,0030074	0,0001072	0,0001071
353,5	0,042857	0,0172984	0,0012755	0,0005184	0,10352	0,0192972	0,0026792	0,0005308
360,5	0,155844	0,0377065	0,002551	0,0007298	0,119816	0,0301677	0,0013932	0,0003845
367,5	0,285714	0	0,0010629	0,0004736	0,285714	0	0,0009645	0,0003204

It should be noted that the mid-point of days between 210 and 216, the hazard rate for male is found to be zero.

In the next part, for investigation and calculation of the effects of sub classifications to the survival analysis the data set is divided into different sub categories according to the gender, age and location. And time to first aplication to health institutes data is compared in respect of specified characteristics.

According to the gender, the value of the log-rank test statistics are equal to 0.942 which concludes that there exists no significant difference in survival times between genders for acute RHD. This result presented in the Table 5.18.

Table 5.18: Log-rank test results for acute RHD

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	,005	1	,942

The comparison is also done with respect to the location. Table 5.19 shows means and medians for survival times for the selected three cities. The median time is the lowest for Ankara, which means the time until half of the population applies to the hospitals is the smallest, and equals to 170 days.

Table 5.19: Means and medians for survival time for the selected cities for acute RHD

Means and Medians for Survival Time

placeofliving	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
6	180,319	9,351	161,990	198,647	170,000	7,593	155,118	184,882
34	181,926	8,436	165,391	198,460	178,000	12,880	152,754	203,246
35	216,918	17,245	183,119	250,718	248,000	3,499	241,141	254,859
Overall	186,475	5,963	174,787	198,163	179,000	10,489	158,442	199,558

The log-rank test shows the survival times for the three selected cities are significantly different as shown in Table 5.20.

Table 5.20: Log-rank test results for the selected cities for acute RHD

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	6,427	1	,011

This can be seen also from the survival and hazard curves respectively. The survival patterns for Ankara and İstanbul seem identical and for İzmir there is a definite difference from the other cities. The decrease in survival proportions is the slowest for İzmir. The hazard rates which focuses on the risk, on the contrary to survival curves, there is an increasing risk for the investigated populations.

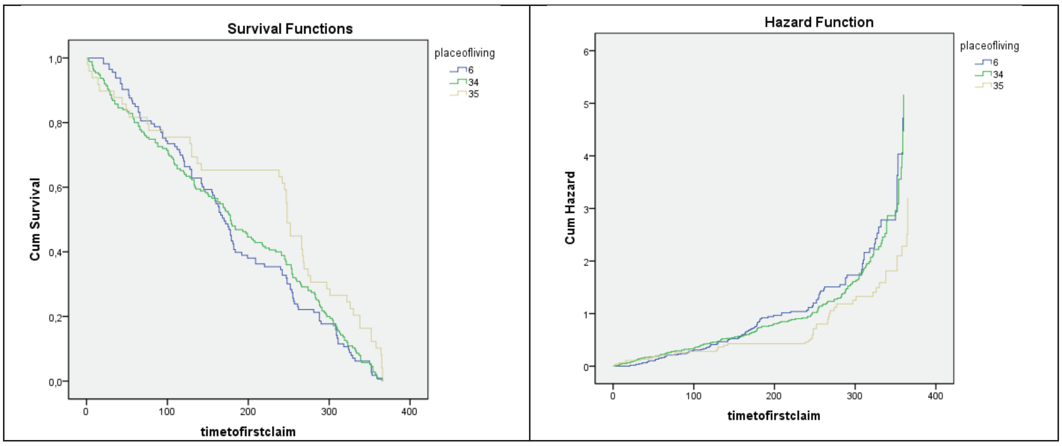


Figure 5.12: Survival and hazard curves of three cities for acute RHD

Lastly, the comparison of survival curves with respect to age is done. For some of these ages survival curves represented to see the difference graphically. In the Figure 5.13 we see the survival curves of the ages 6, 36, 52 and 65. For the ages 6, 36 and 65 the survival probabilities are higher for female, whereas for the age 52, the survival probabilities are higher for male.

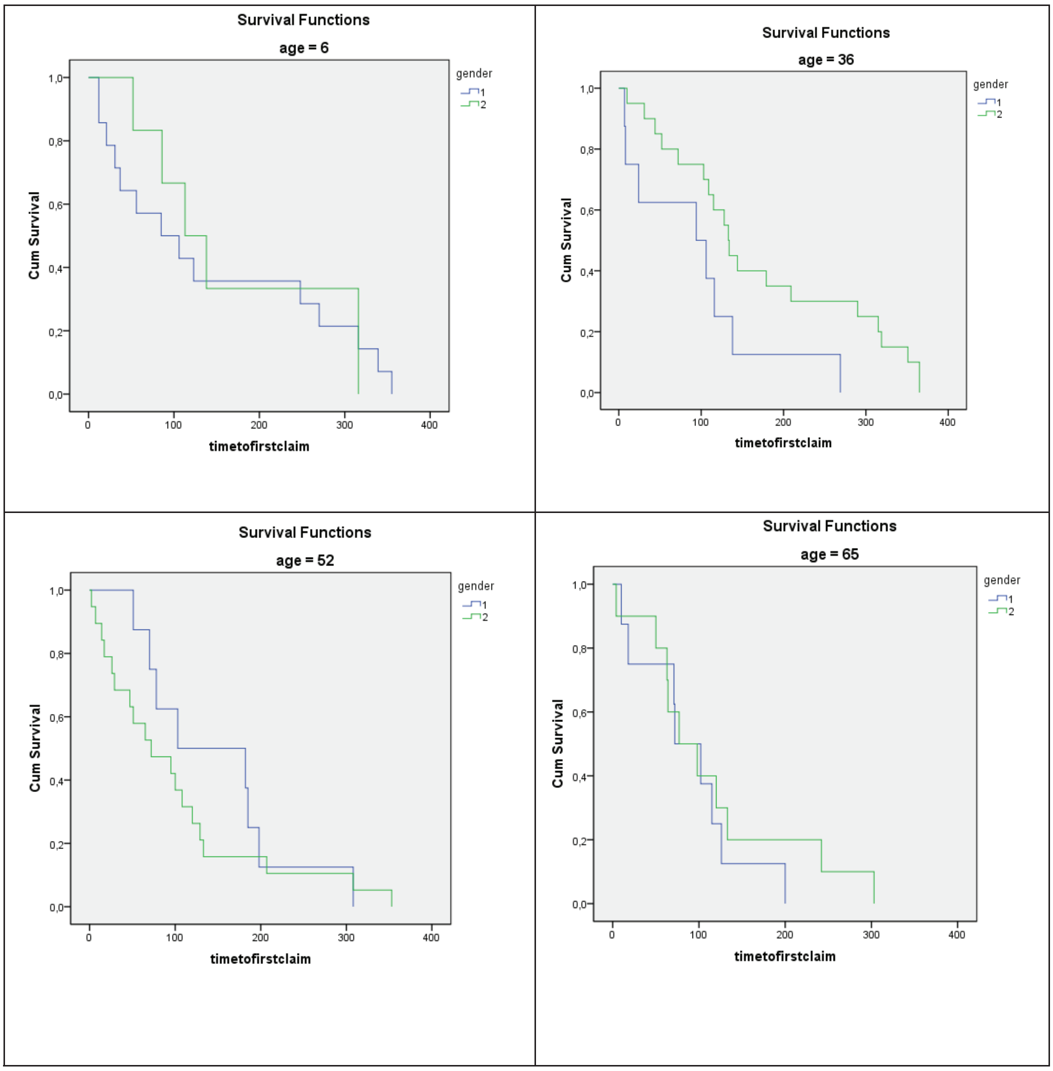


Figure 5.13: Survival curves for ages 6, 36, 52 and 65 for acute RHD

5.3.2.2. Chronic RHD

The survival and hazard curves for chronic RHD are seen in Figure 5.14. For both genders survival curves are decreasing and correspondingly hazard curves are increasing.

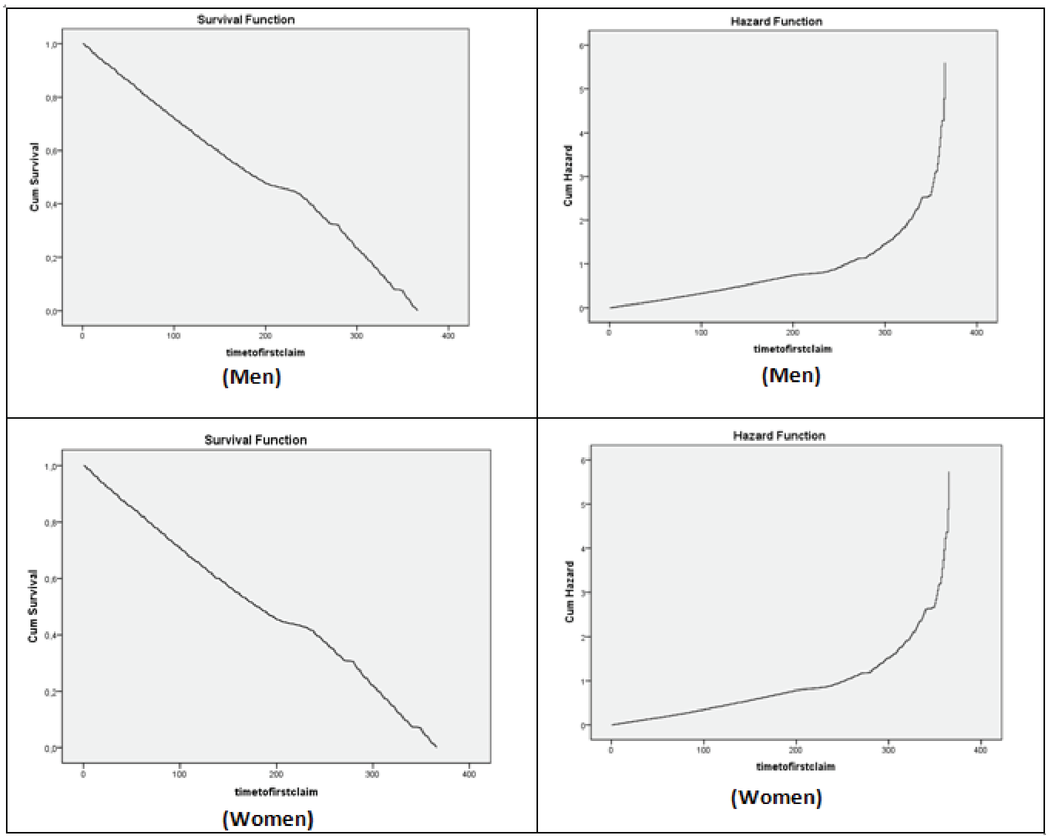


Figure 5.14: Survival probabilities and hazard rates for chronic RHD (M, F)

For each gender and for total life tables are constructed. The tables for male and female are presented in Tables 5.21-5.23, for total in Appendix B10-B12. These tables are:

The first one, represents the number of people at risk and the number of people who failed in each specified period and the conditional probability of failure for each interval (Table 5.21). Second table represents the survival probabilities in each interval (Table 5.22). And the third table represents the hazard and density estimates for each interval (Table 5.23). It should be noted that the intervals for the last life table is the half of the intervals taken for the other two life tables because the hazard estimates is obtained with respect to the half of the interval.

Table 5.21: Life table of chronic RHD for both genders (2008)

Interval		MALE				FEMALE			
Lower	Upper	Number entering	Number failed	conditional probability of failure	standard error	Number entering	Number failed	conditional probability of failure	standard error
0	6	50076	681	0,0136	0,0005176	99482	1489	0,01497	0,000385
7	13	49395	1161	0,0235	0,0006817	97993	2357	0,02405	0,0004894
14	20	48234	993	0,02059	0,0006466	95636	2180	0,02279	0,0004826
21	27	47241	964	0,02041	0,0006505	93456	2119	0,02267	0,0004869
28	34	46277	881	0,01904	0,0006353	91337	2014	0,02205	0,0004859
35	41	45396	1082	0,02383	0,0007159	89323	2185	0,02446	0,0005169
42	48	44314	947	0,02137	0,000687	87138	1912	0,02194	0,0004963
49	55	43367	854	0,01969	0,0006672	85226	1697	0,01991	0,0004785
56	62	42513	1092	0,02569	0,0007673	83529	2108	0,02524	0,0005427
63	69	41421	1045	0,02523	0,0007705	81421	2085	0,02561	0,0005536
70	76	40376	936	0,02318	0,0007489	79336	1960	0,02471	0,0005511
77	83	39440	934	0,02368	0,0007657	77376	2051	0,02651	0,0005775
84	90	38506	973	0,02527	0,0007998	75325	2093	0,02779	0,0005989
91	97	37533	922	0,02457	0,000799	73232	1890	0,02581	0,0005859
98	104	36611	957	0,02614	0,0008339	71342	2041	0,02861	0,0006241
105	111	35654	986	0,02765	0,0008684	69301	2038	0,02941	0,0006418
112	118	34668	793	0,02287	0,0008029	67263	1705	0,02535	0,0006061
119	125	33875	1001	0,02955	0,0009201	65558	1985	0,03028	0,0006692
126	132	32874	949	0,02887	0,0009235	63573	1925	0,03028	0,0006796
133	139	31925	927	0,02904	0,0009397	61648	1919	0,03113	0,0006994
140	146	30998	730	0,02355	0,0008613	59729	1435	0,02403	0,0006266
147	153	30268	913	0,03016	0,0009831	58294	1988	0,0341	0,0007517
154	160	29355	949	0,03233	0,0010323	56306	1722	0,03058	0,0007256
161	167	28406	820	0,02887	0,0009934	54584	1632	0,0299	0,000729
168	174	27586	800	0,029	0,0010103	52952	1635	0,03088	0,0007517
175	181	26786	793	0,02961	0,0010356	51317	1702	0,03317	0,0007905
182	188	25993	763	0,02935	0,001047	49615	1541	0,03106	0,0007788
189	195	25230	750	0,02973	0,0010692	48074	1510	0,03141	0,0007955
196	202	24480	699	0,02855	0,0010645	46564	1461	0,03138	0,0008079
203	209	23781	395	0,01661	0,0008288	45103	872	0,01933	0,0006484
210	216	23386	291	0,01244	0,0007249	44231	531	0,01201	0,0005178
217	223	23095	272	0,01178	0,0007099	43700	549	0,01256	0,0005328
224	230	22823	330	0,01446	0,0007902	43151	691	0,01601	0,0006043
231	237	22493	550	0,02445	0,0010298	42460	1189	0,028	0,0008007
238	244	21943	982	0,04475	0,0013958	41271	1982	0,04802	0,0010525
245	251	20961	1191	0,05682	0,001599	39289	2161	0,055	0,0011502
252	258	19770	1244	0,06292	0,001727	37128	2236	0,06022	0,0012347
259	265	18526	1196	0,06456	0,0018055	34892	2249	0,06446	0,0013146
266	272	17330	1117	0,06445	0,0018654	32643	1996	0,06115	0,0013261
273	279	16213	176	0,01086	0,0008138	30647	277	0,00904	0,0005406
280	286	16037	1487	0,09272	0,0022904	30370	2982	0,09819	0,0017075
287	293	14550	1416	0,09732	0,0024572	27388	2814	0,10275	0,0018347
294	300	13134	1386	0,10553	0,0026808	24574	2731	0,11113	0,0020049
301	307	11748	1110	0,09448	0,0026986	21843	2108	0,09651	0,001998
308	314	10638	1348	0,12672	0,0032252	19735	2634	0,13347	0,0024208
315	321	9290	1251	0,13466	0,0035417	17101	2491	0,14566	0,0026976
322	328	8039	1374	0,17092	0,0041985	14610	2543	0,17406	0,0031369
329	335	6665	1416	0,21245	0,0050104	12067	2636	0,21845	0,0037614
336	342	5249	1250	0,23814	0,0058792	9431	2249	0,23847	0,0043881
343	349	3999	156	0,03901	0,0030618	7182	265	0,0369	0,0022244
350	356	3843	1644	0,42779	0,007981	6917	2862	0,41376	0,0059218
357	363	2199	1505	0,6844	0,0099108	4055	2780	0,68557	0,0072911
364	371	694	694	1	0	1275	1275	1	0

The conditional probability of having a male applicant in the period of 63-69 days is found to be 0.02523. It should be noted that the number of failure for males in chronic RHD is much higher than acute case for the same time interval.

Table 5.22: Survival probabilities for chronic RHD per week per year (M, F) (2008)

	MALE				FEMALE			
Time	Survival probability	standard error	95% confidence interval lower	upper	Survival probability	standard error	95% confidence interval lower	upper
7	0,986401	0,0005176	0,985386	0,987415	0,985032	0,000385	0,984278	0,985787
14	0,963216	0,0008412	0,961567	0,964865	0,96134	0,0006112	0,960142	0,962538
21	0,943386	0,0010327	0,941362	0,94541	0,939426	0,0007563	0,937944	0,940909
28	0,924135	0,0011832	0,921816	0,926454	0,918126	0,0008693	0,916422	0,91983
35	0,906542	0,0013007	0,903993	0,909091	0,897881	0,00096	0,895999	0,899763
42	0,884935	0,001426	0,88214	0,88773	0,875917	0,0010452	0,873869	0,877966
49	0,866024	0,0015222	0,86304	0,869007	0,856698	0,0011109	0,85452	0,858875
56	0,84897	0,0016002	0,845833	0,852106	0,839639	0,0011634	0,837359	0,84192
63	0,827163	0,0016897	0,823851	0,830474	0,81845	0,0012221	0,816054	0,820845
70	0,806294	0,0017661	0,802833	0,809756	0,797491	0,0012741	0,794994	0,799988
77	0,787603	0,0018277	0,784021	0,791185	0,777789	0,0013181	0,775206	0,780372
84	0,768951	0,0018836	0,765259	0,772643	0,757172	0,0013595	0,754508	0,759837
91	0,749521	0,0019363	0,745726	0,753316	0,736133	0,0013973	0,733394	0,738872
98	0,731109	0,0019814	0,727225	0,734992	0,717135	0,001428	0,714336	0,719934
105	0,711998	0,0020236	0,708032	0,715964	0,696618	0,0014575	0,693762	0,699475
112	0,692308	0,0020625	0,688265	0,69635	0,676132	0,0014836	0,673224	0,67904
119	0,676472	0,0020906	0,672374	0,680569	0,658994	0,001503	0,656048	0,661939
126	0,656482	0,0021221	0,652323	0,660641	0,63904	0,0015227	0,636056	0,642025
133	0,637531	0,0021482	0,633321	0,641741	0,61969	0,0015392	0,616673	0,622707
140	0,619019	0,0021701	0,614766	0,623272	0,6004	0,001553	0,597356	0,603444
147	0,604441	0,0021851	0,600159	0,608724	0,585975	0,0015616	0,582915	0,589036
154	0,586209	0,0022009	0,581895	0,590523	0,565992	0,0015714	0,562912	0,569072
161	0,567258	0,0022141	0,562918	0,571597	0,548682	0,0015777	0,54559	0,551774
168	0,550883	0,0022228	0,546526	0,555239	0,532277	0,0015819	0,529177	0,535378
175	0,534907	0,0022289	0,530538	0,539276	0,515842	0,0015845	0,512737	0,518948
182	0,519071	0,0022327	0,514695	0,523447	0,498733	0,0015852	0,495626	0,50184
189	0,503834	0,0022343	0,499455	0,508213	0,483243	0,0015844	0,480138	0,486348
196	0,488857	0,0022338	0,484479	0,493235	0,468065	0,001582	0,464964	0,471165
203	0,474898	0,0022316	0,470524	0,479272	0,453379	0,0015783	0,450285	0,456472
210	0,46701	0,0022295	0,46264	0,47138	0,444613	0,0015755	0,441525	0,447701
217	0,461199	0,0022276	0,456833	0,465565	0,439275	0,0015735	0,436191	0,442359
224	0,455767	0,0022256	0,451405	0,460129	0,433757	0,0015713	0,430677	0,436837
231	0,449177	0,0022228	0,444821	0,453534	0,426811	0,0015682	0,423737	0,429884
238	0,438194	0,0022172	0,433848	0,44254	0,414859	0,0015621	0,411797	0,417921
245	0,418584	0,0022045	0,414263	0,422905	0,394936	0,0015499	0,391898	0,397973
252	0,3948	0,0021844	0,390519	0,399081	0,373213	0,0015334	0,370208	0,376219
259	0,369958	0,0021575	0,365729	0,374186	0,350737	0,001513	0,347771	0,353702
266	0,346074	0,0021259	0,341907	0,350241	0,32813	0,0014887	0,325212	0,331047
273	0,323768	0,002091	0,31967	0,327866	0,308066	0,0014638	0,305197	0,310935
280	0,320253	0,002085	0,316167	0,32434	0,305281	0,0014601	0,30242	0,308143
287	0,290558	0,0020289	0,286582	0,294535	0,275306	0,0014162	0,27253	0,278082
294	0,262281	0,0019657	0,258429	0,266134	0,24702	0,0013674	0,24434	0,2497
301	0,234603	0,0018936	0,230892	0,238315	0,219567	0,0013124	0,216995	0,22214
308	0,212437	0,0018279	0,208855	0,21602	0,198378	0,0012643	0,1959	0,200856
315	0,185518	0,0017371	0,182113	0,188923	0,1719	0,0011962	0,169556	0,174245
322	0,160536	0,0016405	0,157321	0,163751	0,146861	0,0011223	0,144661	0,14906
329	0,133098	0,0015179	0,130123	0,136073	0,121298	0,0010351	0,11927	0,123327
336	0,104821	0,0013689	0,102138	0,107504	0,094801	0,0009288	0,092981	0,096621
343	0,079859	0,0012114	0,077484	0,082233	0,072194	0,0008206	0,070586	0,073802
350	0,076743	0,0011895	0,074412	0,079075	0,06953	0,0008064	0,06795	0,071111
357	0,043913	0,0009157	0,042119	0,045708	0,040761	0,0006269	0,039532	0,04199
364	0,013859	0,0005224	0,012835	0,014883	0,012816	0,0003566	0,012117	0,013515

The survival probability for males in chronic RHD is calculated to be 0.827163 for the end of 63 days in the year 2008.

Table 5.23: Hazard rates for chronic RHD per week per year (M, F) (2008)

	MALE				FEMALE			
time	hazard estimates	standard error	density estimates	standard error	hazard estimates	standard error	density estimates	standard error
3,5	0,001956	0,000075	0,0019428	0,0000739	0,002154	0,0000558	0,0021382	0,000055
10,5	0,003398	0,0000997	0,0033121	0,0000961	0,003478	0,0000716	0,0033847	0,0000689
17,5	0,002972	0,0000943	0,0028328	0,000089	0,003294	0,0000705	0,0031305	0,0000663
24,5	0,002945	0,0000949	0,0027501	0,0000877	0,003276	0,0000712	0,0030429	0,0000654
31,5	0,002746	0,0000925	0,0025133	0,0000839	0,003185	0,000071	0,0028921	0,0000638
38,5	0,003446	0,0001048	0,0030867	0,0000928	0,003538	0,0000757	0,0031377	0,0000664
45,5	0,003086	0,0001003	0,0027016	0,000087	0,003169	0,0000725	0,0027457	0,0000622
52,5	0,002841	0,0000972	0,0024363	0,0000827	0,002873	0,0000697	0,0024369	0,0000586
59,5	0,003717	0,0001125	0,0031153	0,0000932	0,003651	0,0000795	0,0030271	0,0000652
66,5	0,00365	0,0001129	0,0029812	0,0000913	0,003706	0,0000811	0,0029941	0,0000649
73,5	0,003351	0,0001095	0,0026702	0,0000865	0,003573	0,0000807	0,0028146	0,0000629
80,5	0,003424	0,000112	0,0026645	0,0000864	0,003838	0,0000847	0,0029453	0,0000644
87,5	0,003656	0,0001172	0,0027758	0,0000881	0,004025	0,000088	0,0030056	0,000065
94,5	0,003553	0,000117	0,0026303	0,0000858	0,003735	0,0000859	0,0027141	0,0000618
101,5	0,003784	0,0001223	0,0027301	0,0000874	0,004146	0,0000918	0,0029309	0,0000642
108,5	0,004006	0,0001276	0,0028129	0,0000887	0,004264	0,0000944	0,0029266	0,0000642
115,5	0,003306	0,0001174	0,0022623	0,0000797	0,003668	0,0000888	0,0024484	0,0000588
122,5	0,004285	0,0001354	0,0028557	0,0000894	0,004392	0,0000986	0,0028505	0,0000633
129,5	0,004184	0,0001358	0,0027073	0,000087	0,004392	0,0001001	0,0027643	0,0000624
136,5	0,004209	0,0001382	0,0026446	0,0000861	0,004517	0,0001031	0,0027557	0,0000623
143,5	0,003404	0,000126	0,0020825	0,0000765	0,003474	0,0000917	0,0020607	0,000054
150,5	0,004375	0,0001448	0,0026046	0,0000854	0,004956	0,0001111	0,0028548	0,0000634
157,5	0,004694	0,0001524	0,0027073	0,000087	0,004437	0,0001069	0,0024728	0,0000591
164,5	0,004184	0,0001461	0,0023393	0,000081	0,004336	0,0001073	0,0023436	0,0000575
171,5	0,004204	0,0001486	0,0022822	0,00008	0,00448	0,0001108	0,0023479	0,0000576
178,5	0,004293	0,0001524	0,0022623	0,0000797	0,004818	0,0001168	0,0024441	0,0000587
185,5	0,004256	0,0001541	0,0021767	0,0000782	0,004507	0,0001148	0,0022129	0,0000559
192,5	0,004311	0,0001574	0,0021396	0,0000775	0,004559	0,0001173	0,0021684	0,0000554
199,5	0,004138	0,0001565	0,0019941	0,0000749	0,004554	0,0001191	0,002098	0,0000545
206,5	0,002393	0,0001204	0,0011269	0,0000565	0,002789	0,0000944	0,0012522	0,0000422
213,5	0,001789	0,0001049	0,0008302	0,0000485	0,001725	0,0000749	0,0007625	0,000033
220,5	0,001692	0,0001026	0,000776	0,0000469	0,001806	0,0000771	0,0007884	0,0000336
227,5	0,002081	0,0001145	0,0009414	0,0000517	0,002306	0,0000877	0,0009923	0,0000376
234,5	0,003536	0,0001508	0,001569	0,0000665	0,004057	0,0001177	0,0017074	0,0000492
241,5	0,00654	0,0002086	0,0028015	0,0000885	0,007029	0,0001578	0,0028462	0,0000633
248,5	0,008354	0,000242	0,0033977	0,0000973	0,00808	0,0001737	0,0031032	0,000066
255,5	0,009281	0,000263	0,0035489	0,0000994	0,008871	0,0001875	0,0032109	0,0000671
262,5	0,00953	0,0002754	0,003412	0,0000975	0,009515	0,0002005	0,0032296	0,0000673
269,5	0,009514	0,0002845	0,0031866	0,0000943	0,009011	0,0002016	0,0028663	0,0000635
276,5	0,001559	0,0001175	0,0005021	0,0000378	0,001297	0,0000779	0,0003978	0,0000239
283,5	0,01389	0,0003598	0,0042421	0,0001084	0,014751	0,0002698	0,0042822	0,0000772
290,5	0,014614	0,0003879	0,0040396	0,0001058	0,015473	0,0002913	0,0040409	0,0000751
297,5	0,015915	0,0004268	0,003954	0,0001047	0,01681	0,0003211	0,0039217	0,000074
304,5	0,014167	0,0004247	0,0031666	0,000094	0,014486	0,0003151	0,0030271	0,0000652
311,5	0,019327	0,0005252	0,0038456	0,0001033	0,02043	0,0003971	0,0037825	0,0000727
318,5	0,020626	0,0005816	0,0035689	0,0000996	0,022444	0,0004483	0,0035771	0,0000708
325,5	0,026698	0,0007171	0,0039198	0,0001043	0,027236	0,0005376	0,0036518	0,0000715
332,5	0,033958	0,000896	0,0040396	0,0001058	0,035033	0,0006772	0,0037853	0,0000727
339,5	0,038618	0,0010823	0,003566	0,0000996	0,038679	0,0008081	0,0032296	0,0000673
346,5	0,005684	0,000455	0,000445	0,0000356	0,00537	0,0003298	0,0003805	0,0000233
353,5	0,077742	0,001845	0,00469	0,0001138	0,074527	0,0013449	0,0041099	0,0000757
360,5	0,148635	0,0032721	0,0042935	0,000109	0,149022	0,0024115	0,0039921	0,0000746
367,5	0,285714	0	0,0019798	0,0000746	0,285714	0	0,0018309	0,0000509

The comparison with respect to gender, age and location is done also for RHD chronic case. According to the gender, the value of the log-rank test statistics are equal to 0, which concludes that there exists significant difference in survival times between genders for chronic RHD. This result presented in the Table 5.24.

Table 5.24: Log-rank test results for chronic RHD

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	64,104	1	,000

The comparison is done with respect to the three biggest cities in Turkey shows means and medians for survival times for the three cities. In table 5.25 the median time is the lowest for Ankara, which means the time until half of the population applies to the hospitals is the smallest, and equals to 170 days.

Table 5.25: Means and medians for survival time for the selected cities for chronic RHD

Means and Medians for Survival Time

placeofliving	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
6	186,623	,909	184,842	188,405	172,000	1,661	168,744	175,256
34	181,221	,696	179,857	182,584	170,000	1,172	167,702	172,298
35	214,628	,952	212,763	216,493	252,000	1,014	250,013	253,987
Overall	191,360	,482	190,417	192,304	186,000	1,189	183,670	188,330

The log-rank test shows the survival times for the three selected cities are significantly different as shown in Table 5.26.

Table 5.26: Log-rank test results for the selected cities for chronic RHD

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	258,674	1	,000

This can be seen also from the survival and hazard curves respectively. The survival patterns for Ankara and İstanbul seem identical and for İzmir there is a definite difference from the other cities. The decrease in survival proportions is the slowest for İzmir. The hazard rates which focuses on the risk, on the contrary to survival curves, there is an increasing risk for the investigated populations.

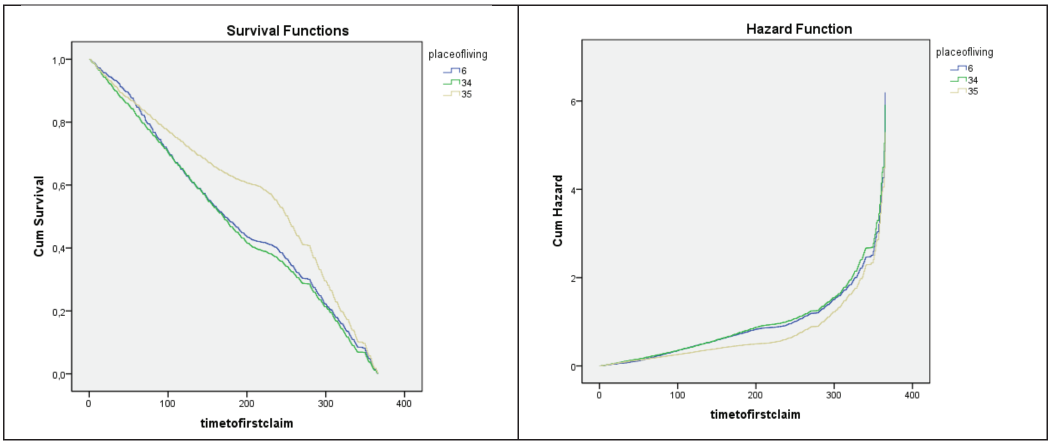


Figure 5.15: Survival and hazard curves of three cities for chronic RHD

Lastly, the comparison of survival curves with respect to age is done. For some of these ages survival curves represented to see the difference graphically. In the Figure 5.16 we see the survival curves of the ages 6, 36, 52 and 65. For the 6 years old people the survival probabilities are higher for female, whereas for the ages 36, 52 and 65 the survival probabilities are higher for male.

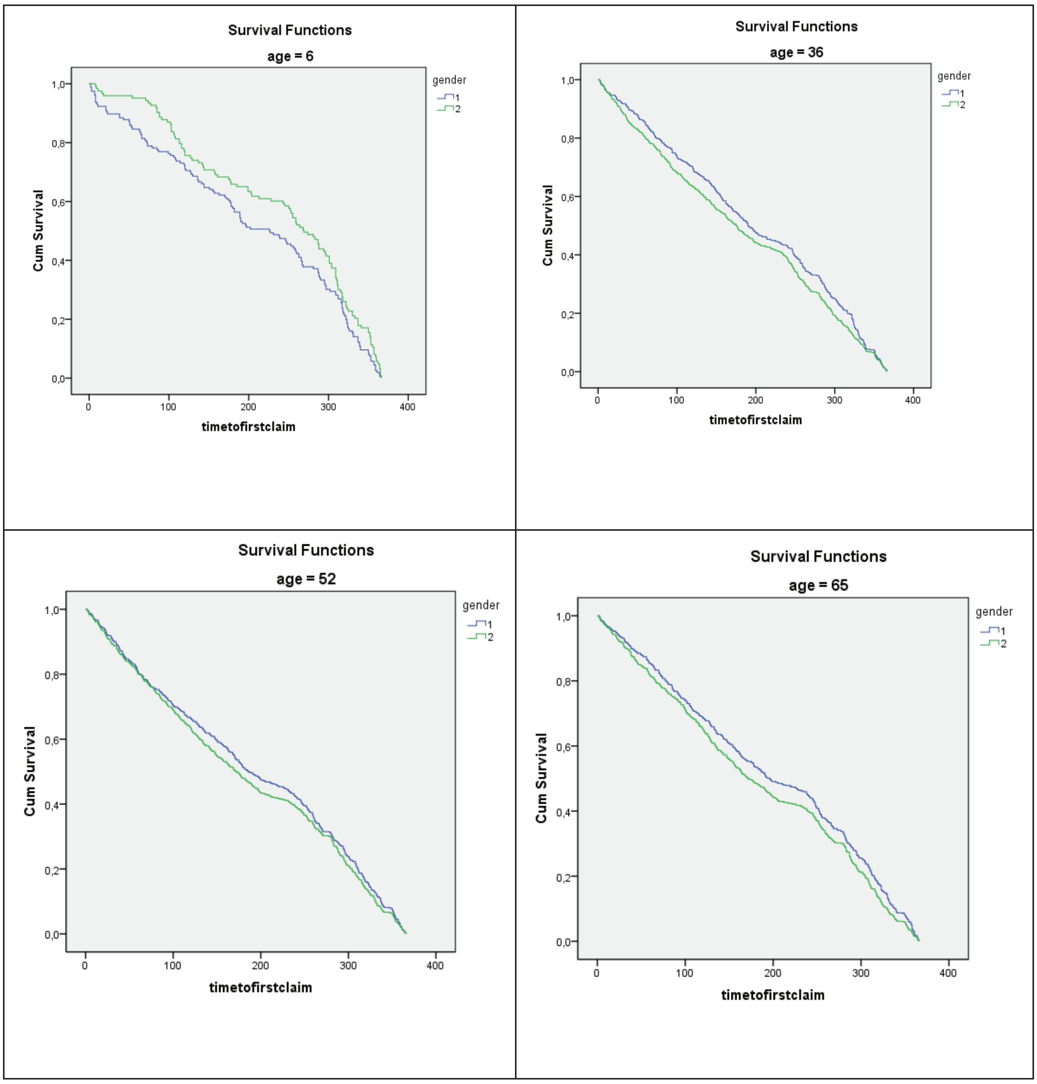


Figure 5.16: Survival curves for the ages 6, 36, 52 and 65 for chronic RHD

5.3.2.3. Comparisons of acute and chronic RHD

In this section, the comparisons are made with respect to acute and chronic RHD morbidities to see the effect of disease type on survival times. The median application time to health institutes is 116 days for acute RHD and 183 days for chronic RHD. This shows acute cases apply health institutes much earlier than chronic cases. This difference could be caused from the seasonal effect of acute RHD. These results are presented in the Table 5.27.

Table. 5.27: Means and medians for survival time for acute and chronic RHD

Means and Medians for Survival Time

ICD-10	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
acute	142,945	2,275	138,486	147,404	116,000	2,228	111,633	120,367
chroni	188,018	,290	187,448	188,587	183,000	,640	181,746	184,254
Overall	187,421	,288	186,856	187,987	182,000	,590	180,844	183,156

The overall comparisons for acute and chronic RHD show that there is a significant diference between the application times for acute and chronic RHD. This can be seen in the log-rank test result in the Table 5.28. This result can also be seen from the survival and hazard curves in the Figure 5.17.

Table.5.28: Log-rank test results for acute and chronic RHD

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	307,484	1	,000

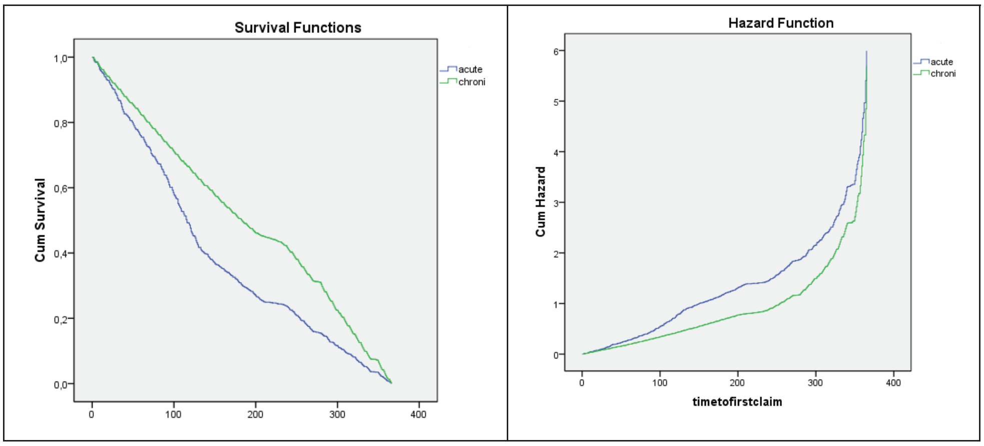


Figure 5.17: Survival and hazard curves for acute and chronic RHD

5.4. Actuarial Premium Calculation for the Time to First Claims

To illustrate the use and impact of estimates obtained by using survival analysis, a simulated data set is used only for poisoning. The reason for simulation is, as no information on distribution of claim frequency and severity have been found due to the scarcity of data. The survival estimates obtained by sub clasifications based on gender, location and age could not taken into account because of lack of information on claim severity and frequency data with respect to these classifications.

As explained in chapter 3, the case application assumes the implementation of individual risk model (IRM) and collective risk model (CRM) to find the impact of survival estimates on actuarial premium calculation. Figure 5.18 represents an assumption on a policy year partitioned into weeks and having n_i, i=1,..52 claims per week. In poisoning case the data shows that there exists totally 19,916 claims in 2008.

```
31/2007   n1          n2                                n52       31/2008
|--------------------|----------------------|----------------------|--------------------|
 0        X1     t1        X2        t2                     X52       t52
```

Figure 5.18: Representation of a policy year in IRM

In other words, let N_i be the total number of time to first claims in the time period having loss amount of X_i, i=1,2,..,19916. And for each interval, there are n_t claims having loss amount of X_t, where t=1,2,..52 and let $X_t = \sum_{i=1}^{n_t} X_i$ and $S = \sum_{i=1}^{52} X_t$

Moreover, the amount of loss is assumed to be Gamma distribution having expected value and variance formulas for S are as follows (Bowers, 1997):

$$E[S] = \sum_{i=1}^{19916} E[X_i] = \sum_{t=1}^{52} n_t x_t q_t \tag{5.1}$$

$$Var[S] = \sum_{i=1}^{19916} Var[X_i] = \sum_{t=1}^{52} n_t x_t^2 q_t (1-q_t) \tag{5.2}$$

where q_t is the probability of having the first claim in the time interval t and $1-q_t$ denoting p_t is the probability of not having the first claim in the time interval t.

The second model that we use to estimate total claims S is the collective risk model. According to the assumptions and the independency between periods in collective risk model, we generate the claim amounts and claim numbers.

31/2007 N_1 N_2 N_{52} 31/2008

|--------------------|---------------------|---------------------|--------------------|

$N_1 \sim Poi(\lambda_1)$

|---|

$N \sim Poi(\lambda)$

Figure 5.19. Representation of a policy year in the CRM

The number of claims occurred in each time interval is a random variable in this approach as presented in the Figure 5.19. Therefore, we generate N, the number of frequency having Poisson distribution which is one of the most common distribution used in literature for modelling claim frequency.

For the claim amount distribution, we assume that severity follows Gamma distribution with the same parameters over one year period. For the estimation of the parameters of the Gamma distribution for poisoning, Akar et al, 2007 gives the estimates for mean and the standard deviation for intentional and unintentional cases. These are given as $\mu_1 = 405.1$, $\sigma_1 = 240.8$ and $\mu_2 = 245.7$, $\sigma_2 = 169$ respectively. As in the analyses there is no such division taken, these parameters are found by averaging the means and standard deviations of intentional and unintentional poisoning cases.

$\hat{\mu}_{avg} = \dfrac{\hat{\mu}_1 + \hat{\mu}_2}{2}$, $\hat{\mu}_{avg} = 325.4$ and $\hat{\sigma}_{avg} = \sqrt{\dfrac{\sigma_1^2 + \sigma_2^2}{2}}$, $\hat{\sigma}_{avg} = 208.02$. This yields the estimates:

$\mu = \alpha\beta$, $\sigma^2 = \alpha\beta^2$, $\hat{\beta} = \dfrac{s^2}{\bar{x}}$, $\hat{\alpha} = \dfrac{\bar{x}^2}{s^2}$ and $\hat{\beta} = 132.98$, $\hat{\alpha} = 2.44$, as Gamma parameters.

Similarly, the frequency distribution N is assumed to be Poisson with parameter λ which is estimated by using the number of first claims within the year.

Therefore, based on the common distributions used in the literature, time to first claims are generated by using Minitab software to immitate the scenario. The steps of the analysis are as follows:

Step I: Estimate the Poisson parameters $\hat{\lambda}_t$ (t=1,2,..52) for each week. Use

$\hat{\lambda}_t = \hat{\lambda} \bullet q_t$

Step II: Generate $N_t \sim Poi(\lambda_t)$ for each time interval t. (by using the $\hat{\lambda}_t$ values found above based on the data)

Step III: Generate $X_t \sim Gamma(\alpha_t, \beta_t)$, where t refers to the time interval.

Step IV: Calculate $E[Loss_t] = \hat{\lambda}_t \bullet E[X_t]$ for each interval t=1,2,..,52

Step V: Calculate $E[Loss_{overall}] = \sum_{t=1}^{52} E[Loss_t]$

Based on the algorithm above, the expected claim amount and its variance are:

$$E[S] = \sum_{t=1}^{52} S_t = E[\sum_{t=1}^{52} x_t N_t] = \sum_{t=1}^{52} x_t \lambda_t \quad (5.3)$$

$$Var[S] = Var(\sum_{t=1}^{52} x_t N_t) = \sum_{t=1}^{52} x_t^2 \lambda_t \quad (5.4)$$

This algorithm is simulated and the net premium is obtained according to the results of simulation work summarized in Table 5. 29.

Table 5.29: Simulation work results for poisoning

	E[S]	**St Dev[S]**
IRM	131,059	6458.96
CRM	123,966	6336.54

The calculated expected losses with respect to IRM and CRM are obtained based on only the first claims from each individual in the population. These expected losses constitute a part of total expected loss for a year, as claims more than one for each individual is not taken into account in the calculation of survival analyses. It can be concluded that CRM results in lower expected loss and standard deviation compared to IRM. This result is expected as CRM includes the randomness in the frequency dimension. It should be noted that, this approach does not take into account the age of the individual which is an important factor in premium calculation. This enables the researcher to estimate the size of expected loss for a portfolio of insurance policy without getting into the specifications of the policy owner expect the gender.

CHAPTER 6

CONCLUSION AND COMMENTS

In this thesis, non-parametric survival models are applied to the time to first claims data for RHD and poisoning morbidities for the purposes of obtaining survival probabilities, comparing the survival rate of those morbidities in terms of gender, age and location; comparing the survival rates with respect to acute and chronic specification; incorporating the time to first claims data to the actuarial premium calculation and comparison of the individual and collective risk models based on the survival probabilities obtained from Kaplan-Meier and Life Table methods.

The survival and hazard probabilities for each specified time interval are determined by using non-parametric methods. To illustrate the use and impact of estimates obtained by survival analysis on actuarial calculation a simulation work is performed only for the poisoning. Based on the common distributions used in literature, time to first claim distributions are generated by using Minitab software and a simulation work is performed. Expected loss with respect to individual and collective risk models are calculated for poisoning.

The data obtained from Social Security Instıtute of Turkey represents only the population covered by this institute. The structure of the data does not allow to make a proper analysis on calculating the incidence rates which would be an important factor to calculate the premium values. In order to eliminate age and other risk factors on calculating actuarial loss, time to first claims are considered.

The aim of this thesis is to investigate and determine the effect of time to first claim on the actuarial calculation of certain morbidities. For this purpose, two different types of morbidities are studied to model the time to first claim within a policy year in order to find the time to first claim for determined periods in one year and its impact on actuarial premium calculation. Based on the results of the proposed study the following conclusions are made:

(i) Kaplan-Meier and life table analysis yield survival rates calculated for both of the morbidities for each gender based on the time to first claim reflects reasonable values.

(ii) The comparisons done with respect to gender in both morbidities yield reasonable results. No gender impact for poisoning is found. However, RHD results show significant difference between genders.

(iii) The comparisons done with respect to location (three metropolitans in Turkey) show considerable differences for both morbidities. For both of the morbidities, İzmir has the highest survival probabilities with respect to Ankara and İstanbul.

(iv) The comparison based on age for both morbidities agree that there exists significant variations for most of the ages.

(v) The comparison based on acute and chronic RHD shows that there is significant difference and chronic RHD has higher survival probabilities than acute RHD as time passes.

(vi) The information collected from the survival analysis based on time to first claims enable researcher to determine expected loss values for any selected period within the policy year.

(vii) This approach enables researcher to implement risk models and make the comparisons based on the distributions assumed for claim and frequency. This also eliminates the age factor in the calculation of actuarial premium value for a portfolio of health insurance under Social Security Institute in Turkey.

As further study, the same analyses for a portfolio taken from an insurance company can be done for comparison and efficiency reasons. The risk factors such as alcohol consumption, smoking status, nutrition, family history, and past information about having the morbidity should be associated in the survival rates which allows the researchers to use Cox- Proportional Regression Analysis.

It should also be mentioned that the data set used in this study reflects the picture for Social Security Institution which is the biggest health insurance supplier in Turkey. A generalization of this study to Turkish population need to be studied.

REFERENCES

[1] Akar et al, 2007. Drug intoxications and hospital costs

[2] Allison, P. D. 1995. Survival Analysis Using SAS: A Practical Guide

[3] http://www.americanheart.gov

[4] Anokute, C. C. 1991. Principals of Epidemiology For the Applied Medical Sciences.

[5] Bovbjerg, R. R. and Hadley, J. 2007. Why health insurance is Important?

[6] Bowers et al, 1997. Actuarial Mathematics, Second Edition. The Society of Actuaries.

[7] Breslow, N. E. and Day, N. E. 1980. Statistical Methods in Cancer Research. The Analysis of Case-Control Studies.

[8] Bronstein et al, 2006. Annual Report of the American Association of Poison Control Centers. National Poison Morbidity System (NPDS). Clinical Toxicology. 45:815-917.

[9] Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. 2008 Web-based Injury Statistics Query and Reporting System (WISQARS) [online].

[10] Czado, C. and Rudolph, F. 2002. Application of survival analysis methods to long-term care insurance. Insurance: Mathematics and Economics 31. 395-413

[11] http://www.cdc.gov . Page visited at various times in 2010

[12] Deshpande, V. J. 2005. Lifetime Morbidity: Statistical Models and Methods, Series on Quality, Reliability and Engineering Statistics.

[13] Dupont, D. W. 2009. Statistical Modeling for Biomedical Researchers. Cambridge University Press.

[14] Finkelstein E, Corso P, Miller T. 2006. The incidence and economic costs of injury in the United States. New York: Oxford University Press.

[15] Fox, J. 2006.Introduction to Survival Analysis. Lecture notes

[16] Greenwood, B.1926."A Report on the Natural Duration of Cancer", Reports on Public Health and Medical Subjects, H.M. Stationery Office, 33, 1.

[17] Gordis, L. 2000. Epidemiology. Second Edition.

[18] Hossack et al, 1983. Introductory Statistics with Applications in General Insurance. Cambridge University Press

[19] Kaas et al, 2001. Modern Actuarial Risk Theory. Kluwer Academic Publishers.

[20] Kalbfleisch, J. D., Prentice, R.L., 2002. The Statistical Analysis of Failure Time Morbidity. New York: John Wiley and Sons.

[21] Kaplan, E. J. and Meier, P. 1958. Nonparametric Estimation From Incomplete Observations. Journal American Statistics Association, (53), 457-481

[22] Kısa, A. and Younis, M. Z. 2006. Financing Health Care for the Poor in Turkey: Is a Temporary Solution Becoming a Permanent Scheme? Public Health Report Nov-Dec; 121(6): 764- 768

[23] Kleinbaum, D. G. 1996. Survival Analysis a Self Learning Text. Springer, 197, New York.

[24] Lawless, J. F. 2003. Statistical Models and Methods for Lifetime Morbidity, John Wiley & Sons.

[25] Lee, E. T., 1992. Statistical Methods for Survival Morbidity Analysis. Second Edition, John Wiley & Sons. Inc., New York.

[26] London, D. 1998. Survival Models and Their Estimation. ACTEX Publications

[27] Lu et al, 2008. Premium Rates Based on Genetic Studies: How Reliable Are They? Science Direct, Insurance: Mathematics and Economics 42, pp.319-331.

[28] MacDonald et al, A Model for Coronary Heart Disease and Stroke with Applications to Critical Illness Insurance Underwriting I: The Model.

[29] Marshall, W. A. and Olkin, I. 2007. Life Distributions: Structure of Nonparametric, Semiparametric, and Parametric Families. Springer Series in Statistics.

[30] METU, IAM. 2010. Constructing Morbidity Tables: Phase I, Report III.

[31] METU, IAM. 2010. Constructing Morbidity Tables: Phase I, Final Report.

[32] Natvig, B. 2002. Reliability Analysis Encyclopedia of Actuarial Science. Statistical Research Report Number: 4, ISBN 82-553-1351-6

[33] Nelson, W. 1982. Applied Life Morbidity Analysis, John Wiley & Sons, Inc., Canada

[34] Rejda, E. G. 2002. Principles of Risk Managemalet and Insurance. Eighth Edition

[35] http://www.saglik.gov.tr . Page visited at various times in 2010

[36] Straub, E. 1971. Application of Reliability Theory to Insurance, Swiss Reinsurance Company, Zurich.

[37] Tableman, M. and Kim, J. S. 2004. Survival Analysis Using S: Analysis of Time-to-Event Morbidity. Chapman & Hall/CRC.

[38] Topçu, C. 2007. Variance Estimation with Greenwood and Kaplan-Meier Methods. (Master Thesis)

[39] http://www.turksagliksen.org.tr . Page visited at various times in 2010

[40] TÜİK, 2009. Turkey's Statistical Yearbook, 2009. Page visited at various times in 2010

[41] Yetkin, B. 2006. Cox Regresyon Analizi ve Bir Uygulaması (Master thesis)

[42] http://www.who.int. Page visited at various times in 2010

[43] http://www.worldbank.org. Page visited at various times in 2010

[44] http://www.who.int/classifications/icd/implemaletation/en/index.html

APPENDIX A

GOODNESS OF FIT TEST RESULTS

Figure: A.1 Goodness of fit outputs of poisoning for male

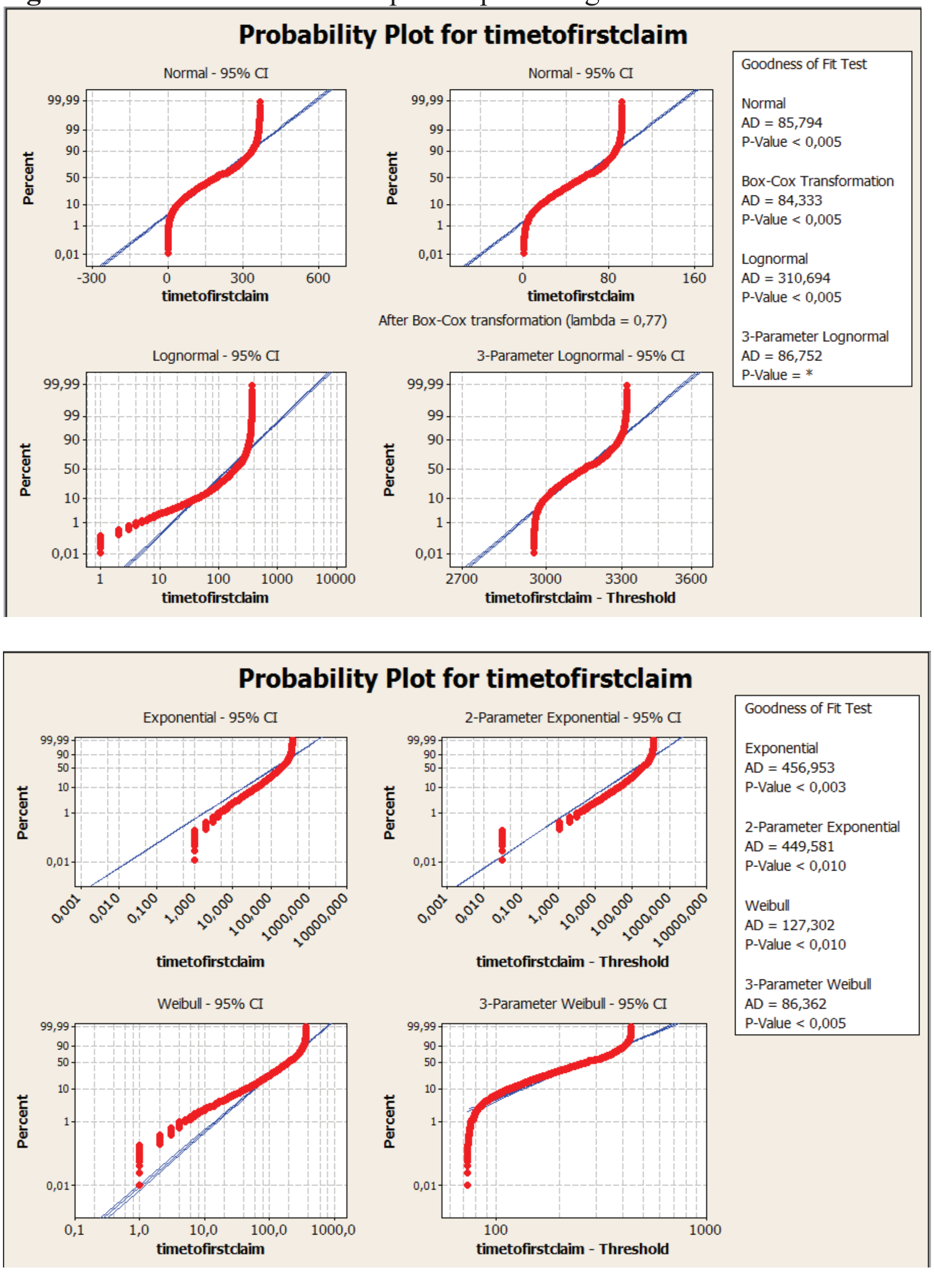
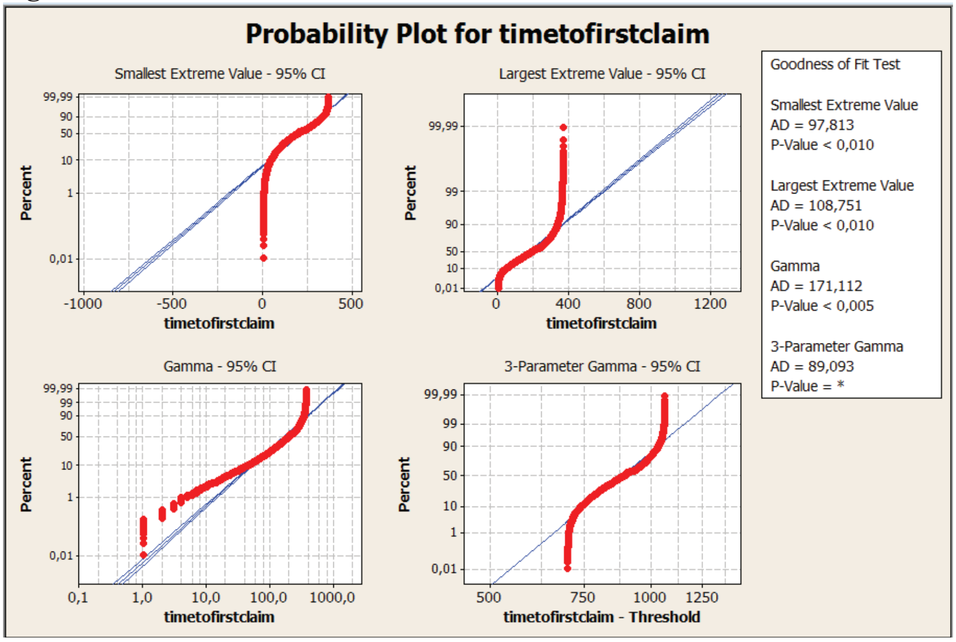


Figure A.1 Continued



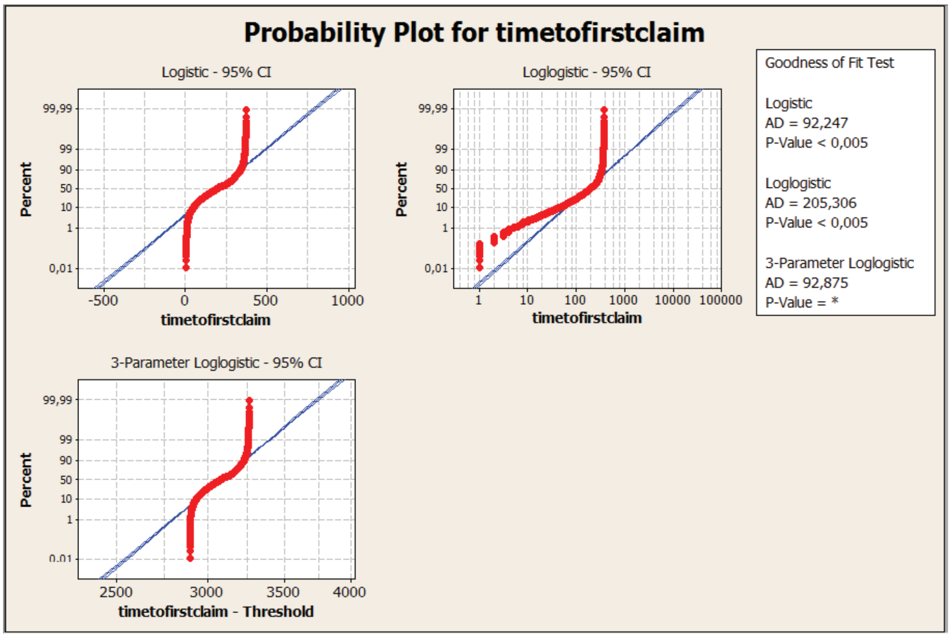
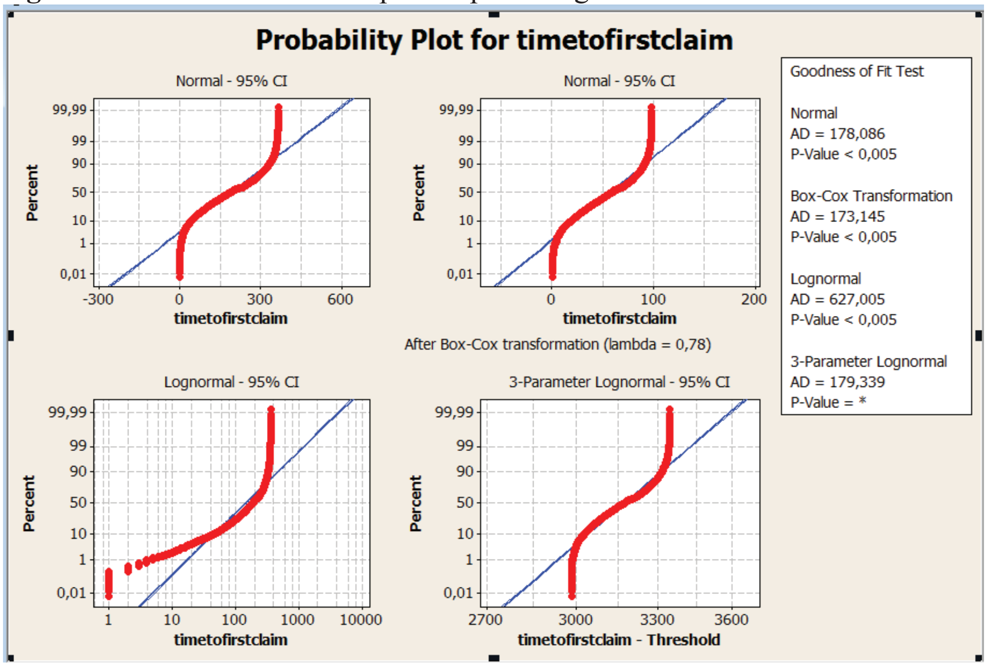


Figure A.2 Goodness of fit outputs of poisoning for female



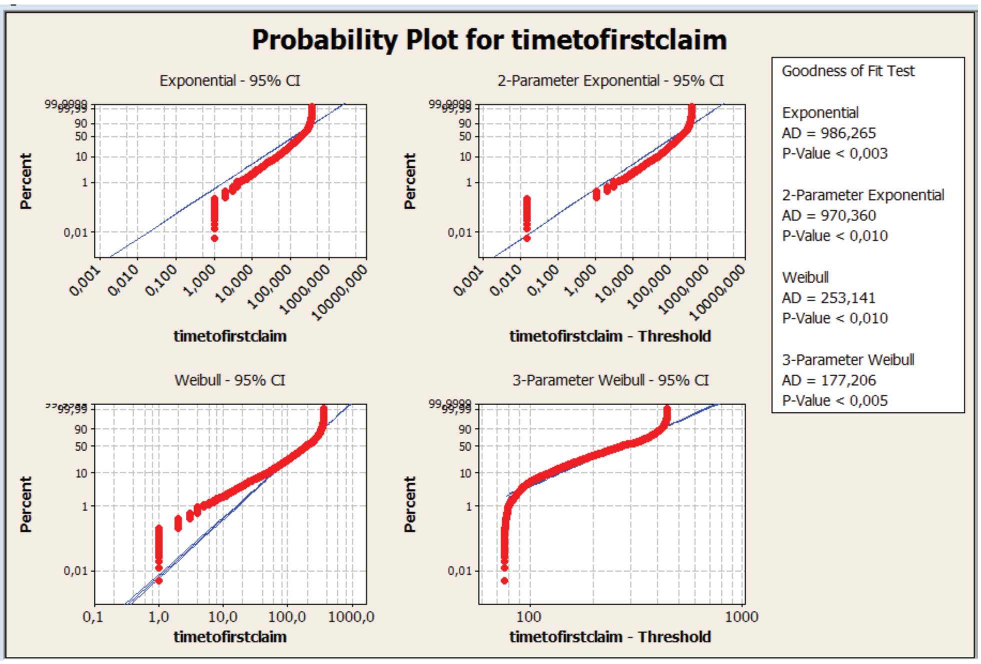
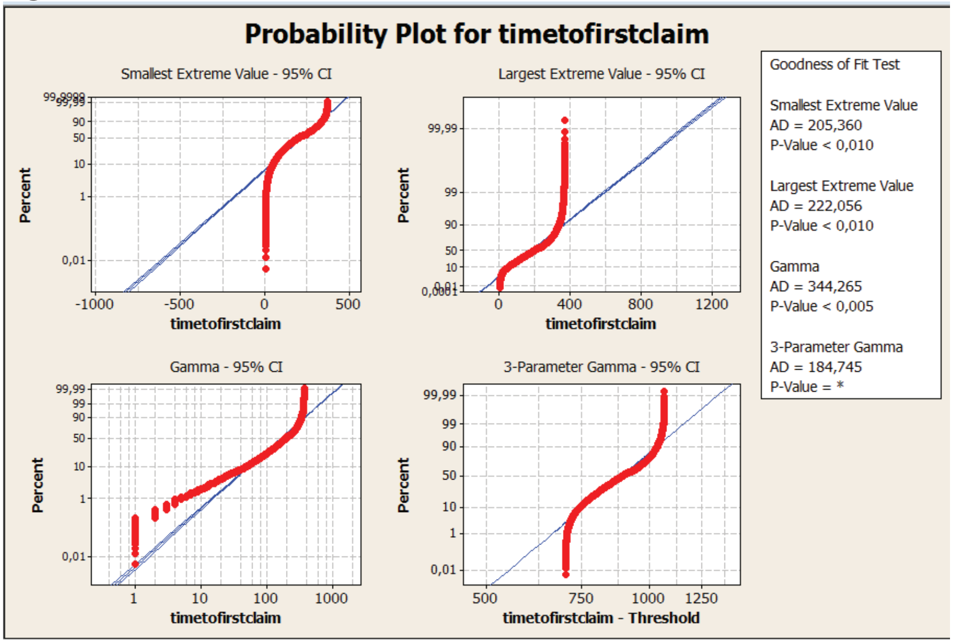


Figure A.2 Continued



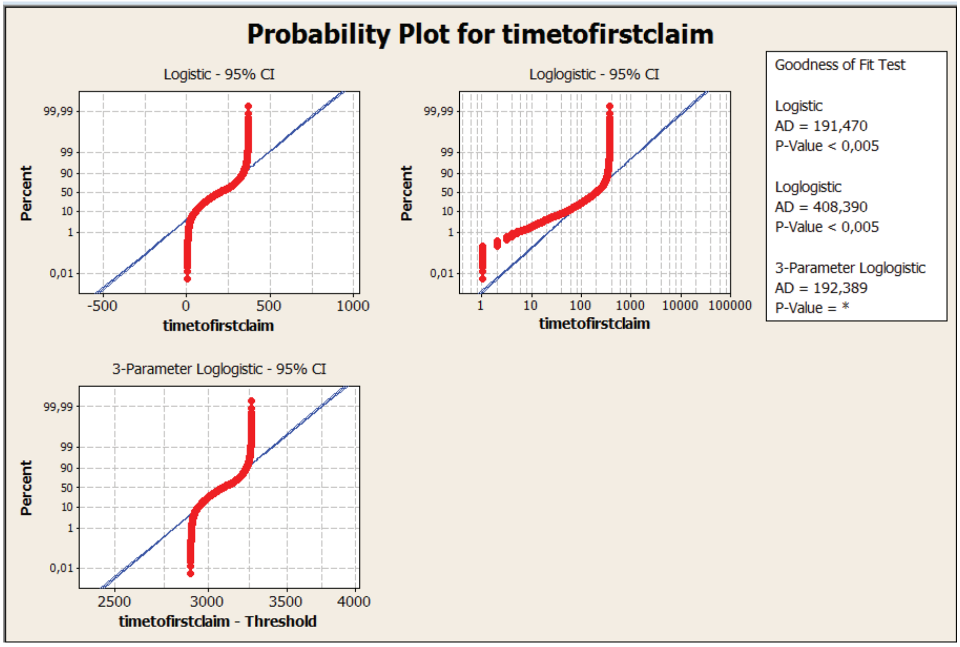
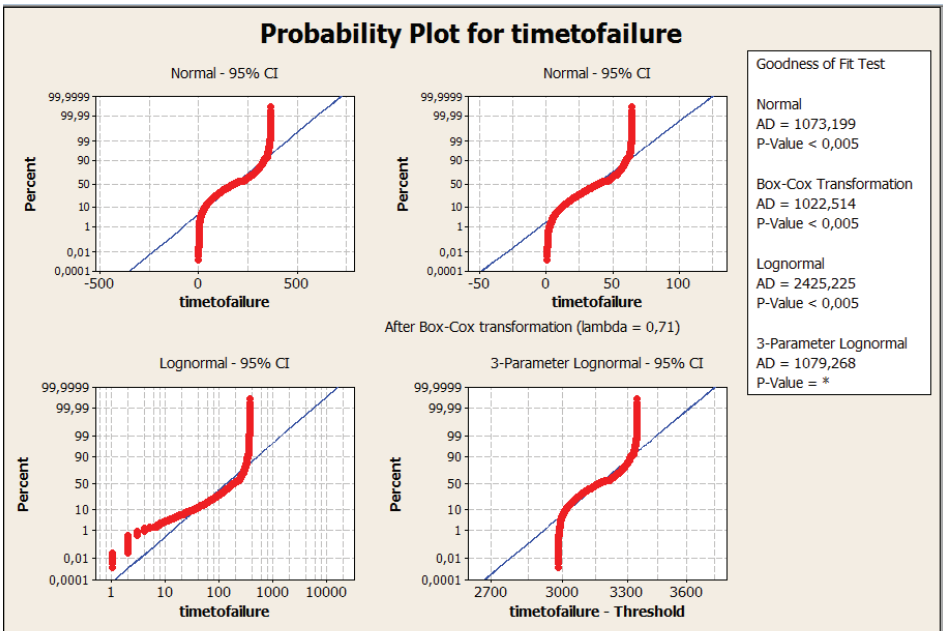


Figure A. 3 Goodness of fit outputs of RHD for male



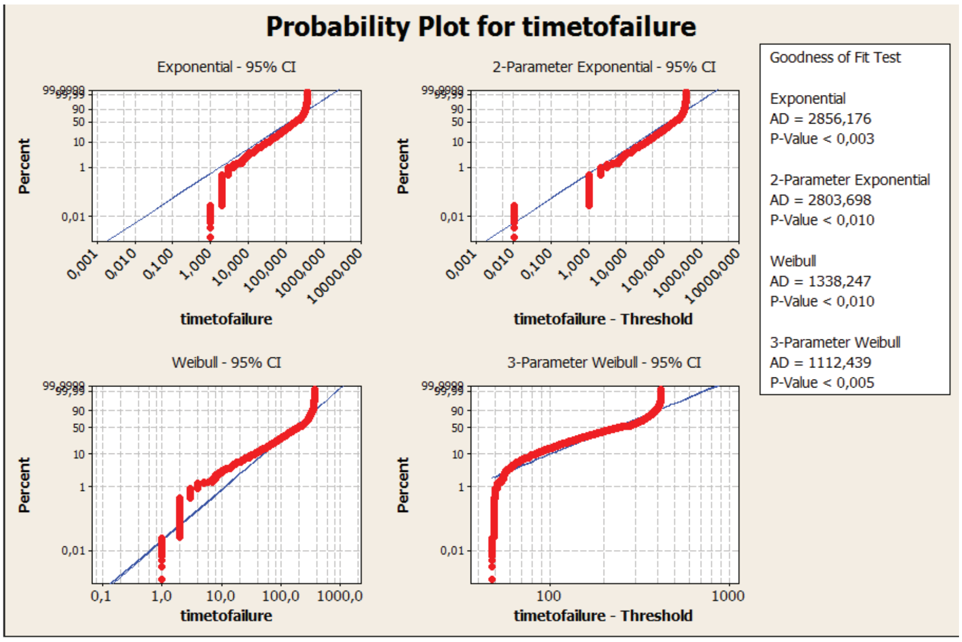
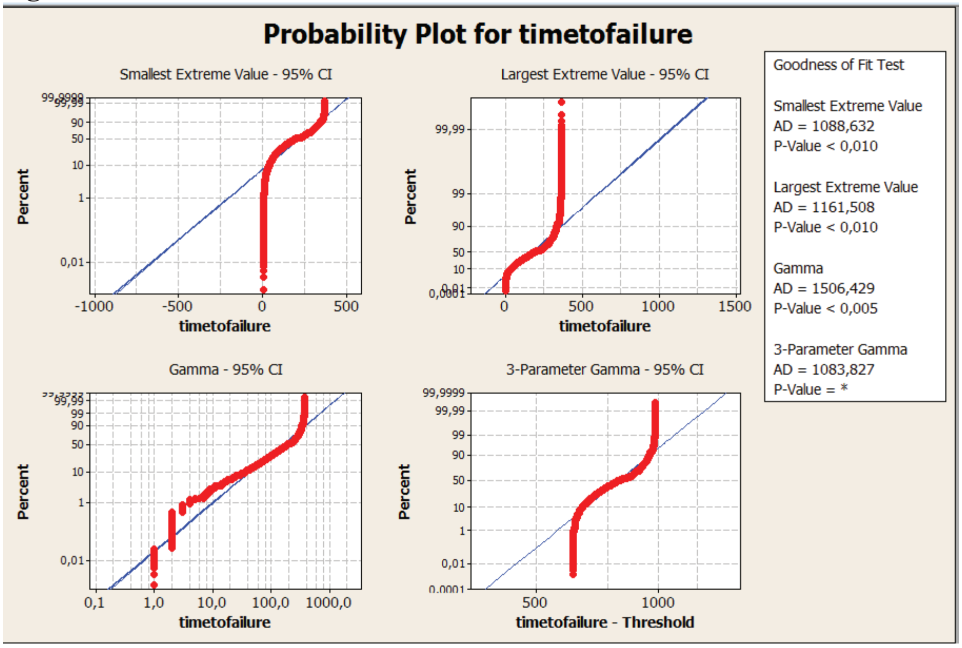


Figure A. 3 Continued



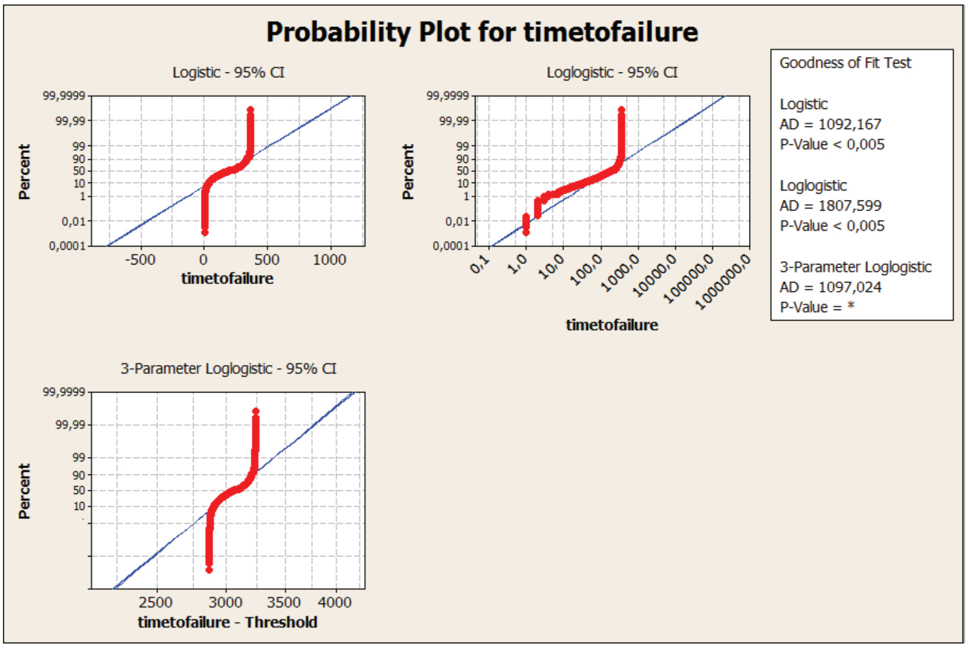
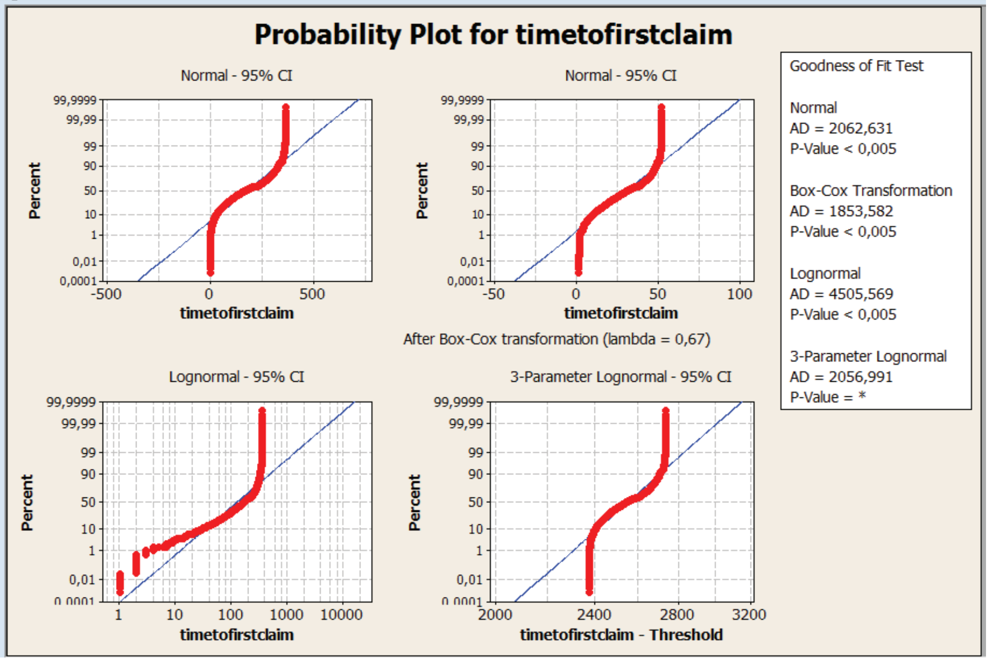


Figure A.4 Goodness of fit outputs of RHD for female



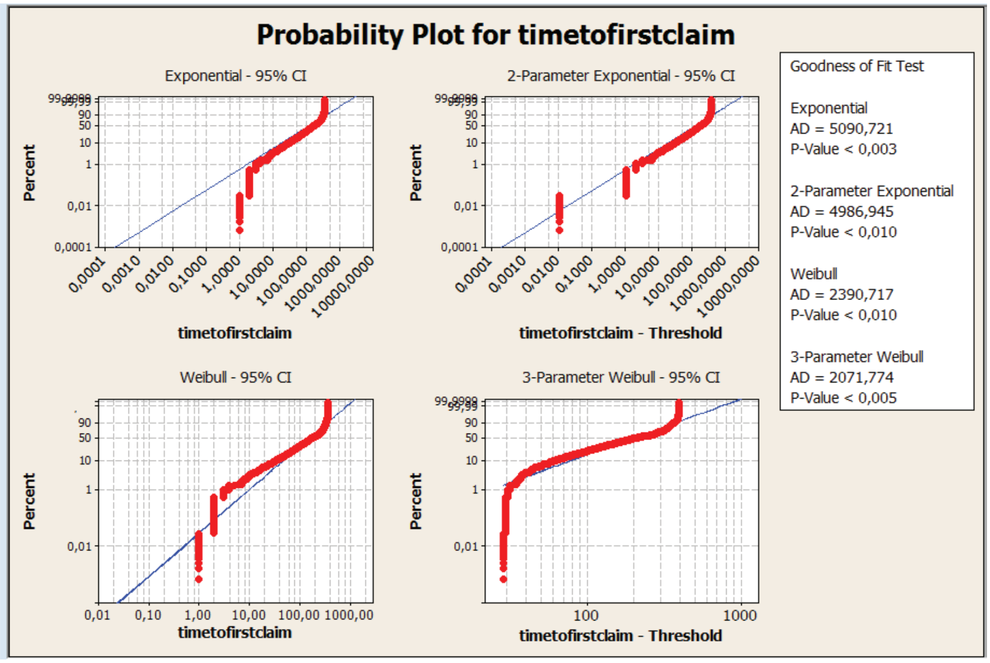
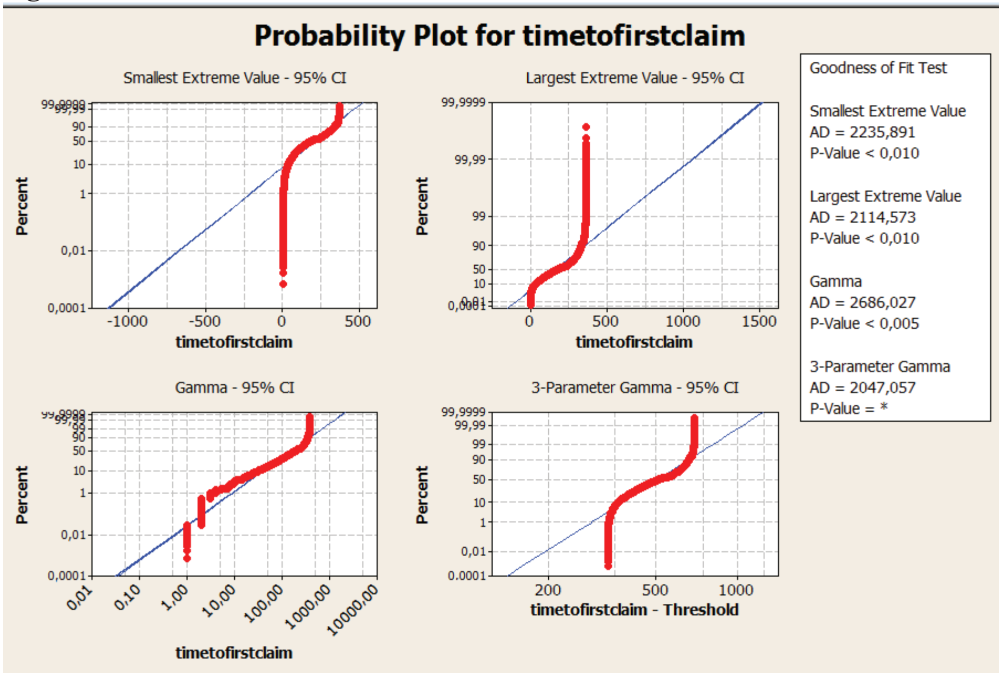
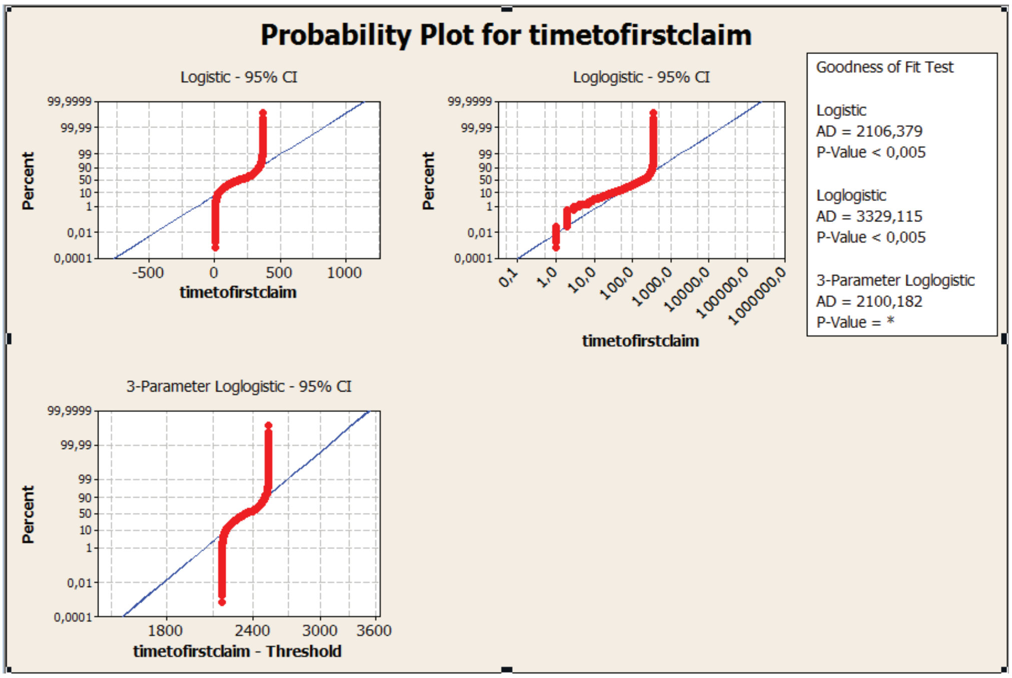


Figure A. 4. Continued





APPENDIX B

Table B.1: Life table of poisoning (Total) (2008)

Interval lower	upper	number entering	number failed	conditional probability of failure	standard error
0	6	19916	261	0,01311	0,0008058
7	13	19655	318	0,01618	0,0008999
14	20	19337	311	0,01608	0,0009046
21	27	19026	329	0,01729	0,0009451
28	34	18697	276	0,01476	0,000882
35	41	18421	294	0,01596	0,0009234
42	48	18127	277	0,01528	0,0009111
49	55	17850	350	0,01961	0,0010378
56	62	17500	352	0,02011	0,0010613
63	69	17148	362	0,02111	0,0010978
70	76	16786	362	0,02157	0,0011212
77	83	16424	355	0,02161	0,0011347
84	90	16069	363	0,02259	0,0011722
91	97	15706	373	0,02375	0,001215
98	104	15333	406	0,02648	0,0012966
105	111	14927	433	0,02901	0,0013737
112	118	14494	388	0,02677	0,0013407
119	125	14106	348	0,02467	0,0013061
126	132	13758	372	0,02704	0,0013828
133	139	13386	397	0,02966	0,0014662
140	146	12989	367	0,02825	0,0014539
147	153	12622	440	0,03486	0,0016327
154	160	12182	422	0,03464	0,0016568
161	167	11760	399	0,03393	0,0016695
168	174	11361	396	0,03486	0,0017208
175	181	10965	379	0,03456	0,0017445
182	188	10586	363	0,03429	0,0017687
189	195	10223	349	0,03414	0,0017959
196	202	9874	658	0,06664	0,0025098
203	209	9216	267	0,02897	0,0017471
210	216	8949	217	0,02425	0,001626
217	223	8732	187	0,02142	0,0015492
224	230	8545	185	0,02165	0,0015744
231	237	8360	277	0,03313	0,0019576
238	244	8083	388	0,048	0,0023777
245	251	7695	393	0,05107	0,0025096
252	258	7302	451	0,06176	0,0028171
259	265	6851	429	0,06262	0,0029271
266	272	6422	415	0,06462	0,0030679
273	279	6007	334	0,0556	0,0029566
280	286	5673	501	0,08831	0,0037673
287	293	5172	485	0,09377	0,0040535
294	300	4687	451	0,09622	0,0043075
301	307	4236	449	0,106	0,0047297
308	314	3787	473	0,1249	0,0053723
315	321	3314	477	0,14393	0,0060976
322	328	2837	458	0,16144	0,0069078
329	335	2379	447	0,18789	0,0080088
336	342	1932	500	0,2588	0,0099643
343	349	1432	371	0,25908	0,0115779
350	356	1061	491	0,46277	0,0153075
357	363	570	422	0,74035	0,0183643
364	371	148	148	1	0

Table B2: Survival probabilities for poisoning per week per year (Total) (2008)

time	survival probability	standard error	95% confidence interval lower	upper
7	0,986895	0,0008058	0,985316	0,98847
14	0,970928	0,0011905	0,968595	0,97326
21	0,955312	0,0014641	0,952443	0,95818
28	0,938793	0,0016986	0,935464	0,94212
35	0,924935	0,0018671	0,921275	0,92859
42	0,910173	0,0020261	0,906202	0,91414
49	0,896264	0,0021606	0,89203	0,9005
56	0,878691	0,0023135	0,874156	0,88323
63	0,861016	0,0024512	0,856212	0,86582
70	0,84284	0,0025789	0,837785	0,8479
77	0,824664	0,0026945	0,819383	0,82995
84	0,806839	0,0027974	0,801356	0,81232
91	0,788612	0,0028931	0,782942	0,79428
98	0,769884	0,0029825	0,764038	0,77573
105	0,749498	0,0030704	0,74348	0,75552
112	0,727757	0,0031541	0,721575	0,73394
119	0,708275	0,003221	0,701962	0,71459
126	0,690801	0,0032749	0,684383	0,69722
133	0,672123	0,0033264	0,665603	0,67864
140	0,652189	0,0033749	0,645575	0,6588
147	0,633762	0,0034138	0,627071	0,64045
154	0,611669	0,0034535	0,6049	0,61844
161	0,59048	0,0034845	0,583651	0,59731
168	0,570446	0,0035076	0,563571	0,57732
175	0,550562	0,0035248	0,543654	0,55747
182	0,531532	0,0035359	0,524602	0,53846
189	0,513306	0,0035417	0,506364	0,52025
196	0,495782	0,0035429	0,488838	0,50273
203	0,462744	0,0035331	0,455819	0,46967
210	0,449337	0,0035247	0,442429	0,45625
217	0,438441	0,003516	0,43155	0,44533
224	0,429052	0,0035071	0,422178	0,43593
231	0,419763	0,0034971	0,412909	0,42662
238	0,405855	0,0034796	0,399035	0,41267
245	0,386373	0,0034503	0,37961	0,39314
252	0,36664	0,0034146	0,359947	0,37333
259	0,343995	0,0033661	0,337397	0,35059
266	0,322454	0,0033121	0,315963	0,32895
273	0,301617	0,0032522	0,295243	0,30799
280	0,284846	0,0031982	0,278578	0,29112
287	0,259691	0,0031069	0,253601	0,26578
294	0,235338	0,0030059	0,229447	0,24123
301	0,212693	0,0028997	0,20701	0,21838
308	0,190149	0,0027807	0,184699	0,1956
315	0,166399	0,0026391	0,161226	0,17157
322	0,142448	0,0024766	0,137594	0,1473
329	0,119452	0,0022981	0,114947	0,12396
336	0,097007	0,0020972	0,092897	0,10112
343	0,071902	0,0018305	0,068314	0,07549
350	0,053274	0,0015914	0,050155	0,05639
357	0,02862	0,0011815	0,026305	0,03094
364	0,007431	0,0006086	0,006238	0,00862

Table B. 3: Hazard rates for poisoning per week per year (Total) (2008)

time	hazard estimates	Standard error	density estimates	standard error
3,5	0,001884	0,0001166	0,0018721	0,0001151
10,5	0,00233	0,0001307	0,002281	0,0001269
17,5	0,002316	0,0001313	0,0022308	0,0001255
24,5	0,002492	0,0001374	0,0023599	0,000129
31,5	0,002124	0,0001279	0,0019797	0,0001183
38,5	0,002298	0,000134	0,0021089	0,0001221
45,5	0,0022	0,0001322	0,0019869	0,0001185
52,5	0,002829	0,0001512	0,0025105	0,000133
59,5	0,002903	0,0001547	0,0025249	0,0001334
66,5	0,003048	0,0001602	0,0025966	0,0001352
73,5	0,003114	0,0001637	0,0025966	0,0001352
80,5	0,003122	0,0001657	0,0025464	0,0001339
87,5	0,003264	0,0001713	0,0026038	0,0001354
94,5	0,003433	0,0001778	0,0026755	0,0001372
101,5	0,003833	0,0001902	0,0029122	0,0001431
108,5	0,004205	0,0002021	0,0031059	0,0001476
115,5	0,003876	0,0001968	0,0027831	0,0001399
122,5	0,003568	0,0001913	0,0024962	0,0001326
129,5	0,003916	0,000203	0,0026683	0,000137
136,5	0,004301	0,0002158	0,0028477	0,0001415
143,5	0,004094	0,0002137	0,0026325	0,0001361
150,5	0,005068	0,0002416	0,0031561	0,0001488
157,5	0,005036	0,0002451	0,003027	0,0001458
164,5	0,004931	0,0002468	0,002862	0,0001418
171,5	0,005068	0,0002546	0,0028405	0,0001413
178,5	0,005025	0,0002581	0,0027186	0,0001383
185,5	0,004984	0,0002616	0,0026038	0,0001354
192,5	0,004962	0,0002656	0,0025034	0,0001328
199,5	0,009848	0,0003837	0,0047198	0,0001809
206,5	0,0042	0,000257	0,0019152	0,0001164
213,5	0,003507	0,000238	0,0015565	0,0001051
220,5	0,003092	0,0002261	0,0013413	0,0000976
227,5	0,003127	0,0002299	0,001327	0,0000971
234,5	0,004813	0,0002892	0,0019869	0,0001185
241,5	0,007026	0,0003566	0,0027831	0,0001399
248,5	0,007487	0,0003776	0,002819	0,0001408
255,5	0,009105	0,0004285	0,003235	0,0001506
262,5	0,009235	0,0004456	0,0030772	0,000147
269,5	0,00954	0,000468	0,0029768	0,0001446
276,5	0,00817	0,0004469	0,0023958	0,00013
283,5	0,013199	0,0005891	0,0035937	0,0001585
290,5	0,014055	0,0006374	0,0034789	0,000156
297,5	0,014441	0,0006791	0,003235	0,0001506
304,5	0,01599	0,0007534	0,0032207	0,0001503
311,5	0,019032	0,0008731	0,0033928	0,0001541
318,5	0,022157	0,0010114	0,0034215	0,0001548
325,5	0,025088	0,0011677	0,0032852	0,0001517
332,5	0,029625	0,0013937	0,0032063	0,0001499
339,5	0,042466	0,0018781	0,0035865	0,0001584
346,5	0,042519	0,0021829	0,0026612	0,0001369
353,5	0,086012	0,0037016	0,0035219	0,000157
360,5	0,167927	0,0066136	0,003027	0,0001458
367,5	0,285714	0	0,0010616	0,0000869

Table B.4: Life table of RHD (Total) (2008)

Interval lower	upper	number entering	number failed	conditional probability of failure	standard error
0	6	151563	2202	0,01453	0,0003074
7	13	149361	3583	0,02399	0,0003959
14	20	145778	3219	0,02208	0,0003849
21	27	142559	3139	0,02202	0,0003887
28	34	139420	2964	0,02126	0,0003863
35	41	136456	3348	0,02454	0,0004188
42	48	133108	2901	0,02179	0,0004002
49	55	130207	2606	0,02001	0,0003881
56	62	127601	3246	0,02544	0,0004408
63	69	124355	3189	0,02564	0,0004483
70	76	121166	2956	0,0244	0,0004432
77	83	118210	3031	0,02564	0,0004597
84	90	115179	3135	0,02722	0,0004795
91	97	112044	2885	0,02575	0,0004732
98	104	109159	3070	0,02812	0,0005004
105	111	106089	3102	0,02924	0,0005173
112	118	102987	2567	0,02493	0,0004858
119	125	100420	3067	0,03054	0,000543
126	132	97353	2944	0,03024	0,0005488
133	139	94409	2885	0,03056	0,0005602
140	146	91524	2199	0,02403	0,0005062
147	153	89325	2934	0,03285	0,0005964
154	160	86391	2695	0,0312	0,0005915
161	167	83696	2477	0,0296	0,0005858
168	174	81219	2462	0,03031	0,0006016
175	181	78757	2530	0,03212	0,0006283
182	188	76227	2333	0,03061	0,0006239
189	195	73894	2283	0,0309	0,0006365
196	202	71611	2194	0,03064	0,000644
203	209	69417	1294	0,01864	0,0005134
210	216	68123	828	0,01215	0,0004198
217	223	67295	826	0,01227	0,0004244
224	230	66469	1027	0,01545	0,0004784
231	237	65442	1750	0,02674	0,0006306
238	244	63692	2991	0,04696	0,0008383
245	251	60701	3385	0,05577	0,0009314
252	258	57316	3512	0,06127	0,0010018
259	265	53804	3481	0,0647	0,0010605
266	272	50323	3144	0,06248	0,0010789
273	279	47179	462	0,00979	0,0004534
280	286	46717	4488	0,09607	0,0013634
287	293	42229	4266	0,10102	0,0014665
294	300	37963	4140	0,10905	0,0015998
301	307	33823	3241	0,09582	0,0016005
308	314	30582	4009	0,13109	0,0019299
315	321	26573	3761	0,14153	0,0021383
322	328	22812	3950	0,17315	0,0025052
329	335	18862	4078	0,2162	0,0029974
336	342	14784	3530	0,23877	0,0035063
343	349	11254	424	0,03768	0,0017949
350	356	10830	4537	0,41893	0,004741
357	363	6293	4310	0,68489	0,0058562
364	371	1983	1983	1	0

Table B.5: Survival probabilities for RHD per week per year (Total) (2008)

time	survival probability	standard error	95% Confidence Interval lower	upper
7	0,985471	0,000307	0,984869	0,9861
14	0,961831	0,000492	0,960866	0,9628
21	0,940592	0,000607	0,939402	0,9418
28	0,919882	0,000697	0,918515	0,9212
35	0,900325	0,00077	0,898817	0,9018
42	0,878235	0,00084	0,876589	0,8799
49	0,859095	0,000894	0,857343	0,8608
56	0,841901	0,000937	0,840064	0,8437
63	0,820484	0,000986	0,818552	0,8224
70	0,799443	0,001029	0,797427	0,8015
77	0,77994	0,001064	0,777854	0,782
84	0,759941	0,001097	0,757791	0,7621
91	0,739257	0,001128	0,737047	0,7415
98	0,720222	0,001153	0,717962	0,7225
105	0,699966	0,001177	0,697659	0,7023
112	0,6795	0,001199	0,67715	0,6818
119	0,662563	0,001215	0,660182	0,6649
126	0,642327	0,001231	0,639914	0,6447
133	0,622903	0,001245	0,620463	0,6253
140	0,603868	0,001256	0,601405	0,6063
147	0,589359	0,001264	0,586882	0,5918
154	0,570001	0,001272	0,567508	0,5725
161	0,552219	0,001277	0,549716	0,5547
168	0,535876	0,001281	0,533365	0,5384
175	0,519632	0,001283	0,517117	0,5221
182	0,502939	0,001284	0,500422	0,5055
189	0,487546	0,001284	0,48503	0,4901
196	0,472483	0,001282	0,46997	0,475
203	0,458008	0,00128	0,455499	0,4605
210	0,44947	0,001278	0,446966	0,452
217	0,444007	0,001276	0,441505	0,4465
224	0,438557	0,001275	0,436059	0,4411
231	0,431781	0,001272	0,429287	0,4343
238	0,420234	0,001268	0,41775	0,4227
245	0,4005	0,001259	0,398033	0,403
252	0,378166	0,001246	0,375725	0,3806
259	0,354994	0,001229	0,352585	0,3574
266	0,332027	0,00121	0,329656	0,3344
273	0,311283	0,001189	0,308952	0,3136
280	0,308235	0,001186	0,30591	0,3106
287	0,278623	0,001152	0,276366	0,2809
294	0,250477	0,001113	0,248295	0,2527
301	0,223161	0,00107	0,221065	0,2253
308	0,201777	0,001031	0,199757	0,2038
315	0,175326	0,000977	0,173412	0,1772
322	0,150512	0,000919	0,148711	0,1523
329	0,12445	0,000848	0,122788	0,1261
336	0,097544	0,000762	0,09605	0,099
343	0,074253	0,000674	0,072933	0,0756
350	0,071455	0,000662	0,070159	0,0728
357	0,041521	0,000512	0,040516	0,0425
364	0,013084	0,000292	0,012512	0,0137

Table B.6: Hazard rates for RHD per week per year (Total) (2008)

time	hazard estimates	Standard error	density estimates	standard error
3,5	0,002091	4,46E-05	0,0020755	4,39E-05
10,5	0,003469	5,79E-05	0,0033772	5,57E-05
17,5	0,00319	5,62E-05	0,0030341	5,29E-05
24,5	0,003181	5,68E-05	0,0029587	5,23E-05
31,5	0,00307	5,64E-05	0,0027937	5,08E-05
38,5	0,003549	6,13E-05	0,0031557	5,39E-05
45,5	0,003148	5,84E-05	0,0027344	5,03E-05
52,5	0,002888	5,66E-05	0,0024563	4,77E-05
59,5	0,003681	6,46E-05	0,0030595	5,31E-05
66,5	0,003711	6,57E-05	0,0030058	5,27E-05
73,5	0,003528	6,49E-05	0,0027862	5,07E-05
80,5	0,003711	6,74E-05	0,0028569	5,14E-05
87,5	0,003942	7,04E-05	0,0029549	5,22E-05
94,5	0,003726	6,94E-05	0,0027193	5,01E-05
101,5	0,004075	7,35E-05	0,0028937	5,17E-05
108,5	0,004239	7,61E-05	0,0029238	0,000052
115,5	0,003606	7,12E-05	0,0024196	4,73E-05
122,5	0,004431	0,00008	0,0028908	5,17E-05
129,5	0,004386	8,08E-05	0,0027749	5,06E-05
136,5	0,004433	8,25E-05	0,0027193	5,01E-05
143,5	0,003474	7,41E-05	0,0020727	4,39E-05
150,5	0,004771	8,81E-05	0,0027655	5,06E-05
157,5	0,004527	8,72E-05	0,0025402	4,85E-05
164,5	0,004291	8,62E-05	0,0023347	4,65E-05
171,5	0,004397	8,86E-05	0,0023206	4,64E-05
178,5	0,004664	9,27E-05	0,0023847	0,000047
185,5	0,00444	9,19E-05	0,002199	4,52E-05
192,5	0,004483	9,38E-05	0,0021519	4,47E-05
199,5	0,004445	9,49E-05	0,002068	4,38E-05
206,5	0,002688	7,47E-05	0,0012197	3,38E-05
213,5	0,001747	6,07E-05	0,0007804	0,000027
220,5	0,001764	6,14E-05	0,0007786	0,000027
227,5	0,002224	6,94E-05	0,000968	3,01E-05
234,5	0,003872	9,25E-05	0,0016495	3,92E-05
241,5	0,00687	0,000126	0,0028192	0,000051
248,5	0,008195	0,000141	0,0031906	5,42E-05
255,5	0,00903	0,000152	0,0033103	5,52E-05
262,5	0,009552	0,000162	0,003281	0,000055
269,5	0,009213	0,000164	0,0029634	5,23E-05
276,5	0,001406	6,54E-05	0,0004355	2,02E-05
283,5	0,014416	0,000215	0,0042302	6,22E-05
290,5	0,015199	0,000232	0,004021	6,07E-05
297,5	0,016478	0,000256	0,0039022	5,98E-05
304,5	0,014378	0,000252	0,0030548	5,31E-05
311,5	0,020041	0,000316	0,0037787	5,89E-05
318,5	0,021759	0,000354	0,003545	5,71E-05
325,5	0,027081	0,000429	0,0037231	5,85E-05
332,5	0,034629	0,000538	0,0038438	5,94E-05
339,5	0,038735	0,000646	0,0033272	5,53E-05
346,5	0,005486	0,000266	0,0003996	1,94E-05
353,5	0,075704	0,001084	0,0042764	6,25E-05
360,5	0,148795	0,001935	0,0040624	0,000061
367,5	0,285714	0	0,0018691	4,17E-05

Table B.7: Life table of acute RHD (Total) (2008)

Interval Lower	Upper	Number entering	Number failed	conditional probability of failure	standard error
0	6	2005	32	0,01596	0,0027988
7	13	1973	65	0,03294	0,0040184
14	20	1908	46	0,02411	0,0035116
21	27	1862	56	0,03008	0,0039581
28	34	1806	69	0,03821	0,0045107
35	41	1737	81	0,04663	0,0050591
42	48	1656	42	0,02536	0,0038635
49	55	1614	55	0,03408	0,0045159
56	62	1559	46	0,02951	0,0042858
63	69	1513	59	0,039	0,0049768
70	76	1454	60	0,04127	0,0052163
77	83	1394	46	0,033	0,0047844
84	90	1348	69	0,05119	0,0060024
91	97	1279	73	0,05708	0,0064868
98	104	1206	72	0,0597	0,0068226
105	111	1134	78	0,06878	0,0075155
112	118	1056	69	0,06534	0,0076048
119	125	987	81	0,08207	0,0087364
126	132	906	70	0,07726	0,0088707
133	139	836	39	0,04665	0,0072938
140	146	797	34	0,04266	0,0071584
147	153	763	33	0,04325	0,0073643
154	160	730	24	0,03288	0,0065997
161	167	706	25	0,03541	0,0069556
168	174	681	27	0,03965	0,0074774
175	181	654	35	0,05352	0,0088006
182	188	619	29	0,04685	0,0084935
189	195	590	23	0,03898	0,0079685
196	202	567	34	0,05996	0,0099708
203	209	533	27	0,05066	0,0094987
210	216	506	6	0,01186	0,0048121
217	223	500	5	0,01	0,0044497
224	230	495	6	0,01212	0,0049184
231	237	489	11	0,02249	0,0067057
238	244	478	27	0,05649	0,0105591
245	251	451	33	0,07317	0,0122625
252	258	418	32	0,07656	0,0130048
259	265	386	36	0,09326	0,0148015
266	272	350	31	0,08857	0,0151871
273	279	319	9	0,02821	0,0092708
280	286	310	19	0,06129	0,0136233
287	293	291	36	0,12371	0,0193011
294	300	255	23	0,0902	0,017939
301	307	232	23	0,09914	0,0196203
308	314	209	27	0,12919	0,0232005
315	321	182	19	0,1044	0,0226654
322	328	163	33	0,20245	0,0314737
329	335	130	26	0,2	0,0350823
336	342	104	31	0,29808	0,0448531
343	349	73	3	0,0411	0,0232341
350	356	70	31	0,44286	0,0593699
357	363	39	25	0,64103	0,0768134
364	371	14	14	1	0

Table B.8: Survival probabilities for acute RHD per week per year (Total) (2008)

Time	Survival probability	standard error	95% confidence interval lower	upper
7	0,98404	0,0027988	0,978554	0,989525
14	0,951621	0,0047919	0,942229	0,961013
21	0,928678	0,0057476	0,917413	0,939943
28	0,900748	0,0066775	0,88766	0,913836
35	0,866334	0,0075997	0,851439	0,881229
42	0,825935	0,0084678	0,809339	0,842532
49	0,804988	0,0088485	0,787645	0,82233
56	0,777556	0,0092879	0,759352	0,79576
63	0,754613	0,0096102	0,735778	0,773449
70	0,725187	0,0099698	0,705647	0,744727
77	0,695262	0,0102797	0,675114	0,71541
84	0,672319	0,0104823	0,651774	0,692864
91	0,637905	0,0107333	0,616868	0,658942
98	0,601496	0,0109339	0,580066	0,622926
105	0,565586	0,0110699	0,543889	0,587283
112	0,526683	0,0111505	0,504829	0,548538
119	0,492269	0,0111651	0,470386	0,514152
126	0,45187	0,0111145	0,430086	0,473654
133	0,416958	0,0110113	0,395376	0,438539
140	0,397506	0,0109293	0,376085	0,418927
147	0,380549	0,0108431	0,359297	0,401801
154	0,36409	0,010746	0,343028	0,385151
161	0,35212	0,0106668	0,331213	0,373026
168	0,339651	0,0105766	0,318921	0,360381
175	0,326185	0,01047	0,305664	0,346705
182	0,308728	0,010317	0,288507	0,328949
189	0,294264	0,0101773	0,274317	0,314211
196	0,282793	0,0100577	0,26308	0,302506
203	0,265835	0,0098661	0,246498	0,285173
210	0,252369	0,0097007	0,233356	0,271382
217	0,249377	0,0096623	0,230439	0,268314
224	0,246883	0,0096298	0,228009	0,265757
231	0,24389	0,0095903	0,225094	0,262687
238	0,238404	0,0095162	0,219753	0,257055
245	0,224938	0,0093249	0,206661	0,243214
252	0,208479	0,009072	0,190698	0,22626
259	0,192519	0,0088053	0,175261	0,209777
266	0,174564	0,0084774	0,157948	0,191179
273	0,159102	0,0081687	0,143092	0,175113
280	0,154613	0,0080741	0,138789	0,170438
287	0,145137	0,0078665	0,129719	0,160555
294	0,127182	0,0074408	0,112598	0,141766
301	0,115711	0,0071438	0,101709	0,129712
308	0,104239	0,0068242	0,090864	0,117615
315	0,090773	0,0064159	0,078198	0,103348
322	0,081297	0,0061033	0,069334	0,093259
329	0,064838	0,0054992	0,05406	0,075616
336	0,05187	0,0049526	0,042163	0,061577
343	0,036409	0,0041831	0,02821	0,044608
350	0,034913	0,0040994	0,026878	0,042947
357	0,019451	0,0030843	0,013406	0,025496
364	0,006983	0,0018596	0,003338	0,010627

Table B.9: Hazard rates for acute RHD per week per year (Total) (2008)

time	hazard estimates	standard error	density estimates	standard error
3,5	0,002298	0,0004063	0,00228	0,0003998
10,5	0,004785	0,0005934	0,0046313	0,0005651
17,5	0,003486	0,000514	0,0032775	0,0004777
24,5	0,004362	0,0005828	0,00399	0,0005257
31,5	0,005564	0,0006697	0,0049163	0,0005816
38,5	0,006821	0,0007576	0,0057713	0,0006282
45,5	0,00367	0,0005662	0,0029925	0,0004569
52,5	0,004953	0,0006677	0,0039188	0,0005211
59,5	0,004278	0,0006307	0,0032775	0,0004777
66,5	0,005682	0,0007395	0,0042038	0,0005392
73,5	0,006019	0,0007769	0,004275	0,0005436
80,5	0,004793	0,0007066	0,0032775	0,0004777
87,5	0,007504	0,0009031	0,0049163	0,0005816
94,5	0,008393	0,0009819	0,0052013	0,0005976
101,5	0,008791	0,0010356	0,00513	0,0005936
108,5	0,010176	0,0011515	0,0055575	0,0006169
115,5	0,00965	0,001161	0,0049163	0,0005816
122,5	0,012225	0,0013571	0,0057713	0,0006282
129,5	0,011481	0,0013711	0,0049875	0,0005856
136,5	0,006824	0,0010923	0,0027788	0,0004406
143,5	0,006227	0,0010677	0,0024225	0,0004119
150,5	0,006315	0,0010991	0,0023513	0,0004059
157,5	0,004775	0,0009746	0,00171	0,000347
164,5	0,00515	0,0010298	0,0017813	0,000354
171,5	0,005778	0,0011118	0,0019238	0,0003677
178,5	0,007855	0,0013273	0,0024938	0,0004178
185,5	0,006853	0,0012723	0,0020663	0,0003809
192,5	0,00568	0,0011841	0,0016388	0,0003397
199,5	0,008831	0,0015138	0,0024225	0,0004119
206,5	0,007425	0,0014284	0,0019238	0,0003677
213,5	0,001704	0,0006957	0,0004275	0,0001743
220,5	0,001436	0,0006421	0,0003563	0,0001591
227,5	0,001742	0,0007112	0,0004275	0,0001743
234,5	0,00325	0,0009799	0,0007838	0,0002357
241,5	0,008304	0,0015974	0,0019238	0,0003677
248,5	0,01085	0,0018874	0,0023513	0,0004059
255,5	0,011372	0,0020087	0,00228	0,0003998
262,5	0,013975	0,0023264	0,002565	0,0004236
269,5	0,013239	0,0023753	0,0022088	0,0003936
276,5	0,004088	0,0013626	0,0006413	0,0002133
283,5	0,009033	0,0020712	0,0013538	0,0003091
290,5	0,018838	0,0031329	0,002565	0,0004236
297,5	0,013494	0,0028105	0,0016388	0,0003397
304,5	0,014901	0,0031029	0,0016388	0,0003397
311,5	0,01973	0,0037879	0,0019238	0,0003677
318,5	0,015735	0,0036044	0,0013538	0,0003091
325,5	0,032179	0,0055661	0,0023513	0,0004059
332,5	0,031746	0,0061874	0,0018525	0,0003609
339,5	0,05004	0,0088486	0,0022088	0,0003936
346,5	0,005994	0,0034599	0,0002138	0,0001233
353,5	0,081258	0,0139917	0,0022088	0,0003936
360,5	0,134771	0,0237671	0,0017813	0,000354
367,5	0,285714	0	0,0009975	0,0002657

Table B.10: Life table of chronic RHD (Total) (2008)

Interval Lower	Upper	Number entering	Number failed	conditional probability of failure	standard error
0	6	149558	51565	0,34478	0,001229
7	13	97993	2357	0,02405	0,0004894
14	20	95636	2180	0,02279	0,0004826
21	27	93456	2119	0,02267	0,0004869
28	34	91337	2014	0,02205	0,0004859
35	41	89323	2185	0,02446	0,0005169
42	48	87138	1912	0,02194	0,0004963
49	55	85226	1697	0,01991	0,0004785
56	62	83529	2108	0,02524	0,0005427
63	69	81421	2085	0,02561	0,0005536
70	76	79336	1960	0,02471	0,0005511
77	83	77376	2051	0,02651	0,0005775
84	90	75325	2093	0,02779	0,0005989
91	97	73232	1890	0,02581	0,0005859
98	104	71342	2041	0,02861	0,0006241
105	111	69301	2038	0,02941	0,0006418
112	118	67263	1705	0,02535	0,0006061
119	125	65558	1985	0,03028	0,0006692
126	132	63573	1925	0,03028	0,0006796
133	139	61648	1919	0,03113	0,0006994
140	146	59729	1435	0,02403	0,0006266
147	153	58294	1988	0,0341	0,0007517
154	160	56306	1722	0,03058	0,0007256
161	167	54584	1632	0,0299	0,000729
168	174	52952	1635	0,03088	0,0007517
175	181	51317	1702	0,03317	0,0007905
182	188	49615	1541	0,03106	0,0007788
189	195	48074	1510	0,03141	0,0007955
196	202	46564	1461	0,03138	0,0008079
203	209	45103	872	0,01933	0,0006484
210	216	44231	531	0,01201	0,0005178
217	223	43700	549	0,01256	0,0005328
224	230	43151	691	0,01601	0,0006043
231	237	42460	1189	0,028	0,0008007
238	244	41271	1982	0,04802	0,0010525
245	251	39289	2161	0,055	0,0011502
252	258	37128	2236	0,06022	0,0012347
259	265	34892	2249	0,06446	0,0013146
266	272	32643	1996	0,06115	0,0013261
273	279	30647	277	0,00904	0,0005406
280	286	30370	2982	0,09819	0,0017075
287	293	27388	2814	0,10275	0,0018347
294	300	24574	2731	0,11113	0,0020049
301	307	21843	2108	0,09651	0,001998
308	314	19735	2634	0,13347	0,0024208
315	321	17101	2491	0,14566	0,0026976
322	328	14610	2543	0,17406	0,0031369
329	335	12067	2636	0,21845	0,0037614
336	342	9431	2249	0,23847	0,0043881
343	349	7182	265	0,0369	0,0022244
350	356	6917	2862	0,41376	0,0059218
357	363	4055	2780	0,68557	0,0072911
364	371	1275	1275	1	0

Table B.11: Survival probabilities for chronic RHD per week per year (Total) (2008)

Time	Survival probability	standard error	95% confidence interval lower	upper
7	0,655217	0,001229	0,652809	0,657626
14	0,639458	0,0012416	0,637024	0,641891
21	0,624881	0,0012519	0,622428	0,627335
28	0,610713	0,0012608	0,608242	0,613184
35	0,597247	0,0012682	0,594761	0,599732
42	0,582637	0,0012751	0,580138	0,585136
49	0,569852	0,0012802	0,567343	0,572362
56	0,558506	0,001284	0,555989	0,561022
63	0,544411	0,0012878	0,541887	0,546935
70	0,53047	0,0012905	0,52794	0,532999
77	0,517365	0,0012921	0,514832	0,519897
84	0,503651	0,0012929	0,501117	0,506185
91	0,489656	0,0012926	0,487123	0,49219
98	0,477019	0,0012915	0,474488	0,47955
105	0,463372	0,0012894	0,460845	0,465899
112	0,449745	0,0012864	0,447224	0,452266
119	0,438345	0,001283	0,43583	0,44086
126	0,425073	0,0012783	0,422567	0,427578
133	0,412201	0,0012728	0,409707	0,414696
140	0,39937	0,0012664	0,396888	0,401852
147	0,389775	0,0012611	0,387304	0,392247
154	0,376483	0,0012528	0,374027	0,378938
161	0,364969	0,0012449	0,362529	0,367409
168	0,354057	0,0012366	0,351633	0,35648
175	0,343124	0,0012276	0,340718	0,34553
182	0,331744	0,0012175	0,329358	0,33413
189	0,321441	0,0012076	0,319074	0,323807
196	0,311344	0,0011973	0,308997	0,313691
203	0,301575	0,0011867	0,299249	0,303901
210	0,295745	0,0011801	0,293432	0,298058
217	0,292194	0,0011759	0,28989	0,294499
224	0,288524	0,0011716	0,286227	0,29082
231	0,283903	0,0011659	0,281618	0,286188
238	0,275953	0,0011558	0,273688	0,278219
245	0,262701	0,001138	0,26047	0,264931
252	0,248252	0,0011171	0,246062	0,250441
259	0,233301	0,0010936	0,231157	0,235444
266	0,218263	0,0010681	0,21617	0,220357
273	0,204917	0,0010437	0,202871	0,206963
280	0,203065	0,0010402	0,201026	0,205104
287	0,183126	0,0010001	0,181166	0,185086
294	0,164311	0,0009582	0,162433	0,166189
301	0,14605	0,0009132	0,144261	0,14784
308	0,131955	0,0008751	0,13024	0,133671
315	0,114344	0,0008229	0,112731	0,115956
322	0,097688	0,0007677	0,096183	0,099193
329	0,080684	0,0007042	0,079304	0,082065
336	0,063059	0,0006285	0,061827	0,064291
343	0,048022	0,0005529	0,046938	0,049105
350	0,04625	0,0005431	0,045185	0,047314
357	0,027113	0,00042	0,02629	0,027936
364	0,008525	0,0002377	0,008059	0,008991

Table B.12: Hazard rates for chronic RHD per week per year (Total) (2008)

time	hazard estimates	standard error	density estimates	standard error
3,5	0,059514	0,0002563	0,0492547	0,0001756
10,5	0,003478	0,0000716	0,0022514	0,000046
17,5	0,003294	0,0000705	0,0020823	0,0000443
24,5	0,003276	0,0000712	0,0020241	0,0000437
31,5	0,003185	0,000071	0,0019238	0,0000426
38,5	0,003538	0,0000757	0,0020871	0,0000443
45,5	0,003169	0,0000725	0,0018263	0,0000415
52,5	0,002873	0,0000697	0,001621	0,0000391
59,5	0,003651	0,0000795	0,0020136	0,0000435
66,5	0,003706	0,0000811	0,0019916	0,0000433
73,5	0,003573	0,0000807	0,0018722	0,000042
80,5	0,003838	0,0000847	0,0019591	0,000043
87,5	0,004025	0,000088	0,0019992	0,0000434
94,5	0,003735	0,0000859	0,0018053	0,0000413
101,5	0,004146	0,0000918	0,0019496	0,0000429
108,5	0,004264	0,0000944	0,0019467	0,0000428
115,5	0,003668	0,0000888	0,0016286	0,0000392
122,5	0,004392	0,0000986	0,0018961	0,0000423
129,5	0,004392	0,0001001	0,0018388	0,0000416
136,5	0,004517	0,0001031	0,001833	0,0000416
143,5	0,003474	0,0000917	0,0013707	0,000036
150,5	0,004956	0,0001111	0,0018989	0,0000423
157,5	0,004437	0,0001069	0,0016448	0,0000394
164,5	0,004336	0,0001073	0,0015589	0,0000384
171,5	0,00448	0,0001108	0,0015617	0,0000384
178,5	0,004818	0,0001168	0,0016257	0,0000392
185,5	0,004507	0,0001148	0,001472	0,0000373
192,5	0,004559	0,0001173	0,0014423	0,0000369
199,5	0,004554	0,0001191	0,0013955	0,0000363
206,5	0,002789	0,0000944	0,0008329	0,0000281
213,5	0,001725	0,0000749	0,0005072	0,000022
220,5	0,001806	0,0000771	0,0005244	0,0000223
227,5	0,002306	0,0000877	0,00066	0,0000251
234,5	0,004057	0,0001177	0,0011357	0,0000328
241,5	0,007029	0,0001578	0,0018932	0,0000422
248,5	0,00808	0,0001737	0,0020642	0,0000441
255,5	0,008871	0,0001875	0,0021358	0,0000448
262,5	0,009515	0,0002005	0,0021482	0,000045
269,5	0,009011	0,0002016	0,0019066	0,0000424
276,5	0,001297	0,0000779	0,0002646	0,0000159
283,5	0,014751	0,0002698	0,0028484	0,0000516
290,5	0,015473	0,0002913	0,0026879	0,0000502
297,5	0,01681	0,0003211	0,0026086	0,0000495
304,5	0,014486	0,0003151	0,0020136	0,0000435
311,5	0,02043	0,0003971	0,002516	0,0000486
318,5	0,022444	0,0004483	0,0023794	0,0000473
325,5	0,027236	0,0005376	0,0024291	0,0000478
332,5	0,035033	0,0006772	0,0025179	0,0000486
339,5	0,038679	0,0008081	0,0021482	0,000045
346,5	0,00537	0,0003298	0,0002531	0,0000155
353,5	0,074527	0,0013449	0,0027338	0,0000506
360,5	0,149022	0,0024115	0,0026554	0,0000499
367,5	0,285714	0	0,0012179	0,000034