TECHNICAL UNIVERSITY OF CRETE

DEPARTMENT OF ELECTRONICS
AND COMPUTER ENGINEERING

Survival Analysis on Ocular Melanoma patients using semi parametric regression models and Neural Networks

UNDERGRADUATE DIPLOMA THESIS

AGRAFIOTI FOTEINI

Examination Committee:

Zervakis Michalis
Liavas Athanasios
Balas Kostas

Chania 2006

Dedicated to my father who taught me how to think.

Abstract

Survival analysis is vital in medical statistics. It is concerned with the modeling of lifetime data. In many biomedical applications the primary interest is the time it takes for an event to occur. Therefore, survival analysis is concerned with studying the time between entry to a study and a subsequent event. Patients suffering from ocular melanoma are studied in order to plot survival curves, which are prognostic curves showing the proportion of people surviving at each time spot. The non parametric Kaplan Meier model (product limit estimate) is used to obtain a survival distribution from which important medical conclusions can be made. The log rank test is used to reject the hypothesis that patients with epithelioid cells and patients without them have the same behavior concerning their survival probabilities. In addition, the semi parametric Cox proportional hazards model has been implemented not only to plot survival and hazard curves but also to identify important prognostic factors. These factors are measurements taken from the patients while they were under observation. This semi parametric model is used to discover the influence ability of each measurement to the evolution of the disease. Finally, a fully parametric model is implemented, the log logistic regression model. In this case the survival distribution is known and regression is applied to identify the effect of covariates. Artificial neural networks are trained as well to learn the behavior of ocular melanoma patients regarding survival. The Cox model is approximated through neural networks, which have the ability to generalize and this is the reason why they are a powerful statistical tool.

Acknowledgments

I would like to thank the people that I had the opportunity to cooperate for their kind help and contribution. First, I appreciate and acknowledge the valuable insights and observations contributed by my professor, Dr. Zervakis Michalis. His guidance in this thesis as well as his advice made it possible for me to overcome the difficulties that came up. In addition, his moral support was more than valuable.

I would also like to thank the examination committee, Dr. Balas Kostas and Dr. Liavas Athanasios for their support. I owe many thanks to the scientists involved in the Biopattern program for their continuing support and contributions. In addition, I appreciate the help and cooperation with Kourkouta Anna Maria during the preparation of this thesis for her suggestions and ideas. Finally, I would like to thank the Msc student Papadakis Kostantinos for sharing with me his knowledge on artificial neural networks.

Contents

Chapter 1
Introduction

This Chapter focuses on analyzing the ocular melanoma disease by explaining its symptoms, epidemiology and treatments. In addition, the idea of survival analysis is introduced and there is a brief overview of the existing models. Furthermore, the advantages of neural networks and their applications are discussed.

1.1 The Ocular Melanoma Disease

The ocular melanoma is disease of adults but might also affect children. The incidence in U.S. and Europe is 5 to 7.5 per million per year. Also for adults over 50 years old the incidence rate is 21 per million per year.

The Ocular melanoma is the melanoma of the eye. Melanoma is a cancer that develops from cells called *melanoytes* .These cells produce dark -colored pigment, the melanin which is considered to be responsible for the color of our skin. These cells are found in many locations in the human body including the eye. There are four types of ocular melanoma. The *Uveal (Choroidal)* is the most commonly faced (85% of ocular melanomas), the *Ciliary body* (7-10% of ocular melanomas) which is unfortunately detected late since the tumour does not impede vision until it reaches a certain size. The *Iris melanoma* (5-8% of ocular melanomas) which is most common at the age of 44, and the *Conjunctiva melanoma* which is the least common, as it is the 2% of the ocular melanomas.

The choroidal is part of the lining of the eyeball and is dark-colored (*pigmented*) to prevent light being reflected around the inside of the eye. The Ciliary body extends from the choroidal and focuses the eye by changing the shape of the lens. The iris is the clearly visible colored disc at the front of the eye, which controls the amount of light entering the eye. All these structures are heavily colored with melanin.

Ocular melanoma is the most common type of cancer to affect the eye, although, generally, it is still quite rare. Approximately 350 new cases of ocular melanoma are diagnosed each year in the UK. The incidence of ocular melanoma increases with age, and most are diagnosed in people in their 60s.

Since the ocular melanoma is a rare type of tumor and as for many other forms of cancer the exact cause is unknown. It is known that exposure to ultraviolet (UV) rays (either from the sun or sun beds) increases the risk of developing melanoma of the skin. People whose skin burns easily are most at risk: typically, people with fair skin, fair or red hair and blue

eyes. However, it is not yet known whether there is any link between UV ray exposure and the development of melanoma of the eye.[4]

Symptoms include blurred vision, flashing lights, shadows and misting of the lens of the eye (*cataract*). Often no symptoms are noticed until the tumor is quite large. All of these symptoms are common to other conditions of the eye, but it is generally possible for an eye specialist (*ophthalmologist*) to diagnose these tumors quite simply and painlessly. Occasionally a *biopsy* (taking a small sample of tissue) is needed to confirm a diagnosis. There are many methods to diagnose the ocular melanoma disease. Some tests that could be performed are:

- **Ophthalmoscopy** A small hand-held microscope (*ophthalmoscope*), similar to those used by opticians during routine eye tests, is used to look at the inside of the eye. This is likely to be the first test that you have
- **Ultrasound scan** A small device which produces sound waves is rubbed over the skin around the eye area. The echoes are then converted into a picture by a computer
- **Fluorescein angiography** A special dye, called flourescein, is injected into a vein in the arm. In a few seconds, the dye travels to the blood vessels inside the eye. A camera with special filters that highlight the dye is used to photograph the flourescein as it circulates through the blood vessels in the retina and choroid
- **CT (computerised tomography) scan** A CT scan takes a series of x-rays to build a three-dimensional picture of the inside of the head. The scan is painless but takes 10 minutes, longer than a standard x-ray. It may be used to find the tumour within the eye or to check for any spread of the disease
- **MRI (magnetic resonance imaging) scan** This type of scanner uses magnetism instead of x-rays to form a series of pictures of the inside of the head. The test can take about 30 minutes and is completely painless, although the machine is noisy and you will be given earplugs or headphones to wear
- **Biopsy** A small sample of tissue may be taken from the suspicious area and examined under a microscope. However, this is not necessary for most ocular melanomas because they have a distinctive appearance and can usually be recognized easily from the x-rays and scans.

For 100 years or longer, the usual treatment for choroidal melanoma has been removal of the eye, or enucleation. If the tumor has not spread to other parts of the body, then removal of the eye rids the patient of the tumor. Since World War II, radiation treatment has been used for choroidal melanoma. During the past 20 years, this method of treatment has been refined. Radiation, at the appropriate dose rates and in the proper physical forms, is intended to eliminate growing tumor cells without causing damage to normal tissue sufficient to require removal of the eye. As the cells die, the tumor shrinks, but it usually does not disappear entirely.

The kind of treatment for the ocular melanoma depends on the size of tumor, the location and the cell type. There are also other factor the affect the treatment methodology concerning the general health of the patient the age and the level of vision in both eyes. Therefore depending on the above limitations the treatment might be internal or external radiotherapy Transpupillary thermotherapy or surgery. The radiotherapy uses high energy rays to destroy the cancer cells, but it might harm a lithe the healthy cells that surround the cancer. The Transpupillary thermotherapy can be used to treat very small ocular melanomas or can be used as an additional treatment after radiotherapy. The surgery is a treatment method that mostly depends on the location of the tumor. If the cancer is very severe then the eyeball must first be removed to access the cancer cells. This is usually takes a lot of time and is painful.

Research into treatments for ocular melanoma is ongoing and advances are being made. Cancer doctors use clinical trials to assess new treatments. Before any trial is allowed to take place an ethics committee must have approved it and agreed that the trial is in the interest of patients. [3]

1.2 Overview of survival analysis

Survival Analysis is central to medical statistics for two reasons. First, because survival is an important medical concern, and second because survival analysis can be used to analyze data with outcomes other than deaths that are otherwise not analyzable. Some examples of such non fatal outcomes include time to tumor recurrence and age of developmental milestones. Survival analysis is concerned with studying the time between entry to a study and a subsequent event. Originally the analysis was concerned with time from treatment until death, but survival analysis is applicable to many areas as well as mortality.

Survival analysis is a statistical procedure that analyses time to event data. It has multiple applications since it can be used to model the time to events such as the diagnosis of a disease, the effect of treatments and drugs, remission of cancer (medicine), felon's time to parole (criminology), lifetime of electronic devices (engineering), length of magazine subscription, effectiveness of employees (marketing), duration of marriage and travel habits (sociology).

In medicine, a survival curve is a statistical picture of the survival experience of some group of patients which is a graph showing the percentage surviving against time. In addition several methods can be implemented to examine the effect of certain parameters, called covariates, on the evolution of the disease. There are many methods to estimate and plot a survival curves as well as to observe a covariate's effect.

Survival analysis methods can be categorized in three main categories concerning the way they implement the survival function. The first group of methods is the non parametric one. These methods are widely used to plot a survival curve. There are no covariates included in such models on the contrary the survival function can be obtained from only time to event observations.

The semi parametric models form the second category of methods in survival analysis. These models make no assumption about the distribution of the survival function but they take advantage of covariates obtained from the patients while they were under observation. These measurements are used in regression procedures and only then can the survival function be plotted.

The third category of methods in survival analysis includes parametric regression models. In these models the form of the survival function is known a priori. Covariates are also examined in such models to formulate their effect on the disease. Using a known survival function, regression methodologies can be applied to check the effect of covariates.

Some well known non parametric methods that have been widely used are the Kaplan Meier estimator, the life table analysis, the Greenwood's formula and the Nelson Aalen's estimator. The Kaplan and Meier estimator (1958) is alternatively called the product limit estimator. Despite censored data the K-M method allows the user to estimate the proportion of subjects in the population whose survival time exceeds a specific time provided that survival is independent of censoring. [12]

The life table method was originally developed by the demographers and actuaries to describe the lifetime of a population. A population life table shows the length of life of a hypothetical group, observed from birth to death, which is assumed to experience the same mortality with the one obtained from the observed population. From the life table analysis it is possible to calculate the expected age of death of an individual of a given age, the probability of surviving from one age to another and other related quantities. [13]

A hazard function can be obtained directly from the survival function in a way that will be proved later. An alternative approach in absence of covariates is the Nelson Aalen estimator [15]. This method is actually an estimator for the cumulative baseline hazard. The method estimates the hazard at each distinct time of death as the ration of the number of deaths to the number exposed to the disease. The cumulative hazard up to a specific time is simply the sum of hazards at all death times up to this specific time. [16]

Another approach for large scale survival data is the *relative survival rate or annual survival ratio.* This method evaluates the survival experience of patients in terms of the general population. Greenwood [18] suggested this approach for to count the effect of cancer treatment. The main idea of the method is that if the average survival time of the patients treated equals that of a random sample of patients of the same age, gender, occupation and other relevant quantities the patients could be considered cured.

The semi parametric models in presence of covariates have a baseline hazard function that is a function common for all patients. The characteristic of this group of models is that the baseline hazard function if cancelled out of the calculations and no form has to be selected a priori. In addition these models perform a regression procedure to count the effect of the covariates on the disease. Such models are the Cox proportional hazards model (1972), the Aalen's additive model (1976), the piecewise exponential model, stratified model, competing risks model, recurrent events model, the models for related observations, the proportional excess hazards and the Cox Aalen model. The additive model makes no assumption about the form of the baseline hazard function. The covariates along with the regression parameters are added to the baseline hazard function.

In the Cox hazards model the covariates and the regression parameter, in an exponential form, are multiplied to the baseline hazard function (The baseline function can be obtained from the Breslow estimator or an extended Kaplan Meier estimator). The form of the baseline hazard is not known and a disadvantage is that it cannot be checked. The simple model requires that the covariates are time independent, but there is an extended Cox model which includes time dependent covariates.

As mentioned above, one characteristic of the Cox proportional hazards model is that the baseline hazard is conditioned out and only the impact of the covariates are estimated by maximizing the partial likelihood. No form of hazard has to be specified which make the Cox model very flexible. On the other hand the parametric models have to specify the functional form of the hazard function. However when the hazard function is of interest it is usually estimated with the Breslow estimator the disadvantage of which is that it lacks the ability to test hypothesis about the shape of the hazard function. The Piecewise Exponential model is a model that is between two extremes. In this model time is divided into several intervals with several different procedures. The hazard in each interval is assumed to be constant but can vary across intervals. It has the flexibility of the Cox model and the ability to statistically check the hazard function that the Cox model lacks due to the Breslow estimator.

The Cox proportional hazards model assumes that the ratio of the hazard functions of any two people with different prognostic covariates is a constant, independent of time. This assumption may not always be met in practical situations due to several reasons. To accommodate these cases, Cox's model is generalized by

the concept of stratification. [5] The *stratified proportional hazards* model was proposed by Kalbfleisch and Prentice at 1980 [29]. In this model the data is categorized in strata by a covariate. If there are two strata then there are two baseline hazard functions and two different regression parameters to satisfy each stratum.

All the methods so far deal with a single type of failure time for each study subject. However, there are cases where a failure may be due to different causes. These different causes of failure are considered as competent events, which introduce *competent risks*. This type of failure is handled by the Competing risks models. The proportional hazards model is used again to identify significant prognostic or risk factors when competing risks are present.

The *recurrences events models* are a class of models that deal with the problem that a failure may be recurrences of the same event. For example the failures of an individual may be recurrences of essentially the same event, such as tumor recurrences after surgeries, or may be successive events of entirely different types, such as strokes and heart attacks. When data include recurrent events, regression models such as the proportional hazards model become more mathematically complicated. A number of regression models have been proposed in the literature that all belong to the class of recurrent events models.

In Cox's proportional hazards model an important assumption is that the time events are independent. However, in many practical situations failure times are observed from related individuals. For example in an epidemiological study of heart disease some of the patients may come from the same family and therefore are not independent. These families with multiple participants are usually called clusters. In this case the regression methods that were so far introduced are inappropriate. A new class of models, the models for related observations, was introduced to solve this issue (Andersen 1993, Liang 1995, Klein and Moeschberger 1997). [5]

Another semi parametric model is the *Cox Aalen* model that was proposed by Scheike and Zhang in 2003 [30]. This model is a combination of the multiplicative Cox model and the Aalen's additive model. This is actually a very flexible model. The main idea is that some covariates may affect additively on the risk while some others multiplicatively. The model allows some covariate effects to be additive non parametric and time varying and other covariates to have constant multiplicative effects.

The *proportional excess hazards models* can be categorized to the semi parametric models as well (Sasieni 1996 [31]). It can be used in a clinical trial setting to accommodate the effect of general population mortality in regression analysis. The effects of covariates on the additional mortality are described by a proportional hazards model. The model may also be used to study cohorts identified by a common exposure which it is assumed will increase the risk of specific diseases, but will not affect the majority of causes of mortality.

The third group consists of the parametric regression models. These models are fully parametric in a sense that hazard and survival follow a given (known) distribution. These models are usually extended to include covariates. If an appropriate model can be assumed then the probability of surviving a given time when covariates are incorporated can be assumed. This class of models includes the log logistic regression model, Weibull model, Exponential model, Gamma model, log normal, Accelerated failure time models and the compared Mackeham model.

Parametric models can be used with a single homogeneous population. Alternatively such models can be fitted ti smaller homogeneous sub-groups and confidence intervals for the fitted parameters will give an objective test of the differences in the lifetime distribution between the groups. However, fully parametric models are difficult to apply without knowledge of the precise form of the hazard function. As a result semi parametric models are most common. A parametric approach is based on the assumption that the lifetime distribution (F (t) = P (T<=t)) belongs to a family of well known parametric distributions.

The exponential regression model assumes that the survival function derives from the exponential cumulative distribution function. The cumulative function is $F(t) = 1 - \exp(-\lambda t)$ and therefore the survival function is $S(t) = 1 - F(t) = \exp(-\lambda t)$. Such a model could be used to reflect the hazard for an individual who remains in good health, where the level of hazard would reflect the risk of death from unnatural causes. Therefore the hazard function is constant $h(t) = \lambda$ and extended to include the covariates for the regression model the hazard function becomes $h(t, z) = \lambda \exp(\beta^{\mathrm{T}} z)$. Here β is a vector of regression and z a vector of covariates.

The Gompertz - Makeham regression models is a fully parametric model too. This regression model could be used over longer time periods to reflect human mortality where the level of hazard increases as age increases. The hazard function in this model follows the following distribution: $h(t) = A + BC^t$.

In addition the Weibull regression model could be used to reflect the hazard for patients recovering from major surgery where the level of hazard is expected to fall as the duration since surgery increases. The survival function follows a Weibull distribution when this model is applied. The probability density function, cumulative density function, survival function and hazard function are respectively: $f(t) = \lambda\gamma(\lambda t)^{\gamma-1}\exp(-(\lambda t)^{\gamma})$, $F(t) = 1 - \exp(-(\lambda t)^{\gamma})$, $S(t) = \exp(-(\lambda t)^{\gamma})$, $h(t) = \lambda\gamma(\lambda t)^{\gamma-1}$. When γ<1 the hazard function is monotonically decreasing, when γ>1 it monotonically increases and when γ=1 the hazard is constant.. In addition the hazard function in this model is extended to include the covariates in an exponential form $h(t,z) = \lambda\gamma(\lambda t)^{\gamma-1}\exp(\beta^{\mathrm{T}} z)$. Here, β is the regression parameter obtained from Weibull regression and z is a vector of covariates.

The log logistic distribution is also used to create the log logistic regression model. This model could be used to reflect the hazard for patients with a disease most likely to cause death in the early stages where the level of hazard increases as the initial condition becomes more severe but then decreases once patients have survived a period of high risk. The probability density function, cumulative function, hazard function and survival function are respectively: $f(t) = \frac{\alpha\gamma t^{\gamma-1}}{(1+\alpha t^{\gamma})^2}$, $F(t) = \frac{\alpha\gamma t^{\gamma}}{1+\alpha t^{\gamma}}$, $h(t) = \frac{\alpha\gamma t^{\gamma-1}}{1+\alpha t^{\gamma}}$, $S(t) = \frac{1}{1+a\cdot t^{\gamma}}$. When γ<1 the hazard function is monotone decreasing from ∞, when γ=1 it decreases from α and for γ>1 the hazard function is non monotonic, but it increases from 0 to a maximum. For the purposes of the log logistic regression model, the survival function can be extended to include the covariates in an exponential form $S(t,z) = \frac{1}{1+a\cdot t^{\gamma}\cdot\exp(z^T\beta)}$ where β is a vector of regression and z a vector consisting of covariates.

The log normal distribution is used in the log normal regression model which is also a fully parametric model. In its simplest form the log normal distribution can be

defined as the distribution of a variable whose logarithm follows a normal distribution. For λ, σ>0 and t>0 the probability density function is $f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp[\frac{-1}{2\sigma^2}(\log(\lambda t))^2]$. The survival function is then $S(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_t^{\infty} \frac{1}{x} \exp[\frac{-1}{2\sigma^2}(\log(\lambda t))^2] dx$. The hazard function is defined as: $h(t) = \frac{f(t)}{S(t)}$. This functions is non monotonic. It begins at h(0)=0 then increases to a maximum and decreases with $\lim_{t\to\infty} h(t) = 0$.

The gamma distribution is also used in parametric models, to satisfy the gamma regression model. The two parameter gamma distribution has a density function that is $f(t) = \frac{\lambda(\lambda t)^{k-1} \exp(-\lambda t)}{\Gamma(k)}$ where Γ (k) is the gamma function and k, λ>0. This distribution like the Weibull one is a generalization of the exponential model, to which it reduces for k=1. The gamma survival function is $S(t) = 1 - \int_0^{\lambda t} u^{k-1} \exp(-u) du / \Gamma(k)$. Therefore the survival function involves an incomplete gamma integral. The gamma regression model assumes a gamma survival function. The hazard function decreases monotonically for k<1, increases monotonically for k>1 and for k=1 hazard is constant.

The class of accelerated failure time models also belongs to the parametric models. The main idea is that the probability that a patient with covariates z will be alive at time t is the same as the probability that a reference subject will be alive at time $t\exp(\beta^{\mathrm{T}} z)$. Here β is the regression parameter and z is a covariates vector. The class of AFT models is a general class of log linear models which assume that the survival time of an individual with covariate z has the same distribution as $T_x = \frac{T_0}{g(z)}$ given that T0 is the survival time under z=0 (baseline) and has a specific known parametric distribution. The function g (z) can have any form provided that g (z)>= 0 and g (0) =1. The functions describing the distribution of failure time in any model of the class given that the covariates act according to $T_x = \frac{T_0}{g(z)}$, satisfy the

following relationships: $S(t,z) = S_0(t \cdot g(z))$, $f(t,z) = g(z) f_0(t \cdot g(z))$, $h(t,z) = g(z) h_0(t \cdot g(z))$.

If the most significant subset of covariates is known, the selection of an appropriate parametric model can be made. The actual lifetime distribution will depend on the assumed form o the hazard rate function. This way the regression problem is reduced to estimating the parameters of the distribution from the observed data. Parametric models can be used with a single homogeneous population. Alternatively, such models can be fitted to smaller homogeneous sub- groups and confidence intervals for the fitted parameters will give an objective test of the differences in the lifetime distribution between the groups. However, fully parametric models are difficult to apply without knowledge of the precise form of the hazard function. As a result, in order to examine covariate's effects semi parametric approaches are more common.

The concept of neural network came up as early as the middle of this century. A neural network is an information processing algorithm, inspired from biological nervous systems. Neural networks, with their remarkable ability to obtain meaning from complicated data can be used to extract patterns and detect trends that are too complex to be noticed by other computer techniques. This fact made neural networks a powerful tool in statistical survival analysis.

Neural networks have been used in various ways for the purposes of analyzing the survival of a group of people. They are not applicable in medicine only but also in engineering, management, sociology and so on.

A restriction of the neural networks is that the dataset used for training has to be sufficient for the network to learn it. In the cases where datasets are too small, neural networks were used again to enlarge them. Resample methods have been used to the original dataset to produce new data with mean and covariance as close as possible to the original dataset. [32]

Another use of the neural networks is the direct classification. Here survival is considered for a fixed time period and the problem is reduced to be binary classification. Censored observations are removed from these models and due to this limitation biases often appear. There is a threshold value at 50% with survival above it indicating that the subject is likely to survive the period while outputs under 50% indicate death. The disadvantage of this network is that no hazard functions can be

plotted for every individual. In addition this kind of neural network does not deal with the censoring problem.

Multiple neural networks were used by Ohno- Machado (1996) to solve the problem of survival analysis. The number of the outputs equals the number of distinct time spots. Therefore each network's output predicts the survival at a specific time point. Censored observations are included until their time of censoring. Therefore, the number of training inputs decreases and this fact makes predictions less reliable. In addition, non monotonic survival curves may result which contradicts with the survival curve definitions that survival curves must be monotonically decreasing. Non monotonic survival curves might mean that the probability of a person to survive two periods might be greater than the probability to survive one period since independencies are not taken into account properly. Furthermore, the problem of combining the multiple neural networks is still under investigation. [34]

A multi layer feed forward network was used be Ravdin and Clark.(1992) This network had one output representing he survival status. A time indicator is used to record the time periods for which a prediction is to be made. A survival status indicator is also used and added to each record. To handle the problem of decreasing training inputs, each uncensored input is replicated for every distinct time unit. A censoring observation is replicated t times, where t is the time of censoring. The survival status is the target of the network and is set to zero when the subject is alive and 1 otherwise. The output of the network is very close to the Kaplan Meier estimator. Again the survival curves are not always monotonically decreasing and the same problem is met. Another disadvantage is that the replication of records leads to large biases as discussed earlier and that this network eventually leads to large datasets with severe scalability problems. [35]

A variation of the approach of Ravdin and Clark was proposed by Biganzoli (1998) at al. The network that is trained is the same, one output and a time indicator. The difference is that uncensored observations are only replicated for the time intervals at which they were actually observed. Thus, subjects that have died are not included after the time interval of death. The network produces discrete hazard rates that are translated to monotonic survival. The disadvantage is that this network is not scalable due to the enormous data replications. [36]

A multi layer strategy to obtain the survival times from the censored cases was proposed by Lapuerta et al.(1995) For each time period a different network is constructed. These networks are trained using only the observations for which the survival status for the corresponding time period is known. Therefore, the networks predict the outcome for the censoring cases.[37] The uncensored cases are used to train a principal neural network which depicts the probability of survival for each time period considered. There are still no guarantees that the survival curve will be monotonic decreasing that is against the survival curve definitions. Since one needs to train as many neural networks as periods considered this approach is not useful for large scale applications. [33]

Faraggi (1995) proposed a neural network close to the Cox proportional hazards model ($h(t,z) = h_0(t)\exp(\beta^{\mathrm{T}} z)$) that replaces the term $\beta^{\mathrm{T}} z$ with the output $g(z,\theta)$ of a network. This network has only one logistic hidden layer and one linear output layer. The θ parameters are estimated with the partial likelihood principle (Newton - Raphson).[38]

Street (1998) used a multilayer perceptron with as many outputs as the distinct observation times. A hyperbolic tangent activation function is used in the output layer such that all output neurons take values between -1 and 1. The first output neuron having a value lower than zero is considered to be the output neuron that predicts the event time. If all output neurons have values greater than one the patient is considered to survive the entire period of study. [39]

A variation of Street's method was proposed by Mani (1999). Again for every observation in the dataset n output units are computed, where n is the number of distinct times. These output units now represent the rate instead of survival probabilities that were used in Street's approach. [40]

Brown (1997) also suggested a single neural network with multiple inputs to predict hazard rate. For the uncensored observations the network's output is set to 0 as long as the subject is alive and 1 when he experiences the event under examination. For the time intervals following the event the hazard is unconstrained. The output values for the censored observations are set to zero until the censoring time. They are also unconstrained for all subsequent intervals. This approach is scalable and results in monotonic survival curves, but there is no extension to deal with time varying inputs. [41]

1.3 Models implemented and thesis structure

This thesis focuses on survival analysis for patients suffering from the ocular melanoma disease. Three models have been implemented for this purpose, each one belonging to a different category. The Kaplan Meier non parametric method has been used to plot survival curves for the whole population. Important conclusions can be made through such a curve since it presents the probability if surviving at each time point.

A semi parametric method is implemented as well. The Cox proportional hazards model for time independent covariates is used to check the effect of certain covariates on the evolution of the disease. The baseline hazard function is approximated in two ways. The Breslow [21] estimator and an extension of the Kaplan Meier estimator to include the covariates are applied for the baseline function to derive. Through a maximization of the partial likelihood the regression parameter derives. This parameter shows the effectiveness of each covariate. A positive value proves a direct effect on the disease. Therefore, the bigger the positive value the larger the effect. A negative value, proves a negative effect in a way that the larger the value the less the effect on the disease. A value close to zero shows that a specific covariate has not effect on the disease.

The Cox proportional hazards model is also used to obtain survival and hazard functions. These hazard functions are different for every patient. The model does not offer directly survival or hazard function for the whole population although there are methodologies to plot such curves. Therefore, the bigger the pick of the hazard function the greater the hazard for this patient. The lower the minimum value of the survival curve the lower the survival probability (and the greater the hazard).

A fully parametric model is implemented as well. A test proved that the log logistic distribution fits the dataset acceptably. Therefore, the log logistic regression model is implemented. The survival and hazard functions are known to follow a log logistic distribution ($h(t)=\frac{\alpha\gamma t^{\gamma-1}}{1+\alpha t^{\gamma}}$, $S(t)=\frac{1}{1+a\cdot t^{\gamma}}$). A linear expression of time versus survival is used for the regression. As it will be shown later, this method cannot be applied for the specific dataset.

Neural networks are also used for the survival analysis of patients who suffer from the ocular melanoma disease. The neural networks are an approximation of the Cox proportional hazards model. The Breslow [21] estimator results are the target of a baseline neural network. The back propagation network is trained to learn the Breslow estimator. In addition, a covariates network is used to perform the regression. The original dataset is divided into three sets for training, testing and validation respectively. The network is trained to learn the exponential part of the Cox model in other words it performs regression.

Chapter 1 is a brief introduction of the ocular melanoma disease as well as an overview of the existing non parametric, semi parametric and parametric models. In chapter two one can find the theoretical background of the models under examination of this thesis. In addition, the ideas of censoring and covariate selection are introduced there. Chapter 3 is the models application on the ocular melanoma dataset. The models used are analyzed and covariates and censoring become clearer for the specific dataset. The chapter is divided by the models categories into non parametric (Kaplan Meier and log rank test), semi parametric (Cox model) and fully parametric (Log logistic regression). In chapter 4 a theoretical background on neural networks is available. Their history, applications and network types. In the same chapter two neural networks are introduced for the Cox proportional hazards model for the specific dataset. Chapter 5 consists of all the results from the models under examination as well as the neural network's results. Finally, in chapter 6 one can find several comparisons among the models of this study and with other relevant studies as well. There are two Appendixes in this thesis. Appendix A provides all programming codes used in environments such as MATLAB, R language and SPSS. In Appendix B there are additional results that might make clearer the models examined.

Chapter 2
Theoretical Background

Chapter 2 extends the theoretical background of the models that are examined in this study. The non parametric Kaplan Meier, the semi parametric Cox model and the fully parametric Log logistic regression are analyzed. In addition, censoring and covariate selection are explained theoretically.

2.1 Survival function and Hazard function

Survival time data measure the time to certain events such as failure (death). These times are subject to random variations from a distribution. This distribution is generally characterized by three functions: 1) The survivorship function 2) the hazard function and 3) the probability density function. All three functions are mathematically equivalent which means that having one of them the other two can be obtained mathematically. In practice the three functions can be used to illustrate aspects of the data. We denote T a non negative random variable which represents the failure time of an individual in a certain population.

Survivorship function. This function is denoted by S (t) and is defined as the probability that an individual survives longer than t. In other words the survival function gives the probability of being alive at duration t. Naturally, S (0) =1 and S(t)=0 as $t \longrightarrow \infty$. S (t) = P (an individual survives longer than t) = P (T>t).

Or S (t) = 1-P (an individual fails before T) = 1- F (t). (Definition 1)

Since P (T<t) is defined as the cumulative function.

$$S(t) = \frac{Number\ \ of\ \ patients\ \ surviving\ \ longer\ \ than\ \ t}{Total\ \ number\ \ of\ \ patients} \qquad \text{(Definition 2)}$$

Definition 1 is not the survival function but an estimate if its values. It is of great importance to mention that this estimator can only be used in data without censoring.

Probability of death density function. Like any other continuous random variable the survival time T has a probability density function that is defined as the limit of the probability of failure in a small interval per unit time.

$$f(t) = \frac{\lim_{\Delta t \to 0} P(An\ \ individual\ \ fails\ \ in\ \ the\ \ Interval\ \ (t, t+\Delta t))}{\Delta t}$$

In practice if there are no censored observations, the probability density function is estimated as the proportion of patients dying in an interval per unit width.

$$f(t) = \frac{number\ \ of\ \ patients\ \ dying\ \ in\ \ the\ \ Interval\ \ beginning\ \ at\ \ time\ \ t}{(total\ \ number\ \ of\ \ patients) \times (\text{int}\, erval\ \ width)}$$

(Definition 3)

Definition 2 is also an estimator and cannot be used for censored data. The cumulative density function is: $F(t) = \int_0^t f(x)dx$

Hazard function. An alternative formation of the distribution of T is given by the hazard function. In literature the hazard function is also called the force of mortality, the mortality intensity function or the failure rate. This function shows the probability that an individual will experience an event within a time interval given that he has survived up to the beginning of the interval.

$$h(t) = \frac{\lim_{\Delta t \to 0} P(An\ \ individual\ \ fails\ \ in\ \ the\ \ time\ \ Interval(t, t+\Delta t)\ \ /\ \ The\ \ individual\ \ has\ \ survived\ \ to\ \ t)}{\Delta t}$$

(Definition 4)

The cumulative hazard function is: $H(t) = \int_0^t h(x)dx$

Hazard function is defined in terms of the cumulative function and the probability function as:

$$h(t) = \frac{f(t)}{1 - F(t)} \text{ (Definition 5)}$$

Relationships of the Survival Functions

The three functions that were defined in the previous section are mathematically equivalent. Given any one of them the other two can be derived.

a. From definitions 1 and 5: S (t) = 1- F (t) and $h(t) = \frac{f(t)}{1-F(t)}$ it can be derived that $h(t) = \frac{f(t)}{S(t)}$ (Definition 6)

b. Since the probability density function is the derivative of the cumulative density function then $f(t) = \frac{d[1-S(t)]}{dt} = -S'(t)$

c. Integrating from 0 to t and using S (0)=1 we have

$$-\int_0^t h(x)dx = \log S(t)$$

$$H(t) = -\log S(t) \quad \Rightarrow \quad S(t) = \exp[-H(t)]$$

$$S(t) = \exp[-\int_0^t h(x)dx]$$

(Definition 7)

- From definitions 6 and 7: $f(t) = h(t)\exp[-H(t)]$

2.2 Models Overview

2.2.1 *Non parametric*: Product limit estimates of survivorship function (Kaplan Meier)

The survival function is the function that is most widely used. This chapter introduces the Kaplan Meier estimator, a method of estimating the survivorship function. With the increased availability of computers this method is applicable to small, moderate and large samples. However, if the sample size is vary large (thousands) the life table analysis might be more useful for the reasons explained on chapter 1.

The Kaplan Meier estimate is a simple way of computing the survival curve. It involves computing the number of people who died at a certain time point, divided by the number of people who are still at risk in the study. The reason why this estimator is called product limit is because these probabilities are multiplied by any earlier computed probabilities. The formula of the estimator obeys the 2nd definition of the survival analysis: $S(t) = \frac{Number\ \ of\ \ patients\ \ surviving\ \ longer\ \ than\ \ t}{Total\ \ number\ \ of\ \ patients}$

The Kaplan Meier survival curve is often illustrated as a staircase with steps downward at the time of death of each individual subject. To compute a survival curve first a notation of the time of occurrence of events is needed. It is possible for two or more events to occur at the same time. The failure times should be sorted from smallest to largest.

The basic computations for the Kaplan Meier survival curve rely on the computation of conditional survival probabilities. In particular, the probability $P[T \geq t_i \mid T \geq t_{i-1}]$. This probability can be interpreted as the probability of surviving to a specific time given the survival to the previous time. A more important probability is the unconditional probability of survival: $P[T \geq t_i] = 1 - F(t_i)$. This represents the simple probability of survival to a specific time.

A relationship between the unconditional and the conditional probability is: $P[T \geq t_i] = P[T \geq t_i \mid T \geq t_{i-1}] \times P[T \geq t_{i-1}]$. Applying this rule again we get:

$$P[T \geq t_i] = P[T \geq t_i \mid T \geq t_{i-1}] \times P[T \geq t_{i-1} \mid T \geq t_{i-2}]$$
$$\times P[T \geq t_{i-2} \mid T \geq t_{i-3}] \times P[T \geq t_{i-3} \mid T \geq t_{i-4}] \times .... \times P[t \geq t_0]$$

The last probability presents the probability of surviving at the beginning of the study. This probability has to be one. Therefore, the unconditional probability equals the cumulative product of conditional probabilities.

The conditional probability of survival can be formed to be: $P[T \geq t_i \mid T \geq t_{i-1}] = 1 - \frac{d_i}{n_i}$

where d_i is the number of deaths at each specific time and n_i is the number of patients at risk at the same time. The unconditional probability of the survival times is simply the cumulative product of the conditional probabilities: [9]

$$P[T \geq t_i] = \prod_{j=1}^{i} 1 - \frac{d_j}{n_j}$$

$$S(t) = \prod_{i=1}^{N} (1 - \frac{d(i)}{n(i)})$$

2.2.2 *Semi parametric*: Cox proportional Hazards model

Survival analysis takes the survival times of a group of patients and produces a survival curve, which shows how many of the patients remain alive over time. Survival time is usually defined as the length of the interval between diagnosis and death, although other events can be examined.

We often want to know whether the survival is influenced by one or more factors, called "covariates" which may be categorical or continuous. For simple situations involving a single factor with just two values there are methods for comparing the survival curves for the two groups of patients. But for more complicated situations a special kind of regression is needed that investigates the effect of each predictor on the shape of the survival curve.

In order to understand the method of proportional hazards, one must first consider a baseline survival curve. This is a survival curve of a hypothetically average patient, with variables that are close to their mean values. This baseline survival curve may have any shape as long as it starts at 1 at time 0 and is monotonically decreasing.

The baseline survival curve is then systematically 'flexed' up or down by each of the predictor variables, while still keeping its general shape. The proportional hazards method computes a coefficient for each predictor variable that indicates the direction and degree of flexing for each predictor on the survival curve. Zero means that a variable has no effect on the curve, and therefore it is not a predictor. A positive

variable means that larger values of the variable are associated with higher mortality. Negative value means that the covariate affects the disease negatively (The greater the value the less the affect on disease). Knowing these coefficients, a customized survival curve for any particular combination of predictor values can be constructed. [11]

Survival analysis examines and models the time it takes for an event to occur. The most important event is death, but the scope of application of survival analysis is broader. Survival analysis focuses on the distribution of survival times. Although there are well known methods for estimating unconditional survival distributions, most interesting survival modeling examines the relationship between survival and one or more predictors, usually named "covariates" in the survival literature. The subject of this chapter is the Cox proportional hazards model, a broadly applicable tool in survival analysis.

The Cox regression model was first introduced in 1972 [24] and it is now well recognized in the analysis of survival data. The model is widely used for exploring the relationship between risk and a set of explanatory variables. These variables can either describe treatment or prognostic factors taken from clinical trials.

The covariates are assumed constant over time, as is typical with treatment, sex and age. In the case of the Cox's model the hazard depends on both time and covariates. This dependence is though provided through two separate factors: The first is h_0 (t) which is a function of time only and is assumed to be the same for all patients. The second is a quantity that depends on covariates through the vector β, which is the vector of regression. Cox suggested that the hazard function has the following form:

$$h(t, z_i) = h_0(t) \cdot \exp(z_i^T \beta) \quad (1)$$

The hazard function is a measure of the potential for the event to occur at a particular time *t*, given that the event did not yet occur. Larger values of the hazard function indicate greater potential for the event to occur. In Eq 1 h_0 (t) is the underlying hazard function, which is a function of time in the case where covariates are not considered. The baseline hazard function measures this potential independently of the covariates. The shape of the hazard function over time is defined by the baseline hazard, for all cases. The baseline hazard function can be approximated from the Breslow [21] estimator or the extended Kaplan Meier [12] estimator. The value of the hazard is equal to the product of the baseline hazard and a covariate effect. While the baseline hazard is dependent upon time, the covariate effect is the same for all time points. Thus, the ratio

of the hazards for any two cases at any time period is the ratio of their covariate effects. This is the proportional hazards assumption. In Eq. 1 the covariates simply help to determine the overall magnitude of the function. $z_i{}^T = (x_{1i}\ \ x_{2i}\ \ x_{3i}\ldots\ x_{ki})$ where i indicates the patient and k is the number of covariates under examination. $\beta^T = (\beta_1\ \ \beta_2\ \ \beta_3\ \ \ldots\ \ \beta_\kappa)$ is the k x 1 vector of regression. These coefficients indicate the magnitude of the effects of their corresponding covariates.

Maximum Likelihood Estimation (MLE) is generally used in survival estimation. MLE produces estimators that are consistent, asymptotically efficient and asymptotically normal. Cox (1972) proposed a partial likelihood function that depends only on the unknown parameter β. The maximization of this partial likelihood offers a rough estimate of the regression parameter. The interpretation of β's sign is:

- If $\beta=0$ the covariate has no effect on survival
- If $\beta<0$ the covariate affects the survival inversely. This means that the higher the value of the examined covariate the lower the hazard.
- If $\beta>0$ the covariate affects the survival. A high value of the covariate would mean high hazard.

The Cox model is not a fully parametric model since it does not specify the form of the baseline hazard function. It does however specify the hazard ratio of any two individuals and for this reason it is defined as a semi parametric model. Therefore the hazard ratio of two different patients with covariate vectors z_1 and z_2 is:

$$\frac{h(t,z_1)}{h(t,z_2)} = \frac{h_0(t)\exp(z_1^T\beta)}{h_0(t)\exp(z_2^T\beta)} \quad \Rightarrow \quad \frac{h(t,z_1)}{h(t,z_2)} = \frac{\exp(z_1^T\beta)}{\exp(z_2^T\beta)} \quad \Rightarrow \quad \frac{h(t,z_1)}{h(t,z_2)} = \exp[(z_1^T - z_2^T)\beta] \qquad (2)$$

Given that this ratio does not depend on time, so that it is constant then the failure rates of any two individuals are proportional. Due to the equation above the Cox's regression model is called a proportional hazards model. The equation 2 can have a logarithmic form like this:

$$\frac{h(t,z_1)}{h(t,z_2)} = \exp[(z_1^T - z_2^T)\beta] \quad \Rightarrow \quad \ln[\frac{h(t,z_1)}{h(t,z_2)}] = \ln[\exp[(z_1^T - z_2^T)\beta]] \quad \Rightarrow$$

$$\Rightarrow \ln[h(t,z_1)] - \ln[h(t,z_2)] = (z_1^T - z_2^T)\beta \qquad (3)$$

The assumption in this case is that there is a constant difference between the logarithmic hazard functions of two different individuals.

Moreover, the Cox proportional model makes one more assumption. This one has to do with the exponential link of the covariates to the hazard. The assumption is that two completely different covariates may affect hazard in a multiplicative way, like in equation 2 or in an additive way like in equation 3.

An extension of the Cox proportional hazards model allows analyzing time dependent covariates. These covariates appear as functions of time. It is still possible to analyse the effect of time dependent covariates using the simple Cox model if the value of these covariates does not change greatly over the period of the investigation and if the main effect of these covariates depends on the value at a specific point in time.

However, of the variables change rapidly throughout the study the extended Cox model can be used. Then, the Cox model is easily extended to include such variables. The most important change is the vector of covariates which is altered in order to express covariates through time ($h(t, z_i) = h_0(t) \cdot \exp(z_i^T(t)\beta(t))$).

2.2.3 *Fully parametric*: Log logistic regression model

The Cox model is a method that belongs to semi parametric models, since the form of the baseline hazard functions is not known during the regression part. Another set of models is the parametric one. Parametric models perform the regression procedure given that the distribution of the hazard function is known. Therefore, several parametric regression models can be defined. When covariates are considered, we assume that the survival time, or generally a function of it, has an explicit relationship with the covariates. Furthermore, when a parametric model is considered we assume that the survival time follows a given theoretical distribution and has an explicit relationship with the covariates.

The log logistic distribution can be used for the purposes of a parametric model. In this case the survival time is accepted to follow a log logistic distribution a priori. The general distribution of the survival time is: $S(t) = \frac{1}{1 + a \cdot t^{\gamma}}$. This can be extended in the exact same way that the Kaplan Meier estimator was extended in the Cox model to include the covariates. Therefore the distribution becomes: $S(t, z) = \frac{1}{1 + a \cdot t^{\gamma} \cdot \exp(z^T \beta)}$. There are obvious similarities with the Cox proportional hazards model, since the covariates appear again in an exponential form and they act also multiplicatively on the

survival. This distribution can become the original log logistic survival function in the absence of the covariates (β=0). As a consequence the distribution function becomes:

$$F(t,z) = 1 - S(t,z)$$

$$F(t,z) = 1 - \frac{1}{1 + a \cdot t^{\gamma} \cdot \exp(z^T \beta)}$$

$$F(t,z) = \frac{a \cdot t^{\gamma} \cdot \exp(z^T \beta)}{1 + a \cdot t^{\gamma} \cdot \exp(z^T \beta)}$$

With the following transformations the so called odds ($\frac{F(t,z)}{1-F(t,z)}$) can be obtained:

$$F(t,z) = 1 - \frac{1}{1 + a \cdot t^{\gamma} \cdot \exp(z^T \beta)} \Rightarrow F(t,z) = \frac{1 + a \cdot t^{\gamma} \cdot \exp(z^T \beta) - 1}{1 + a \cdot t^{\gamma} \cdot \exp(z^T \beta)} \Rightarrow F(t) = \frac{a \cdot t^{\gamma} \cdot \exp(z^T \beta)}{1 + a \cdot t^{\gamma} \cdot \exp(z^T \beta)} \Rightarrow$$

$$\Rightarrow F(t,z) = a \cdot t^{\gamma} \cdot \exp(z^T \beta) \cdot S(t,z) \Rightarrow \frac{F(t,z)}{S(t,z)} = a \cdot t^{\gamma} \cdot \exp(z^T \beta) \Rightarrow$$

$$\frac{F(t,z)}{1-F(t,z)} = a \cdot t^{\gamma} \cdot \exp(z^T \beta) \qquad (12)$$

Conditional that all covariates are ignored (β=0), Eq.12 reduces to the baseline odds function. The baseline function is multiplied with a factor $\exp(z^T \beta)$ for any individual with covariate z. The characteristics of the log logistic regression model are proportional to the formation of the odds failure. Therefore, the plot of the odds failure versus time is expected to be monotonically increasing and depends only on the parameter γ. Additionally the covariates act multiplicatively on the odds function. This is obvious when regarding any two individuals with covariates z1 and z2. The ratio is expected to be constant.

$$\lambda = \frac{odds(individual1)}{odds(individual2)} = \frac{\frac{F(t,z_1)}{1-F(t,z_1)}}{\frac{F(t,z_2)}{1-F(t,z_2)}} = \frac{a \cdot t^{\gamma} \cdot \exp({z_1}^T \beta)}{a \cdot t^{\gamma} \cdot \exp({z_2}^T \beta)} = \exp[({z_1}^T - {z_2}^T)\beta]$$

An empirical check of the suitability of the log logistic regression model for the analysis of a specific dataset in the presence of covariates can be derived when logarithms are applied to Eq. 12

$$\frac{F(t,z)}{1-F(t,z)} = a \cdot t^{\gamma} \cdot \exp(z^T \beta) \Rightarrow$$

$$\Rightarrow \log[\frac{F(t,z)}{1-F(t,z)}] = \log[a \cdot t^{\gamma} \cdot \exp(z^T \beta)] \Rightarrow$$

$$\Rightarrow \log[\frac{F(t,z)}{1-F(t,z)}] = z^T \beta + \log(\alpha) + \gamma \log(t) \qquad (13)$$

Equation 13 proves that there is a linear relationship between the log odds and the logarithm of time. The slope is γ and there is a clear dependence on $z^T\beta$ which moves the plot on the y axis. For the model to be suitable we expect that a plot of the *non parametric* log odds (that can be log odds derived from the non parametric Kaplan Meier estimator) against the log t is linear for a homogeneous population. In other words patients that appear with a particular covariate pattern (homogeneous) should have almost linear plots. In addition two different patterns for example z1 and z2 should correspond to roughly parallel lines the slope of which gives an estimate of γ. If the two patterns are different in one covariate only say x1 and all other covariates have the same values then the distance between the two lines gives an estimate of β. If the distance is not constant then the proportional odds model is not appropriate. [6]

In the biomedical research we are mostly interested in the hazard functions. The hazard function n the log logistic regression model is:

$$S(t,z)=\exp[-H(t,z)] \quad \Rightarrow \quad H(t,z)=-\log S(t,z) \quad \Rightarrow$$

$$H(t,z)=-\log\frac{1}{1+a\cdot t^{\gamma}\exp(z^T\beta)} \quad \Rightarrow \quad H(t,z)=\log(1+at^{\gamma}\exp(z^T\beta))$$

$$H(t,z)=\int_0^t h(x)dx \qquad \Rightarrow h(t,z)=\frac{dH(t,z)}{dt}=\frac{\alpha\gamma t^{\gamma-1}\exp(z^T\beta)}{1+\alpha t^{\gamma}\exp(z^T\beta)}$$

$$h(t,z)=\frac{\alpha\gamma t^{\gamma-1}\exp(z^T\beta)}{1+\alpha t^{\gamma}\exp(z^T\beta)}$$

Equation 12 could be a member of class models if the covariates didn't appear exponentially but in a variety of ways say g (z). The *class of proportional odds models (PO)* suggests that $\frac{F(t,z)}{1-F(t,z)}=\omega_0(t)\cdot g(z)$ [Eq. 14]. Here g (z) is a function of the covariate vector but not of time. In addition $\omega_0(t)$ is any function of time not dependent on time with $\omega_0(t)>0$.

2.3 Censoring

In medical studies censoring is common and needs to be handled by some techniques. There are two major sources of censoring. In the first case a patient might be lost to follow up. It is known that the patient is alive at the last meeting but his subsequent status is unknown. This might happen if a patient stops coming to the clinic or moves away for undefined reasons. In addition censoring occurs even if patients die from other causes (competing risks) before the closure of the study. Time to last contact will be taken as censoring time. [2]

In the second case where a clinical study is closed after a fixed study period, some of the patients are alive at the time of study closure. Therefore time to study closure will be taken as censoring time. In this case we only know that the time to event is greater than a certain value, the time from entry to the end of the study. There are right censored data. Although the exact time of the outcome event is not known the fact that it does not precede the censoring time is useful information for the survival analysis.

Some other censoring cases come about if the length if the follow up varies due to random entry. Therefore, we can not observe the event for those individuals with insufficient follow up time. Censoring form staggered entry may be different from censoring due to the other reasons mentioned earlier. In addition, left censoring happens when the time of entry of certain individuals is not known and even worse, interval censoring happens when neither the time of entry nor the event time are known.

There is an important assumption in survival analysis that individuals who are censored are at the same risk of subsequent failure as those who are still alive and uncensored. The risk set at any time point (which is defined as the set of individuals who are still alive and uncensored) should be representative of the entire population alive at the same time. Statistically this assumption is equivalent to the one that the censoring process is independent of the survival time. If censoring only occurs because of staggered entries then the assumption of independent censoring is satisfied. However, when censoring results from loss to follow up or death from a competing risk, then this assumption is more suspect.

If the study period is long enough to observe the survival time of all subjects, as in some animal experiments, one may prefer to use more common methods to analyze the survival (t- test, least square regression). However in studies of human subjects there is censoring (due to the reasons mentioned above) and the outcome cannot be analyzed by the usual methods of continuing data. [1]

2.4 Covariate Selection

In survival analysis it is often useful to examine the influence of certain covariates on the probability of the examined event to occur. There are various measurements that can be examined in such models. Survival analysis for travel habits (sociology) could examine the effects of the weather, roads, location and so on. Survival analysis of newspapers subscribers (management) could examine the effects of social status, marriage, job description and so on.

In the medical field survival analysis focuses on predicting the survivability given several measurements. From the moment a disease is diagnosed and the patient enters the study, several measurements are taken that are related to the disease. For example, in the case where tumors are under examination, an appropriate measurement could be the location of the tumor and its diameter.

In addition, several kinds of treatments could be examined to figure out the most effective one. In this case, the event under examination is the recovery and not death. There are various studies that perform survival analysis on patients suffering from a disease and certain treatments are applied on them. For example, in the case of tumors, radiotherapy, chemotherapy, drugs, placebos and surgeries are some of the treatment methodologies that are investigated. Survival analysis answers the issue of selecting the more appropriate treatment.

These measurements are called covariates. Several models in survival analysis include covariates in the estimations. The semi parametric and fully parametric models are able to check the influence of a certain covariate on the disease. The covariates can either be constant or time dependent and all models can be extended to include time dependent covariates. A covariate is any quantity recorded in respect of each life under observation, which is likely to affect the future lifetime distribution. Such covariates can be:

1. A direct quantitative measure which is numerical (for example age, tumour diameter, weight)
2. An indicator which is usually binary (e.g. 0 for male 1 for female, 0 for non metastatic death, 1 for metastatic death)
3. A quantitative interpretation of a qualitative measure. This covariate is numerical but it expresses quality (e.g. severity of symptoms ranging from 0 to 10, with 0 representing no symptoms and 10 representing extremely severe symptoms).

Chapter 3

Models application on specific dataset

In the beginning of this chapter there is an extended analysis of the specific dataset under examination, the ocular melanoma dataset. Then, this chapter analyzes the way that the models introduced in previous chapters are applied on the specific ocular melanoma dataset. In addition, for the purposes of the Kaplan Meier, censoring on the specific dataset is defined and the log rank test is introduced analytically. For the purposes of the Cox proportional hazards model covariate selection is also defined.

3.1 Overview of the ocular melanoma dataset

The data for this study come from patients who suffer from the ocular melanoma disease. The data were actually collected from patients who suffer from one of the four kinds of ocular melanoma that where mentioned earlier in chapter 1. There were 1735 patients who were studied from the moment the disease was diagnosed. From this collection we have information about the following measurements: Date of birth, Date of death (if they died), Submission date which is the date when the disease was diagnosed, management date, an indicator of metastatic death, the survival time in years and sex. Furthermore there is information on certain characteristics of the tumour which were obtained after medical observation. These are: anterior tumour margin, ultrasound height, longest ultrasound basal dimension and epithelioid cellularity.

The date of birth was obtained from the hospital records while the date of death from death certificates, the family, a medical practitioner or any other valid source. The management day was the day of primary treatment of intraocular tumour. The death is a binary variable obtained from any valid source. Logical one indicates a dead patient while logical 0 indicates a patient still alive. The sex is a binary variable as well with one indicating male and zero indicating a female. Another binary variable is the epithelioid cellularity. Zero indicates that there were no epithelioid melanoma cells and one the contrary.

Another measurement is concerned with the tumour height. The ultrasound height is measured by ultrasonography. Distance is measured from internal scleral surface to tumour apex, taking care to measure thicknest part of tumour, excluding overlying retinal detachment and avoiding oblique cuts. The MD is an indicator of the existence or not of a metastatic

death. One indicates that the death was provoked from metastasis, while zero indicates the opposite. Obviously patients who are still considered to be alive have a value of MD equal to zero.

A measurement that deals with the ultrasound height is the LUDB which stand for the 'Longest Ultrasound Basal Dimension '. This defines the longest basal tumour dimension as measured by ultrasonography. Care must be taken to avoid over-estimating this measurement because of retinal detachment or underestimating it because of tapering tumour margins.

The anterior tumour margin is estimated by ophthalmoscopy, slit-lamp examination and ultrasonography. This variable is called ANT MAR, Anterior Tumour Margin. The equator is located 15.6 mm from the fovea in an emmetropic eye and corresponds to a point just anterior to the vortex vein ampulla. This is a categorical variable with 14 indicators of the location of the tumour.

Value	**Location**
1	Disc and Fovea
2	Fovea
3	=<1DD Fovea
4	1-2DD Fovea
5	Disc
6	=<1DD Disc
7	1-2DD Disc
8	Posterior Choroid
9	Anterior Choroid
10	Pars plana
11	Pars plicata
12	Iris
13	Angle
14	Cornea

The survival time is a variable measured for each patient in years. It counts the date from the first ocular melanoma treatment to the date of death. In the case where a patient is still alive (when dead=0) then the survival time is counted from the date of birth to the end of study (10th February 2005). This variable is going to be used as the time variable in all of our plots.

Database ID	Submission Date	Last Update Date	DOB	antmar
7690	21/3/2005 16:43	21/3/2005 16:43	25-Feb-1933	9
7691	21/3/2005 16:43	21/3/2005 16:43	17-Sep-1907	9

Mand	DOD	*Lubd*	*Uh*	*epi*
19-Jan-2005	missing	13.60	6	1.00
19-Mar-1993	missing	10.00	5	0.00

Md	death	survy	Sex	Rand10
0.00	0.00	0.06	0.00	1
0.00	0.00	11.90	0.00	1

3.2 *Non Parametric*: The product limit estimator (Kaplan Meier)

Kaplan and Meier [12] proposed an estimator that computes survival probabilities. When these probabilities are plotted versus the survival time, very important conclusions can be extracted concerning the probability of survival of a patient by comparing it with the experience of other patients.

The model was introduced in Chapter 2. The Product limit estimator is an estimator for survival functions which is implemented in order to analyze the survivability of patients suffering from the ocular melanoma disease. This specific dataset has also censoring cases. These cases are patients that were cancelled out of the study for various reasons but are still alive at the end of the study.

At this point it is important to examine which cases of this particular study should be considered as censored. After the closure of the study there are some patients who are still alive (those with dead=0). This is obviously a censoring case: "*In the second case where a clinical study is closed after a fixed study period, some of the patients are alive at the time of study closure. Therefore time to study closure will be taken as censoring time*." Thus, whenever the model suggests that a set includes patients at risk, the censored patients will be cancelled out of the calculations.

The Kaplan Meier estimator is: $S(t) = \prod_{i=1}^{N} (1 - \frac{d(i)}{n(i)})$. This estimator produces survival results for each time spot. Here N is the number of time spots, d (i) is the number of deaths at each time spot and n (i) is the risk set at this time.

At the beginning, the risk set has to be defined. The risk set is calculated for each specific time to include patients that have survived up to this time and after it. For this study the "survy" (survival years) numerical variable will be considered as time. Time zero is the time when all patients become subjects to the study (They are considered to enter the study at the same time). Therefore, the risk set at time 8.00 includes all patients that are alive after the 8th year. Taking a subset of our dataset the risk set idea can become clearer:

Patient number	*Survival years*	*Death indicator*	*Risk set (patients)*
1	0.02	1	2,3,4,5,6
2	0.08	1	4,5,6
3	0.08	1	4,5,6
4	0.16	1	5,6
5	0.24	1	6
6	0.28	1	-

Generally a formula to obtain the risk set for each specific time spot is t R (t_j)= {j: $t_j >= t_i$ }. In the ocular melanoma dataset very often two or more patients experience the event at the same time. Therefore attention should be made when considering that moment's dataset.

In the ocular melanoma dataset there are also censoring cases, which can be easily handled with the Kaplan Meier estimator. Patients might be censored because they abandoned the study or because they died due to other (even natural) reasons. The censoring indicator is the binary "death" variable. In order for statisticians to have a clear view over the group that they follow up, the study has to go on until all patients involved die. Consequently, when "death" indicates 0, this doesn't mean that the patient is still alive and under observation. Therefore, death=0 indicates a censoring case.

This fact has a direct effect in the risk set of specific time spot. To be more specific, in order to obtain the risk set at time ti, not only do we have to leave aside the deaths at that time but also the censoring cases. Taking another subset into consideration this becomes clearer:

Patient number	*Survival years*	*Death indicator*	*Risk set (patients)*
1	0,01	0	2,3,4,5,6,7,8,9,10,11,12
2	0,02	1	3,4,5,6,7,8,9,10,11,12
3	0,04	0	4,5,6,7,8,9,10,11,12
4	0,05	0	6,7,8,9,10,11,12
5	0,05	0	6,7,8,9,10,11,12
6	0,07	0	7,8,9,10,11,12
7	0,08	1	8,9,10,11,12
8	0,1	0	10,11,12
9	0,1	0	10,11,12
10	0,15	1	11,12
11	0,16	1	12
12	0,17	1	-

Using these rules the risk set is obtained for each time. The Kaplan Meier estimator can then be easily implemented using $S(t) = \prod_{i=1}^{N} (1 - \frac{d(i)}{n(i)})$

3.2.1 The log rank Test

Often it is useful to compare the survival experience of two or more groups of individuals. In the case of the ocular melanoma dataset such stratification could have some point if we compared the survival curves of patients who have epithelioid cellularity and of those who don't. The epithelioid cellularity could be examined as a comparing factor. Epi is a binary covariate indicating whether or now a patient has it.

In clinical research one is concerned not only with estimating the cumulative probability of surviving but also with the comparison of the experience of groups of subjects differing on a given characteristic or on different treatments. Unlike the reliability analysis in the technological field where the physical properties of the production process can suggest the theoretical failure function, in the biomedical field it is extremely difficult to have a priori knowledge to make hypothesis on the underlying theoretical survival functions. Therefore, the non parametric approach is usually adopted to compare survival curves. Among the various non parametric tests the Mantel Haenzel test (1959) currently called the log rank test will be applied to our dataset. [6]

The log rank test is the most popular method of comparing the survival of groups, which takes the whole follow up period into account. An advantage o this test is that it does not require to know anything about the shape of the survival curve or the distribution of survival times.

The log rank test is used to test the null hypothesis that there is no difference between the populations in the probability of an event at any time point. The analysis is based on the time of events. It is based on the same assumptions as the Kaplan Meier curve, that censoring is unrelated to prognosis and that the survival probabilities are the same for subjects recruited early and late in the study. Deviations from these assumptions are more important if they are satisfied differently in the groups being compared; for example if censoring is more likely in one group than another.

The log rank test is most likely to detect a difference between groups when the risk of an event is consistently higher for one group than another. It is unlikely to detect the difference when survival curves cross as can happen when comparing a medical with a surgical intervention. Since the log rank test is a test of significance it cannot provide an estimate of the size of the difference between the groups investigated. [10]

3.3 *Semi parametric*: The Cox Proportional Hazards model

The Cox proportional hazards model was introduced in chapter 2. It is a semi parametric model that includes the covariates whether they are time dependent or not. This model assumes a baseline hazard function that is not obtained parametrically but derives from an estimator. The general formula of the Cox model hazard is:

$$h(t, z_i) = h_0(t) \cdot \exp(z_i^T \beta)$$

The covariates act multiplicatively on hazard. Z is a vector of covariates and β is the regression parameter. In addition, h_0 (t) is the baseline hazard function, that is a common function for all patients.

For the purposes of the Cox proportional model, three covariates are selected to examine their effect on the evolution of the disease. These covariates are time independent and therefore the proportionality assumption introduced in chapter 2 is met. In other words the variables used are constant over time.

In addition the three variables that are selected have some serious medical interest. Doctors should need the conclusions of this study in order to diagnose the severity of the ocular melanoma given a set of prognostic factors. The covariates selected are: *ultrasound height, longest ultrasound basal dimension* and the indicator of *epithelioid cellularity.* These variables have missing values for certain cases in the dataset. There are techniques to fill in the missing gaps (even with a neural network) nut it was preferred that patients with missing values were extracted from the calculations. This leaves us with a dataset of 743 patients.

The Cox model can examine the effect of time dependent variables as long as they don't change value over the study period. For example, age has a value that changes over time. If the study period is a few months then age can be considered constant. In our case the study period has a year scale and therefore age can not be a variable under examination.

Cox proposed that the parameter β can be estimated by the partial likelihood method. Consider a total of N subjects and suppose that k failures occur (k<N). R (t) is a set that contains all patients who are still alive when the ith patient experiences the event ($R(t_j) = \{j: t_j >= t_i\}$). Conditioned on the fact that one individual is observed to fail at time t_i the probability that he is the patient with covariates z_i is: (from equation 1).

$\frac{h(t_i, z_i)}{\sum_{j \in R(t_i)} h(t_i, z_j)} = \frac{h_0(t)\exp(z_i^T \beta)}{\sum_{j \in R(t_i)} h_0(t)\exp(z_j^T \beta)} = \frac{\exp(z_i^T \beta)}{\sum_{j \in R(t_i)} \exp(z_j^T \beta)}$. Here the baseline function is canceled out.

Assume we have k failures (deaths) then the partial likelihood function over all failure times is: (independent probabilities)

$$L(\beta) = \prod_{i=1}^{k} \frac{\exp(z_i^T \beta)}{\sum_{j \in R(t_i)} \exp(z_j^T \beta)} \quad (4)$$

Equation 4 indicates that L (β) depends on the unknown values of the regression parameter and the known vector of covariates. Apparently, there is no contribution to the estimate of β from the gaps between successive failure times due to the absence of the baseline hazard function. In other words, since this partial likelihood involves patient's covariates it makes sense only for specific time points and not for the spots between them. The Log partial likelihood is:

$$\log L(\beta) = \log \left[\prod_{j=1}^{k} \frac{\exp(z_j \cdot \beta)}{\sum_{i \in R(t_j)} \exp(z_t \cdot \beta)} \right]$$

$$l(\beta) = \sum_{j=1}^{k} \left(\log \left(\frac{\exp(z_j \cdot \beta)}{\sum_{i \in R(t_j)} \exp(z_t \cdot \beta)} \right) \right)$$

$$l(\beta) = \sum_{j=1}^{k} \left(\log(\exp(z_j \cdot \beta)) - \log\left(\sum_{i \in R(t_j)} \exp(z_t \cdot \beta) \right) \right)$$

$$l(\beta) = \sum_{j=1}^{k} \left((z_j \cdot \beta) - \log\left(\sum_{i \in R(t_j)} \exp(z_t \cdot \beta) \right) \right)$$

The maximum partial likelihood estimation of β is the solution of the equation:

$$\frac{\partial\, l(\beta)}{\partial \beta} = 0$$

$$\left[\sum_{j=1}^{k}\left((z_j \cdot \beta) - \log(\sum_{i \in R(t_j)} \exp(z_t \cdot \beta))\right)\right]' = 0$$

$$\sum_{j=1}^{k}\left((z_j \cdot \beta)' - \left[\log(\sum_{i \in R(t_j)} \exp(z_t \cdot \beta))\right]'\right) = 0$$

$$\sum_{j=1}^{k}\left(z_j - \frac{1}{\sum_{i \in R(t_j)} \exp(z_t \cdot \beta)} \cdot \left[\sum_{i \in R(t_j)} \exp(z_t \cdot \beta)\right]'\right) = 0$$

(5)

From this log likelihood, Cox showed that a valid value of β can be estimated. One of the nonlinear algorithms to compute this maximization is the **Newton-Raphson** iteration. In this algorithm we start with an initial guess of β and iteratively determine β^m with the formula $\beta^{(m)} = U^{-1}(\beta^{(m-1)})s(\beta^{(m-1)})$ where $U(\beta) = -\mathrm{N} \cdot Hessian(\beta)$).

The baseline hazard function can be estimated in various ways. The most common way is to use the Breslow estimator or an extension of the Kaplan Meier estimator. In addition, there are many cases in literature where the baseline function is approximated theoretically or by taking advantage only of patients that have low numerical values in their covariates (This makes sense since theoretically the baseline function is the hazard in absence o covariates).In this study, the Breslow and the extended Kaplan Meier will be used to obtain the baseline hazard function.

Although the values of the regression parameters are derived right form the dataset, the over all survival behaviour can not be understood without knowing the baseline hazard function. One way to understand the baseline hazard function is to specify it.

By assuming that the baseline hazard function is constant between each pair of successive observed failure times, Breslow [21] has proposed the following estimator of baseline cumulative hazard function:

$$H_0(t) = \sum_{t(i) \le t} \frac{d(i)}{\sum_{l \in R(t_i)} \exp(z_l^T \beta)}$$

With the estimated regression coefficient β, the Breslow estimator for the cumulative hazard can be implemented to obtain the values of H_0 (t). This estimator is not enough by itself to produce meaningful hazard or survival curves since there is no meaningful proportion. The number of deaths divided by an exponential sum of covariates and regression parameter can not give by itself an appropriate estimate of hazard or survival. Only after this estimator is used in the Cox model can the graphs make sense.

A very important point is that this estimator is used for every unique survival time but only those when a death happens. Thus, a value of the cumulative hazard function is obtained for each death time. This doesn't mean that censoring cases are cancelled out of the calculations. The censored patients are included in the risk set. The idea of the risk set was analyzed in chapter 2. For any time point a risk set includes patients that are still alive and not censored.

The Breslow estimator is an extension of the Nelson Aalen estimator:

$$H_0(t) = \sum_{t(i) \le t} \frac{d(i)}{n(i)}$$

The Breslow estimator includes the covariates and the regression parameter β. On the contrary the Nelson Aalen estimator takes advantage only of the number of deaths d(i) and the risk set n(i). The two estimators are equivalent in the absence of the covariates, that is when β=0.

Alternatively the Breslow estimator can be transformed to calculate the values of the cumulative survival. This transformation is performed by applying Definition 7 from the 2nd chapter to the Breslow estimator:

$$H_0(t) = \sum_{t(i)\le t} \frac{d(i)}{\sum_{l\in R(t_i)} \exp(z_l{}^T \beta)} \quad (Breslow) \quad , \qquad S_0(t) = \exp[-H_0(t)] \quad (Definition \ 7)$$

$$S_0(t) = \exp[-\sum_{t(i)\le t} \frac{d(i)}{\sum_{l\in R(t_i)} \exp(z_l{}^T \beta)}] \quad \Rightarrow \quad S_0(t) = \prod_{t(i)\le t} \left[\exp[\frac{-d(i)}{\sum_{l\in R(t_i)} \exp(z_l{}^T \beta)}] \right] \quad (6)$$

The Kaplan Meier estimator is also an estimator of a cumulative survival function in the absence of covariates. This survivorship function shows how the probability of surviving changes through time, living aside the existence of any kind of covariates. Thus, there is a theoretical relationship of the Kaplan Meier's estimator for the survival function, with the baseline hazard function.

The formula of the Kaplan Meier estimator is $S(t) = \prod_{i=1}^{N} (1 - \frac{d(i)}{n(i)})$. (Eq. 10)

Here d (i) is the number of events that take place at the same time with the event of the ith patient. n (i) is the number of patients who are still at risk when the ith patient experiences the event. In other words, n (i) indicate the number of patients that have more survival years than the ith patient and are not censored.

Of course we can not cancel out completely the fact there are some covariates that affect the survivorship. The main idea is that the Kaplan Meier survival function can be transformed to include somehow the effects of the covariates. This implementation is completely analogue to the Nelson Aalen and Breslow estimators in chapter 6.1.2. Therefore, an analog of the Kaplan Meier estimator can be derived for S (t) by thinking if h_0 (t) in terms of a discrete hazard having mass $\frac{d(i)}{\sum_{l\in R(t_i)} \exp(z_l{}^T \beta)}$ at each failure time. This could be interpreted as the probability that m individuals with covariates x=0 fail at t_i conditional on the sets of covariates observed individuals at risk at t_i. Therefore the analogue if the Kaplan-Meier estimator is: $S_0(t) = \prod_{t(i)\le t} \left[1 - \frac{d(i)}{\sum_{l\in R(t_i)} \exp(z_l{}^T \beta)} \right]$ (Eq. 11)

This estimator is actually the same with the Kaplan Meier (Eq. 10) for β=0 which theoretically means that no covariates affect the survivorship of an individual. Obviously the extended Kaplan Meier is an estimator and not a function used to obtain separate estimations of the cumulative survival at each death time.

The Breslow estimator for survival $S_0(t) = \prod_{t(i)\le t} \left[\exp[\frac{-d(i)}{\sum_{l\in R(t_i)} \exp(z_l^T \beta)}] \right]$ and the extended Kaplan Meier estimator $S_0(t) = \prod_{t(i)\le t} \left[1 - \frac{d(i)}{\sum_{l\in R(t_i)} \exp(z_l^T \beta)} \right]$ have similar numerical results. There s a connection between these two estimators that derives from a Taylor series of an exponential function: $\exp(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + ... \qquad -\infty < x < +\infty$

3.4 *Fully parametric*: Log logistic regression model

In the beginning it is essential to provide a brief overview of the log logistic distribution since it is widely used both in the Cox model application on the specific dataset (as it fits well the survival function obtained from the Cox model) and the log logistic regression model.

The survival time T has a log logistic distribution if log (T) has a logistic distribution. The two parameter distribution has a hazard function that is:

$$h(t) = \frac{\alpha\gamma t^{\gamma-1}}{1+\alpha t^{\gamma}}$$

This distribution differs form the Weibull hazard for the denominator. The log logistic distribution is characterized by two factors α and γ. The median of the log logistic is $\alpha^{-\frac{1}{\gamma}}$. When γ>1 the log logistic hazard has the value zero at time zero, increases to a peak and then declines. When γ=1 the hazard starts at $\alpha^{\frac{1}{\gamma}}$ and declines monotonically. When γ<1 the hazard starts at infinity and then declines which is similar to the Weibull distribution. Therefore the log logistic distribution can be used to describe a first increasing and then decreasing or a monotonically decreasing hazard.

The probability density function and the survival function of the log logistic distribution are: $f(t)=\frac{\alpha\gamma t^{\gamma-1}}{(1+\alpha t^{\gamma})^2}$, $S(t)=\frac{1}{1+a\cdot t^{\gamma}}$. The cumulative hazard function can derive: $S(t)=\exp[-H(t)] \Rightarrow H(t)=-\log S(t) \Rightarrow H(t)=-\log\frac{1}{1+a\cdot t^{\gamma}} \Rightarrow H(t)=\log(1+at^{\gamma})$

An example taking all possible values of the parameter γ in the hazard and density function is: [5]

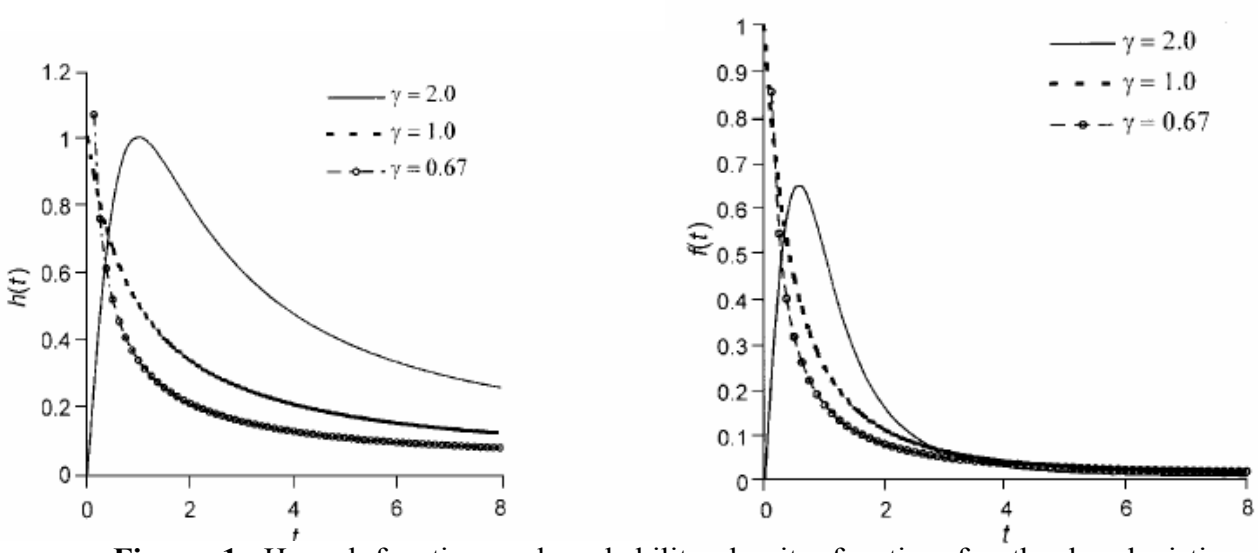


Figure 1. Hazard function and probability density function for the log logistic distribution. [5]

The suitability of the log logistic distribution for the survival analysis of a specific dataset can be empirically checked, using a linear relationship which can be derived from S (t) and F (t) = 1- S (t).

$$F(t)=1-\frac{1}{1+a\cdot t^{\gamma}} \Rightarrow F(t)=\frac{1+a\cdot t^{\gamma}-1}{1+a\cdot t^{\gamma}} \Rightarrow F(t)=\frac{a\cdot t^{\gamma}}{1+a\cdot t^{\gamma}} \Rightarrow$$

$$\Rightarrow F(t)=a\cdot t^{\gamma}\cdot S(t) \Rightarrow \frac{F(t)}{S(t)}=a\cdot t^{\gamma} \Rightarrow \frac{F(t)}{1-F(t)}=a\cdot t^{\gamma} \Rightarrow$$

$$\Rightarrow \log[\frac{F(t)}{1-F(t)}]=\log[a\cdot t^{\gamma}] \Rightarrow$$

$$\Rightarrow \log[\frac{F(t)}{1-F(t)}]=\log(\alpha)+\gamma\log(t)$$

From this relationship the log logistic distribution is a linear model for the log odds of failure over the logarithm of time with slope γ. If a plot of the log odds ($\log[\frac{F(t)}{1-F(t)}]$) versus the logarithm of time appears to be approximately a straight line then the log logistic distribution is suitable to the dataset. The log odds can be calculated through the Kaplan Meier estimator, $S(t)=\prod_{i=1}^{N}(1-\frac{d(i)}{n(i)})$ Here d (i) is the number of events that take place at the same time with the event of the ith patient. n (i) is the number of patients who are still at risk when the ith patient experiences the event. In other words, n (i) indicates the number of patients that have more survival years than the ith patient and are not censored.

The Log logistic regression model was introduced in chapter 2. Before starting to plot the hazard functions, the model suitability should be checked. In order to do that, the non parametric odds must be calculated to be plotted versus the log t. The log odds are:

$$\frac{F(t,z)}{1-F(t,z)}=a\cdot t^{\gamma}\cdot\exp(z^{T}\beta)\Rightarrow$$

$$\Rightarrow\log[\frac{F(t,z)}{1-F(t,z)}]=\log[a\cdot t^{\gamma}\cdot\exp(z^{T}\beta)]\Rightarrow$$

$$\Rightarrow\log[\frac{F(t,z)}{1-F(t,z)}]=z^{T}\beta+\log(\alpha)+\gamma\log(t)\qquad(Eq.13)$$

The reason why the odds are used is obvious. They provide a linear representation of the cumulative density function F (t) against the log t. When the plot of the log odds against the log t is almost linear then the log logistic distribution is considered suitable for the description of the dataset. Theoretically only the odds are expected to be linear against the log t. If the plot doesn't follow this theoretical hypothesis as it is shown in equation 13, then log logistic distribution is not suitable.

The values of the survival function will derive from the Kaplan Meier estimator because the model requires non parametric estimates. The values of the density function can derive from F (t) =1 – S (t). The logarithmic ratio of those two will be plotted against the log t to check linearity and calculate the regression parameter β.

The Kaplan Meier estimator is: $S(t)=\prod_{i=1}^{N}(1-\frac{d(i)}{n(i)})$ where d (i) is the number of events that take place at the same time with the event of the ith patient. n (i) is the number of patients who are still at risk when the ith patient experiences the event. In other words, n (i) indicate the number of patients that have more survival years than the ith patient. After

having obtained the estimates from the Kaplan Meier, the values of the density function derive. The next step if to obtain the log odds ($\log[\frac{F(t,z)}{1-F(t,z)}]$) and plot them against the logarithm of time.

It is important to notice that this procedure has to be performed for a homogeneous population. In other words, patients with the same pattern of covariates will be plotted together. If the two populations are different in one covariate only for example z_1 and all the others are the same, the distance between the two lines gives a rough estimate of β_1. Two covariates will be used for the log logistic regression model, the epithelioid cellularity and the ultrasound height. The selection wasn't random. In order for the population to be homogeneous, it is divided into to two sets of patients, to those who have epithelioid melanoma (epi=1) and ultrasound height bigger than 10 and those who have epithelioid melanoma and uh<10. No other classification could be made since the ultrasound height and the longest ultrasound basal dimension have real numerical values.

The characteristics of those two groups of patients are applied in the Kaplan Meier estimator and their log odds are plotted against the log t. If a plot shows departure from linearity another model from the class of proportional odds models can be used. If linearity is satisfied but the distance between the lines is not constant a model assuming proportional odds model is not appropriate. In the case where linearity is satisfied and distance is constant this distance is an estimate of the regression parameter β.

Chapter 4
The application of Neural Network models

Chapter 4 analyzes neural networks only. There is an extended theoretical background in the beginning, explaining their applications and the existing network types. In addition this chapter introduces the two networks of this study for the Cox proportional hazards model.

4.1 Theoretical Background

The concept of the neural network came up as early as the middle of this century. A neural network is an data processing algorithm that is inspired by the way biological nervous systems process information. A neural network is an assembly of simple processing elements, the nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the inter-unit connection strengths, the weights, which are obtained by a learning process. [7]

Neural networks are used in statistical analysis and data modeling, in which their role is to perform a non linear regression or cluster analysis. Therefore they are typically used in problems of classification, image and speech recognition. In addition neural networks are applied in domains of human expertise such as medical diagnosis. As explained in chapter 1 neural networks, with their remarkable ability to derive meaning from complicated data can be used to extract patterns that are too complex to be noticed by other computer techniques. This fact made neural networks a powerful tool in statistical survival analysis.

The type of networks that is most commonly used in survival analysis is the multilayer feedforward network. The class consists of one or more hidden layers whose nodes are called hidden nodes. The function of hidden nodes is to intervene between the input and the network output in some useful manner. By adding one or more hidden layers the network can extract higher statistics which is particularly valuable if the size of the input layer is large.

The most commonly used training algorithm for the multilayer feed forward networks is the Backpropagation training algorithm. It is a supervised learning rule which offers a solution to the multilayer training problem. Backpropagation is a generalization of the Widrow -Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer functions. Input vectors and target vectors are used to train a network until it can approximate a function, associate the input to the output vectors, or classify input vectors in an appropriate way as defined by the designer.

Standard Backpropagation is a gradient descent algorithm, as is the Widrow-Hoff learning rule, in which the network weights are moved along the negative of the gradient of the performance function. The term Backpropagation refers to the way in which the gradient is computed for nonlinear multilayer networks. There are a number of variations on the basic algorithm that are based on optimization techniques, such as conjugate gradient and Newton methods. [8]

The basic idea is to give the input vector and calculate the output of each layer in the forward direction. For the output layer the desired values are known (supervised learning) and therefore the weights can be adjusted according to the gradient descent rule.

To calculate the weights in the hidden layer the error in the output layer is transferred back to these layers according to their connecting weights. This process is repeated for each sample in the dataset. One cycle through the dataset is called an epoch.

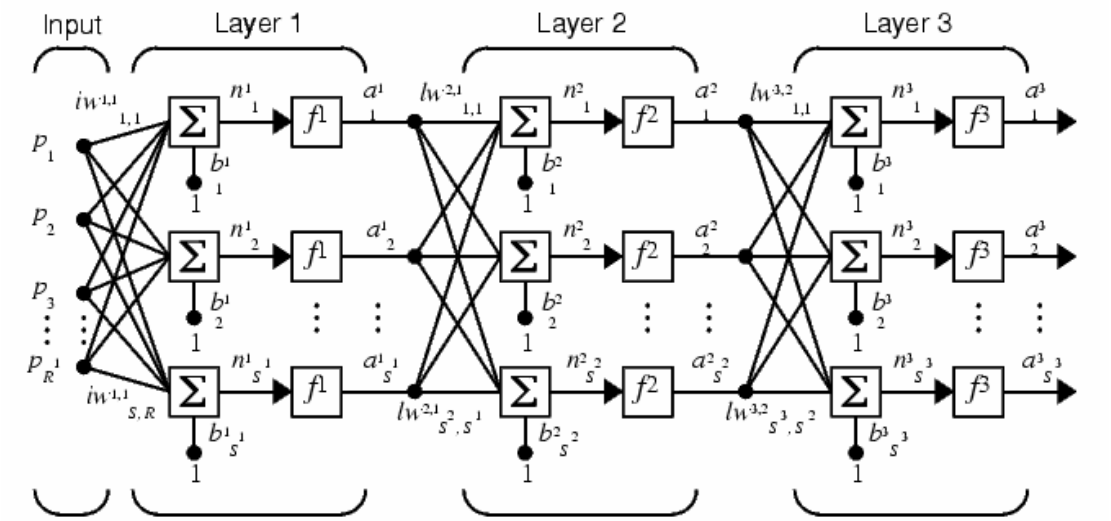
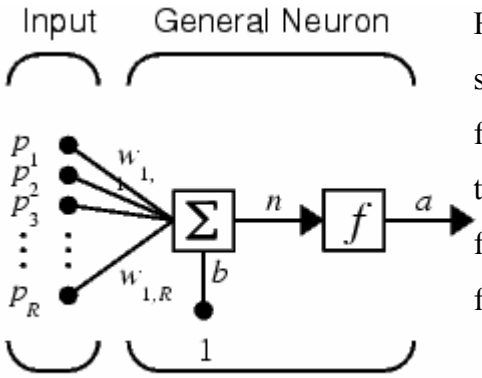


Figure 2. An example of a multilayer network with three hidden layers. [8]

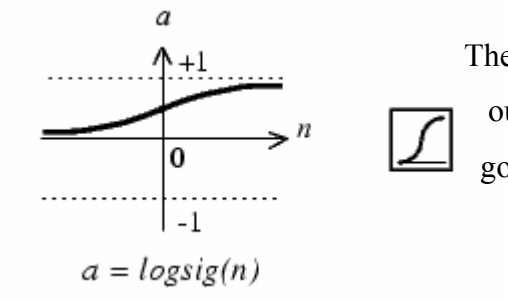
The transfer functions that can be used in the Backpropagation training algorithm are the purelin the tansig and the logsig (MATLAB environment). A simple neuron is shown in the picture below.



Here p1 to p_R are consist the input vector, w show the weights, b the bias f the transfer function. The sum of the weighted inputs and the bias forms the input to the transfer function f. Neurons may use any differentiable transfer function f to generate their output.

Figure 3. A layer of a network. [8]

The argument of the transfer function is the sum of the bias plus the weighted inputs. $n = w_{11}p_1 + w_{12}p_2 + ... + w_{1R}p_R + b$ Where R is the number of inputs. For the Backpropagation algorithm the transfer functions that can be used are:



The log sigmoid transfer function generates outputs between 0 and 1 as the neuron's net input goes from negative to positive infinity

Figure 4. The log sigmoid transfer function [8]

The tan sigmoid transfer function,

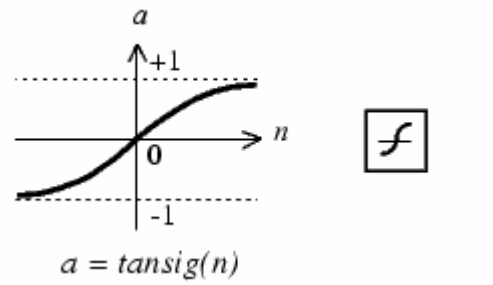
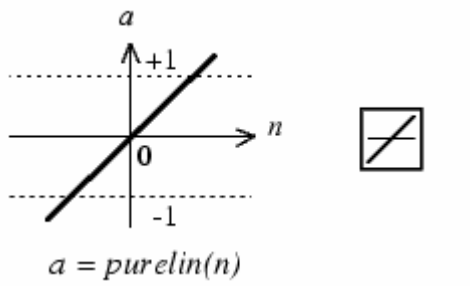


Figure 5. The tan sigmoid transfer function [8]



Occasionally the linear transfer function purelin is used in Backpropagation networks

Figure 6. The linear transfer function. [8]

There are many variations of the Backpropagation algorithm. In the simplest implementation, the network weights and biases change in the direction in which the performance function decreases most rapidly which is the negative of the gradient. An iteration of this algorithm is $x_{k+1} = x_k - a_k g_k$ where x_k is the vector of current weights and biases, a_k is the learning rate and g_k is the current gradient. There are two different ways in which this gradient descent algorithm can be implemented: incremental mode and batch mode. In the incremental mode, the gradient is computed while the weights are updated after each input is applied to the network. In the batch mode all of the inputs are applied to the network just before the weights are updated.

In multilayer feedforward networks the input neurons supply input data to the first hidden layer whose outputs constitute to the second hidden layer and so on. The outputs of the final layer represent the response of the network to the inputs fed by the input neurons of the input layer.

Learning process can either be supervised or unsupervised. In supervised learning the inputs and the outputs are provided. The network then tries to learn a set of input output pairs which constitute the training set. Therefore the weights are adjusted such as the network responds to any input by the desired output or with an accepted accuracy or error. This process occurs repeatedly as the weights are continually changed. During the training of the network the same set of data is processed many times as the connection weights are refined.

For faster training the Levenberg- Marquardt algorithm can be used to train the network. Like the quasi-Newton methods, the Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares, then the Hessian matrix can be approximated as $H = J^T J$ and the gradient descent can be estimated through $g = J^T e$ where J is the Jacobian matrix which contains the derivatives of the network errors. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$x_{k+1} = x_k - [H + \mu I]^{-1} J^T e$$

When the scalar μ is zero, this is just Newton's method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to turn towards Newton's method as quickly as possible. Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a step would increase the performance function. In this way, the performance function will always be reduced at each iteration of the algorithm. [8]

4.2 Neural Networks for the Ocular Melanoma dataset

The Cox proportional hazards model can be implemented with the use of the Backpropagation multilayer neural network. Two different nets will be modeled, one to simulate the baseline hazard function and another one to simulate the covariates (the exponential part of the model) Taking into consideration the general formula of the Cox model the two networks will be:

$$h(t, z_i) = h_0(t) \cdot \exp(z_i^T \beta)$$

Baseline network

Covariates network

The two networks are multilayer feed forward with the Backpropagation algorithm. The learning type is supervised in both cases, since the output target is known. For faster training the Levenberg- Marquardt algorithm is be used to train the networks too. The transfer and activation functions as well as the training epochs vary for each network.

The baseline function approximation as introduced in chapter3 can be performed with two different estimators. The first one was the Breslow estimator and the second one was the extended Kaplan Meier. Both of these estimators were extensions of simple estimators in the absence o covariates (Nelson Aalen, Kaplan Meier). Since the results were approximately equal only one of those two will be implemented with a neural network. The estimator that is most widely used in the Cox model for the baseline hazard approximation is the Breslow estimator.

The neural network will be a multilayer net with hidden nodes on which the Backpropagation training algorithm will be applied. The inputs and the output of the network

will be defined from the Breslow estimator $H_0(t) = \sum_{i(i)\le t} \frac{d(i)}{\sum_{l\in R(t_i)} \exp(z_l^T \beta)}$. The unknown parameters are the number of deaths at each time and the sum of $\exp(z_l^T \beta)$ for each patient at risk at the same time. Much attention must be given to the censoring times since d (i) refers only to deaths but $\exp(z_l^T \beta)$ refers to censored patients too. When censoring takes place, the patient is considered at risk at that time but not at the following times. The network couldn't be trained to include or cancel out censored patients because a censoring variable is not present at the estimator.

Therefore, the network is created to have one input, which would be a vector of two elements, the number of deaths d (i) and the sum $\exp(z_l^T \beta)$ of patients at risk. The output of the network is obviously one, the values of the cumulative hazard of the baseline hazard function. The input vector has length 240, since there are 240 discrete times when deaths take place among the 743 patients. The output is of the same length. In addition, 6 hidden layers were applied for the outcome to be reasonable since the input is vector is large. The first layer has 2 nodes, the next two hidden layers have 5 nodes each, the fourth 20 and the following two 5 nodes each. The transfer function applied to the input and the 5 hidden layers is the tan sigmoid while the output layer has a linear transfer function.

In order to have a network with inputs and outputs and training according to the error of each epoch, a Backpropagation rule was used. With this learning rule the network weights are moved along the negative of the gradient of the performance function.

The training algorithm is the Levenberg- Marquardt which was analyzed earlier in the chapter. The learning function is the Gradient descent momentum weight/bias learning function. This function calculates the weight change dW for a given neuron for the neuron's input, the error, the weight, the learning rate and the momentum constant MC, according to the gradient descent with momentum: dW = mc*dWprev + (1-mc)*lr*gW. The previous weight change dWprev is stored and read from the learning state LS.

The main idea is that in the future one could simulate this network by supplying the inputs (number of deaths and the sum exponential for patients at risk) to easily obtain the values of the cumulative baseline hazard function.

The input dataset was divided into two sets, one for simulating and one for training. The two sets had to be equal. When the original input (240 elements) was split into two sets, one containing the first 120 elements and the second the rest, simulation had very bad results. The training set and the simulation set must have same deviations and for this reason, the original set was divided with a rule that supports both big and small values in the two sets.

The Covariates Network is referred to the exponential part of the Cox proportional hazards model. The main idea is to create a network that provides the values of the exponential part for each patient. In the future any one would be able to obtain the exponential part by entering the covariate values.

$$h(t, z_i) = h_0(t) \cdot \exp(z_i^T \beta)$$

Given the formula of the Cox proportional hazards model, the inputs of the neural network are three. These will be the three covariates under examination, longest ultrasound basal dimension, ultrasound height and the epithelioid cellularity. The input values can either be numerical or binary. The regression parameter is not an input to the neural network so it won't be necessary in later studies to approximate it in order to simulate the network. The output of the network is the value of the exponential part for each patient. In other words, this network performs a regression, since the regression parameter β can be easily obtained from network's output.

The network that has been evaluated is a feed forward multilayer network with the Backpropagation learning rule. The learning is supervised again since the target output is known. The training function is the Levenberg- Marquardt which was analyzed earlier in the chapter. The learning function is the Gradient descent momentum weight/bias learning function.

In order to achieve a good performance our data is divided into three sets that will be used for training, testing and validation of the network correspondingly. The user can select the number of data vectors that each set will consist of. This is done by giving specific values at the variables NTR (training) and NTST (testing). The remaining data will be used for validation. Suppose a user enters 100 data vectors for training. In order to achieve efficient training, these 100 data vectors should not be sequential. Therefore an algorithm was created that selects 100 data vectors normally distributed across the entire data set. Hence, they are not sequential and will be used for training only. This means that a certain column of the input vector will be part of only one these three sets. This process is handled by indexing

on the input matrix. In other words we pick the sets as equally spaced points throughout the original data.

After defining the indexes for each set, 6 matrices are constructed. The first two are the input and output data for validation, the second is the input and output testing data and the rest is the input and output training data.

The Backpropagation Neural Net was chosen as it is one of the most powerful neural net types. Generally this network has the same structure as the Multi Layer Perceptron and it uses the Backpropagation learning algorithm. This Neural net type is a feedforward which means that that there are connections only between two different layers. Our purpose is to approximate a function effectively.

Chapter 5
Implementation Aspects and Results.

Chapter 5 provides the results of the three statistical methods under examination. Several implementation aspects are explained for the Kaplan Meier, the Cox model and the Log logistic regression model. In addition, the results from the log rank test are provided after the Kaplan Meier method since they are methodologically close. Finally, the neural networks results are presented and the correlations of the network's outputs and targets are provided.

5.1 Non parametric: Product limit estimate (Kaplan Meier).

The Kaplan Meier estimate is a simple way of computing the survival curve. It involves computing the number of people who died at a certain time point, divided by the number of people who are still at risk in the study.

The formula of the estimator is $S(t) = \prod_{i=1}^{N} (1 - \frac{d(i)}{n(i)})$. In this formula, N is the number of time points, d (i) is the number of deaths at each time point and n (i) is the risk set at this time. The risk set was analyzed in chapter 3.

The Kaplan Meier estimator was not applied to the whole original dataset. The reason is that in order to have analogue results with the Cox proportional hazards model the dataset had to be reduced to 743 patients (The Cox model examines covariates and there are missing values in the dataset, which leaves us with 743 patients with full records).

The estimator was programmed with the help of MATLAB environment the code of which is available in APPENDIX A. In addition it was programmed with the SPSS statistical tool the code of which is in APPENDIX A and the results in APPENDIX B.

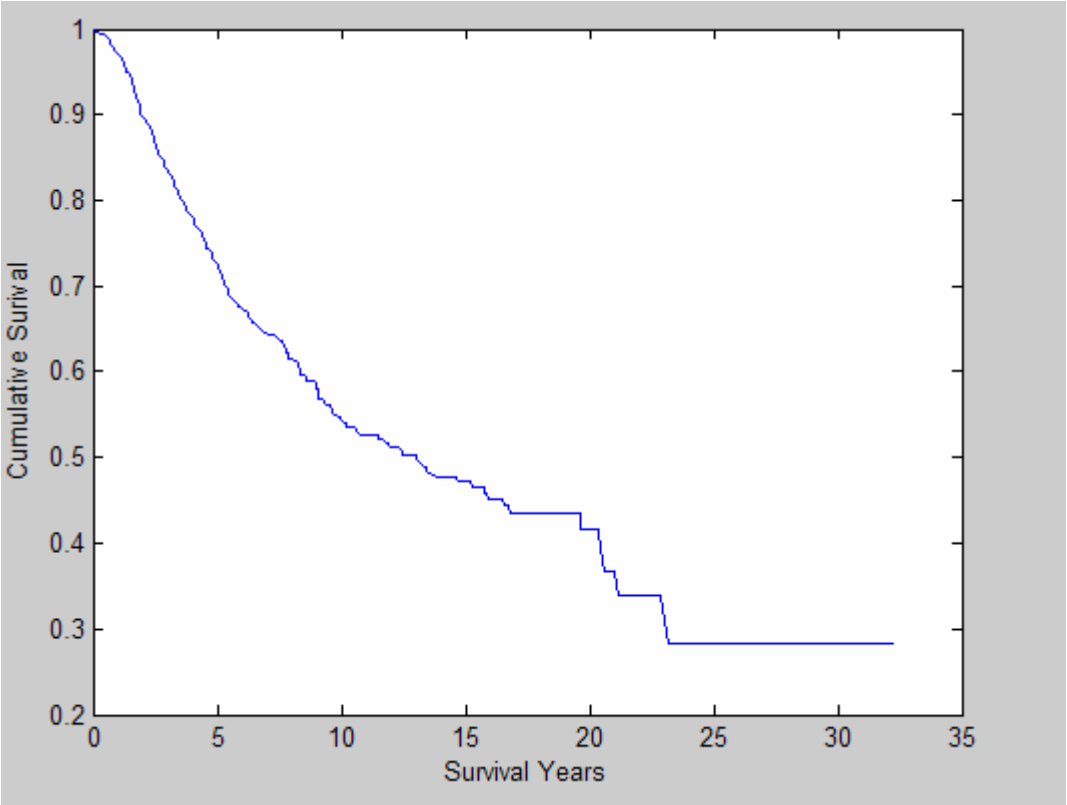


Figure 7. Survival curve for 743 patients from the ocular melanoma dataset obtained from the Kaplan Meier estimator

Figure 7 is a probabilistic plot that is very useful in survival analysis as it expresses the experience of 743 patients suffering from the ocular melanoma disease.

The vertical Y axis gives the proportion of people surviving. The value is a fraction which runs from 1 at the top to zero at the bottom representing 100% survival to zero percent survival at the bottom. Often the actual percentage is used rather that the proportion.

The horizontal X axis gives the time after the start of the observation experiment. Even if the observation o different patients started at different times the curve represents the experience of each patient from the time that observation began for that patient.

Any point of the curve gives the proportion or percentage surviving at a particular time after the beginning of the observation. A survival curve always starts out with 100% survival time at zero, the beginning. From there it can descend or stay level but it can never increase.

In the survival curve of figure 7, it is obvious that in the end of the study period there have been observed sudden deaths. These deaths alter the plot in a way that they affect the conclusions. After the 23rd year of the disease there is a big step down due to this deaths. These deaths effect might be statistically wrong and can be cancelled out. In the case where they are cancelled out the survivability at the 23rd year is 32% and not 29%.

Most survival curves are portrayed as staircase curves with a 'step' down to mean that there is a death. At the moment of each death the proportion of survivors decreases and it cannot change at any other time.

There are two main types of survival curves. First, curves which flatten to a level plateau and which suggest that patients are being cured, and second curves which descent all the way to zero implying that no one is cured.

The median survival is the time at which the percentage surviving is 50%. If more than half patients are cured there is no such point in the survival curve and the median is undefined. For the ocular melanoma disease the median survival is at 17th year of the disease. This suggests that after the 17th year half the patients die.

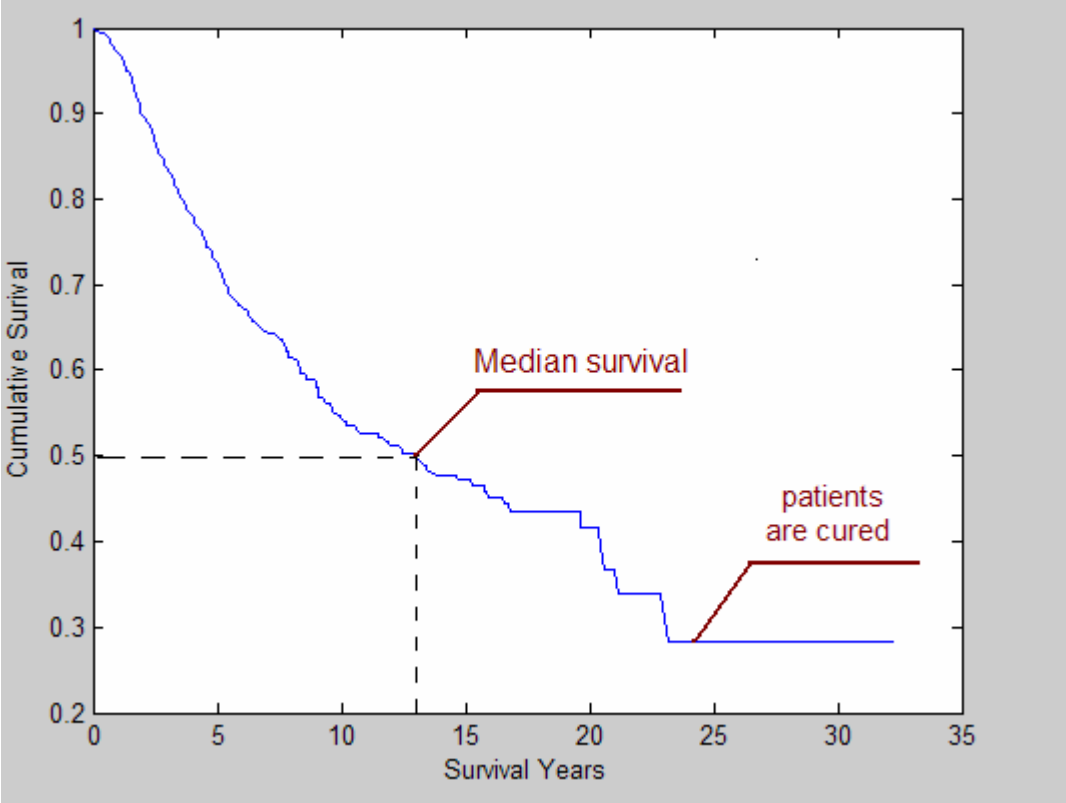


Figure 8. Kaplan Meier plot showing the Median survival.

The Kaplan Meier estimator can be implemented to obtain the survival curves of two groups, patients with epi 1 and epi 0. First the dataset is divided into two homogeneous sets concerning the epi variable. The first group has 464 patients with epi 1. The second group has 279 patients with epi 0. It is important to notice that not all patients from the original dataset are used, because there were missing values concerning the covariates. Therefore 743 patients with known covariates were selected to be part of the calculations.

In the first group there are totally 184 deaths and 280 censoring cases. In the second group there are 89 deaths and 190 censoring cases. The Kaplan Meier estimator is used ($S(t) = \prod_{i=1}^{N} (1 - \frac{d(i)}{n(i)})$) to obtain the following survival curves.

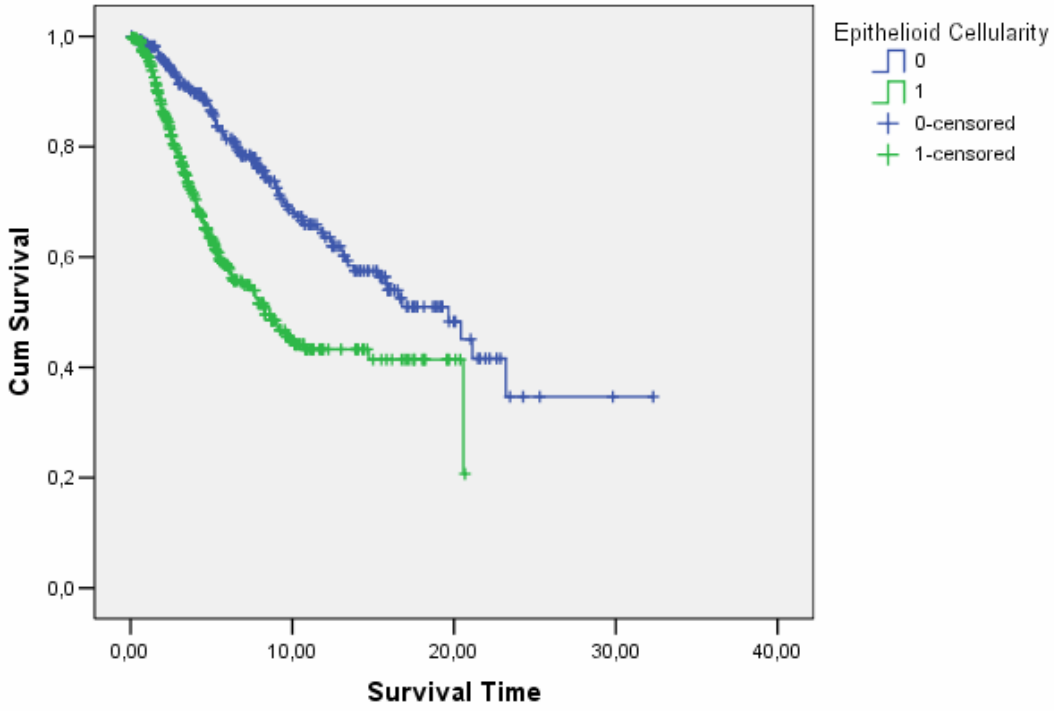


Figure 9. Kaplan Meier plots for two populations. The first group has epi=1 and the second epi=0. The curve is obtained from the SPSS environment.

From figure 9 it is clear that at any time spot that the patients with epithelioid cellularity are less likely to survive than patients without it. The median survival for patients with epi1 is at the 10^{th} years while for patients with epi 0 is at the 20^{th} year.

For the purposes of the log rank test, the null hypothesis has to be formulated. The groups under observation are two; the first one includes patients with epithelioid cellularity and the second patients without it. The null hypothesis is that the two groups are equally likely to survive at any given time point.

Null Hypothesis	$S_A(t)=S_B(t)$

The properties of each group under examination are:

	Group A (epi=1)	**Group B (epi=0)**	**Total**
Deaths	184	89	273
Censoring	280	190	470
Total	464	279	***743***

For every time point the expected number of deaths is calculated if there were in reality no difference between the groups and the null hypothesis was true. The real number of deaths is also calculated. For example, the first death occurred at time 0.02, when one patient from the first group (epi =1) died. At the start of this time period there were 743 patients alive in total, so the risk of death was 1/743. There were 184 patients in group A, so if the null hypothesis were true the expected number of deaths would be 464x1/743= 062. Likewise, in group B the expected number of deaths is 279x1/743= 0.37. The same calculations occur each time an event occurs. If a survival time is censored, that individual is considered to be at risk of dying at that time but not in subsequent times. This way of handling censored observations is the same as for the Kaplan Meier survival curve.

In the end of these calculations, the expected numbers of deaths for each group are summed as well as the true number of deaths. The chi square can be used to test the null hypothesis. The degrees of freedom are the number of groups minus one that is 1. The test statistic is computed through the formula: $K=\frac{(O_A-E_A)^2}{E_A}+\frac{(O_B-E_B)^2}{E_B}$. Here O is the real number of deaths for each group and E the expected number of deaths. The test statistic is a measure of difference between what was observed and what we'd expected if there were no difference between the populations. Asymptotic argument suggests that if there were no population differences the statistic K has approximately a chi square distribution with one

degree of freedom. Comparing k with chi square tables enables us to obtain a p value which measures the strength of the hypothesis of no population difference. The log rank and the chi square results are performed with the help of MATLAB environment the code of which is in APPENDIX A.

The test statistic is 30,525 (Also computed by the SPSS, APPENDIX B). This statistic is checked at the chi square distribution with 1 degree of freedom.

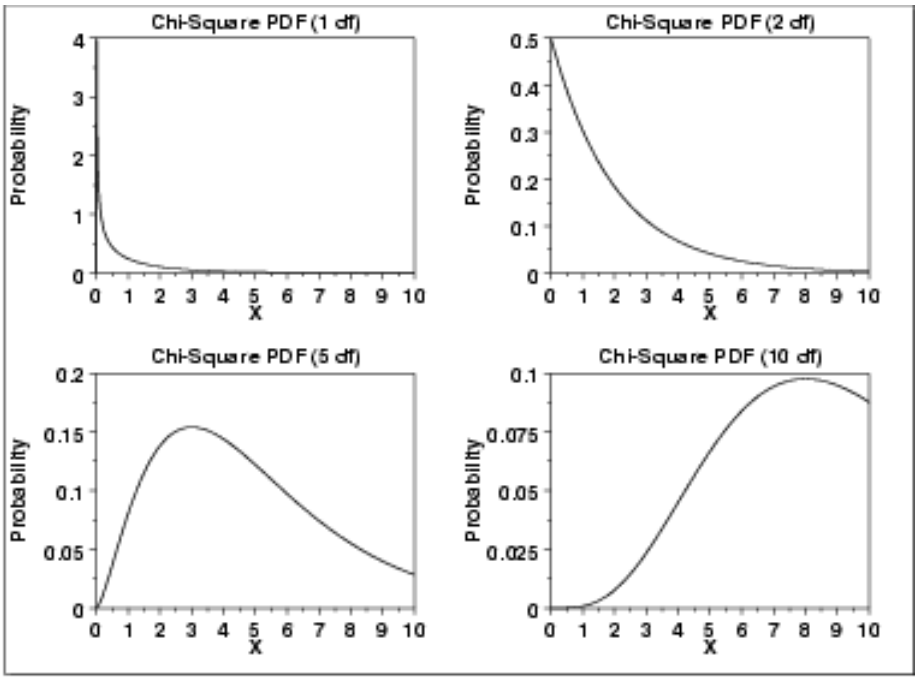


Figure 10. The chi square distribution for 1 degree of freedom.

Degree of freedom	***Probability p***				
/	**0.99**	**0.95**	**0.05**	**0.01**	**0.001**
1	0.000	0.004	3.84	6.64	10.83

Table 1. Chi square values for one degree of freedom.

Literature in chi square tests mentions that if the estimated K value is greater that the chi square value in the 0.05 column (at specific degrees of freedom) then there is a significant difference from the null hypothesis. This would mean that there is only 5% probability that the estimated K value would occur by chance.

The test statistic computed for the specific dataset is 30.525 with one degree of freedom. And it exceeds the chi square value in the 0.001 probability columns. This can make us even more confident that the null hypothesis can be rejected.

From the chi square distribution the corresponding probability is p<0.00000003295 and we can sensibly reject the null hypothesis. The p values is the probability that a test statistic at least as significant as the one observed would be obtained assuming the null hypothesis were true. The smaller the P value the stronger the evidence against the null hypothesis. The reason why the chi square is used in the log rank test is because the chi square distribution is non parametric and it does not require the data to be more or less normally distributed. In addition chi square can be applied only when observations are independent.

5.2 *Semi parametric*: The Cox Proportional Hazards Model

The semi parametric Cox model is a model that includes the covariates to examine their influence on the disease. For the purposes of this model, three covariates were selected to examine their effect on the evolution of the ocular melanoma disease. There are no time dependent variables in the ocular melanoma dataset and for this reason the simple version of the model is used. Since the covariates are time independent so that the proportionality assumption is met. The proportionality assumption is that the hazard ratio of two different patients with covariate vectors z_1 and z_2 is constant.

$$\frac{h(t,z_1)}{h(t,z_2)}=\frac{h_0(t)\exp(z_1^T\beta)}{h_0(t)\exp(z_2^T\beta)} \quad \Rightarrow \quad \frac{h(t,z_1)}{h(t,z_2)}=\frac{\exp(z_1^T\beta)}{\exp(z_2^T\beta)} \quad \Rightarrow \quad \frac{h(t,z_1)}{h(t,z_2)}=\exp[(z_1^T-z_2^T)\beta]$$

Since the covariates vectors z are time independent the expression $\exp[(z_1^T-z_2^T)\beta]$ is also constant over time and therefore the proportionality assumption is met.

In addition the three variables that are selected have to have some serious medical interest. Doctors should need the conclusions of this study in order to diagnose the severity of the ocular melanoma given a set of prognostic factors. The covariates selected were: *ultrasound height, longest ultrasound basal dimension* and the indicator of *epithelioid cellularity.* There are missing values in these three variables at the original dataset which is the reason why the dataset was reduced to include patients with known covariates. The new sub set dataset involves 743 patients.

The formula of the Cox proportional hazards model is $h(t, z_i) = h_0(t) \cdot \exp(z_i^T \beta)$. Z is the vector of covariates and β is the regression parameter. In addition $h_0(t)$ is the underlying baseline hazards function, that is a function of time which is covariates independent.

The baseline hazards function will be approximated with two different but close estimators, the Breslow estimator and an extension of the Kaplan Meier estimator. The regression parameter is estimated with maximum likelihood estimation as analyzed in chapter 3.

The regression was performed with the SPSS environment. SPSS took advantage only of uncensored patients (patients whose death indicator is 1). It is important that SPSS has a variable which indicates if a patient is censored or uncensored. In our case this variable is "death" which has binary values, 1 to indicate death and 0 to indicate censored. The SPSS also performs an Omnibus test. It first analyzes the data to quantify the asymmetry of the distribution and then calculates how much each of these values differs from the value expected from a Gaussian distribution and computes a single value from the sum of the squares of these discrepancies. The code for the SPSS and alternatively a MATLAB code are provided at APPENDIX A.

The estimated regression parameter is $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0.127 \\ 0.033 \\ 0.664 \end{pmatrix}$

The regression parameter β could have both positive and negative values. The values are all positive which means that the prognostic factors affect the disease. The higher these values are the higher is the hazard and eventually the lower is the survival probability. A value closer to 0 means that this variable affects the disease less. From the results above, the longest ultrasound basal dimension and the epithelioid cellularity affect the disease more. This means that a higher value of these values will result in higher hazard or risk of death.

A very important conclusion can be made here. A patient that appears to have epithelioid cellularity is less likely to survive than a person who doesn't. Therefore, the presence or not of epithelioid cellular can be a serious prognostic factor for this kind of disease.

Since the regression parameter is estimated, the Breslow estimator can be used to approximate the baseline hazard function. The formula of this estimator is

$H_0(t) = \sum_{t(i)\le t} \frac{d(i)}{\sum_{l\in R(t_i)} \exp(z_l^T \beta)}$ where d (i) is the number of deaths at each distinct time, and R (t) is the risk set at each specific time point. The idea of the risk set was analyzed in chapter 2. The estimator is programmed in the MATLAB environment the code of which is available in APPENDIX A. It has also been performed with the help of the R language, the code of which is in APPENDIX A and results in APPENDIX B.

Obviously Breslow's formula is an estimator and not a function used to obtain separate estimations of the cumulative hazard at each death time. In order to obtain the distribution from this estimator‘s output MATLAB environment was used to perform the fitting. To make sure that there is only one distribution and a standard selection of its parameters, the data was divided into three different sets at random. Each set was tested to examine the quality of fitness to a certain distribution. The main idea is that once the same distribution fits properly three different sets, then it will certainly fit their union.

To examine the quality of fitness some goodness parameters were estimated to indicate a good or bad fit. Those parameters were:

Sum of squares due to error. This statistic measures the total deviation of the response values. It is also called the summed square of residuals and is usually labeled as SSE. A value closer to 0 indicates a better fit. $SSE = \sum_{i=1}^{n} w_i (y_i - \widehat{y_i})^2$

R-square. This statistic measures how successful the fit is in explaining the variation of the data. In other words it is the square of the correlation between the response values and the predicted response values. It can take any value between 0 and , with a value closer to 1 indicating a better fit. $R-square = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$

Where $SST = \sum_{i=1}^{n} w_i (y_i - \overline{y})^2$ and $SSR = \sum_{i=1}^{n} w_i (\widehat{y_i} - \overline{y})^2$

Adjusted R square. This statistic is generally considered to be the best indicator of the fitness, for model with coefficients. It can take values less than or equal to 1, with 1 indicating the best fit. $Adjusted \;\; R-square = 1 - \frac{SSE(n-1)}{SST(u)}$

Root mean square error. This one is also known the fit standard error of the regression. $RMSE = \sqrt{\frac{SSE}{u}}$

(Note: n = number of response values, m = number of fitted coefficients, u = n-m)

All relevant code is available in APPENDIX A. The Breslow estimator is used to compute the cumulative hazard values and then these values are fitted with the help of MATLAB in a known distribution. During the fitting process the parameters of the distribution are estimated.

The quality of fitness is more than accepted since the parameters which were explained earlier have acceptable values:

Measure	**Value**	**Good Fit values**
Sum of squares due to error	0,0022721	Close to 0
R-square	0,98346	Close to 1
Adjusted R- square	0,98339	Close to 1
Root mean square error	0,0030897	Close to 0

The distribution that turned out to fit the data best is $H_0(t) = \log(1 + at^{\gamma})$ and the parameters α and γ are estimated to be $H_0(t) = \log(1 + 0.005176t^{1.036})$. The parameters are estimated as long as the function is fitted.

It can be proved that $H_0(t) = \log(1 + at^{\gamma})$ is the *log logistic Cumulative Hazard.*

$$H(t) = \log(1 + at^{\gamma}) \ , \ S(t) = \exp[-H(t)] \ \Rightarrow \ S_0(t) = \frac{1}{1 + a \cdot t^{\gamma}} \quad (7)$$

$$H(t) = \int_0^t h(x)dx \quad \Rightarrow h(t) = \frac{dH(t)}{dt} = \frac{\alpha\gamma t^{\gamma-1}}{1 + \alpha t^{\gamma}} \quad (8)$$

From the definitions, the relationship of the hazard function to the cumulative survival function is: $h(t) = \frac{f(t)}{S(t)}$ (9).

Where f (t) is the probability density function (in the case of the log logistic distribution the probability density function is known to be: $f(t) = \frac{\alpha\gamma t^{\gamma-1}}{(1 + \alpha t^{\gamma})^2}$) Equations 7 and 8 are replaced in equation 9.

$$h(t) = \frac{f(t)}{S(t)} \quad \Rightarrow f(t) = h(t) \cdot S(t)$$

$$f(t) = \frac{\alpha\gamma t^{\gamma-1}}{1 + \alpha t^{\gamma}} \cdot \frac{1}{1 + \alpha t^{\gamma}}$$

$$f(t) = \frac{\alpha\gamma t^{\gamma-1}}{(1 + \alpha t^{\gamma})^2}$$

This is indeed the probability density function of the well known log logistic distribution.

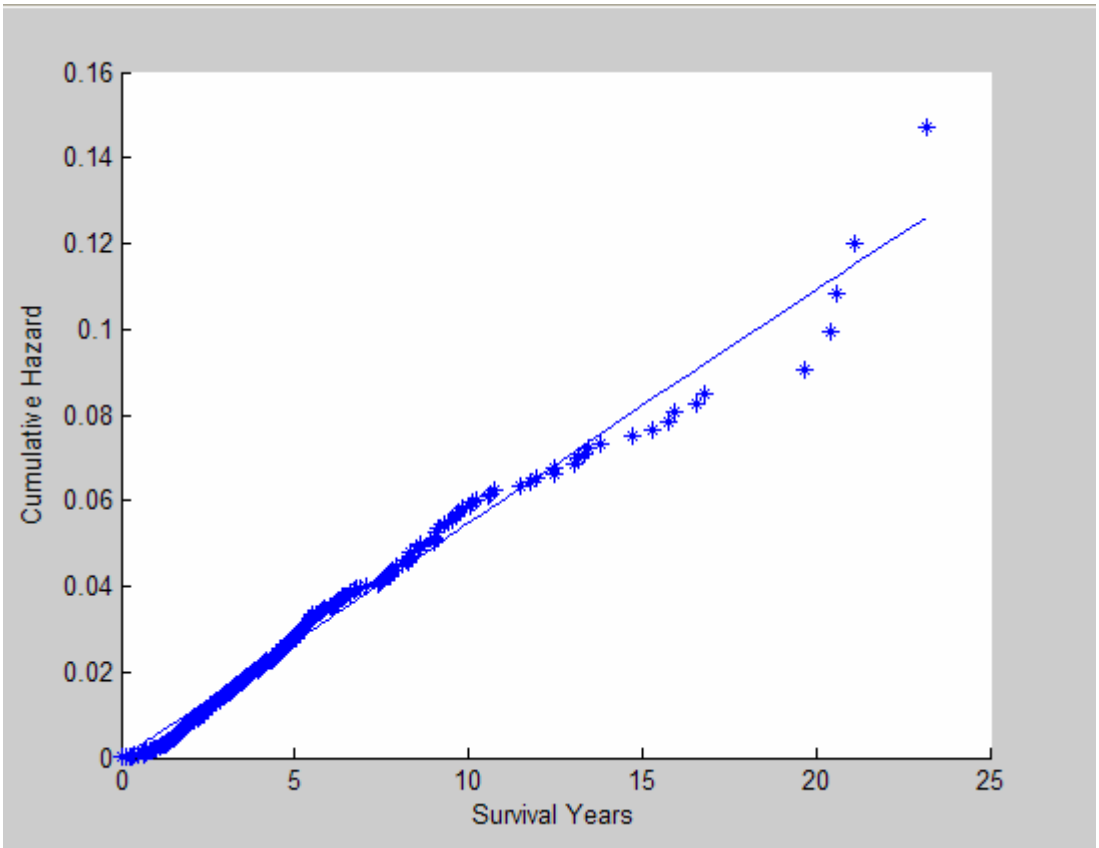


Figure 11. Cumulative hazard obtained from the Breslow estimator and it's fit to the log logistic cumulative hazard function.

The baseline hazard function can be derived form definition 7 of chapter 2 (where the hazard function derives from the survival function $-\int_0^t h(x)dx = \log S(t)$). Therefore, the baseline hazard function is $h_0(t) = \frac{\alpha\gamma t^{\gamma-1}}{1+\alpha t^{\gamma}}$.

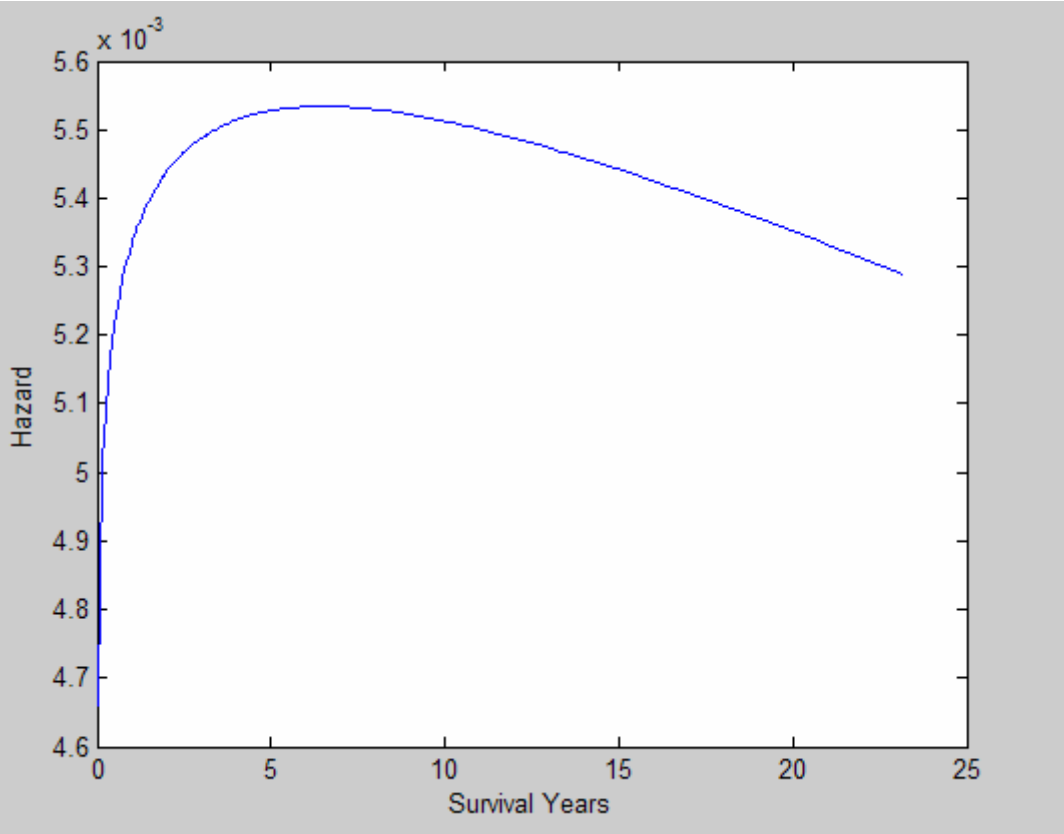


Figure 12. The baseline hazard function obtained from the Breslow estimator and definition 7 of chapter 2

]Having obtained the cumulative hazard function, the cumulative survival can be easily obtained using $S(t) = \exp[-H(t)] \Rightarrow S_0(t) = \frac{1}{1 + a \cdot t^{\gamma}}$ and with the estimated parameters $S_0(t) = \frac{1}{1 + 0.005176 \cdot t^{1.036}}$. The plot of the Cumulative survival function (derived from the Breslow estimator) is:

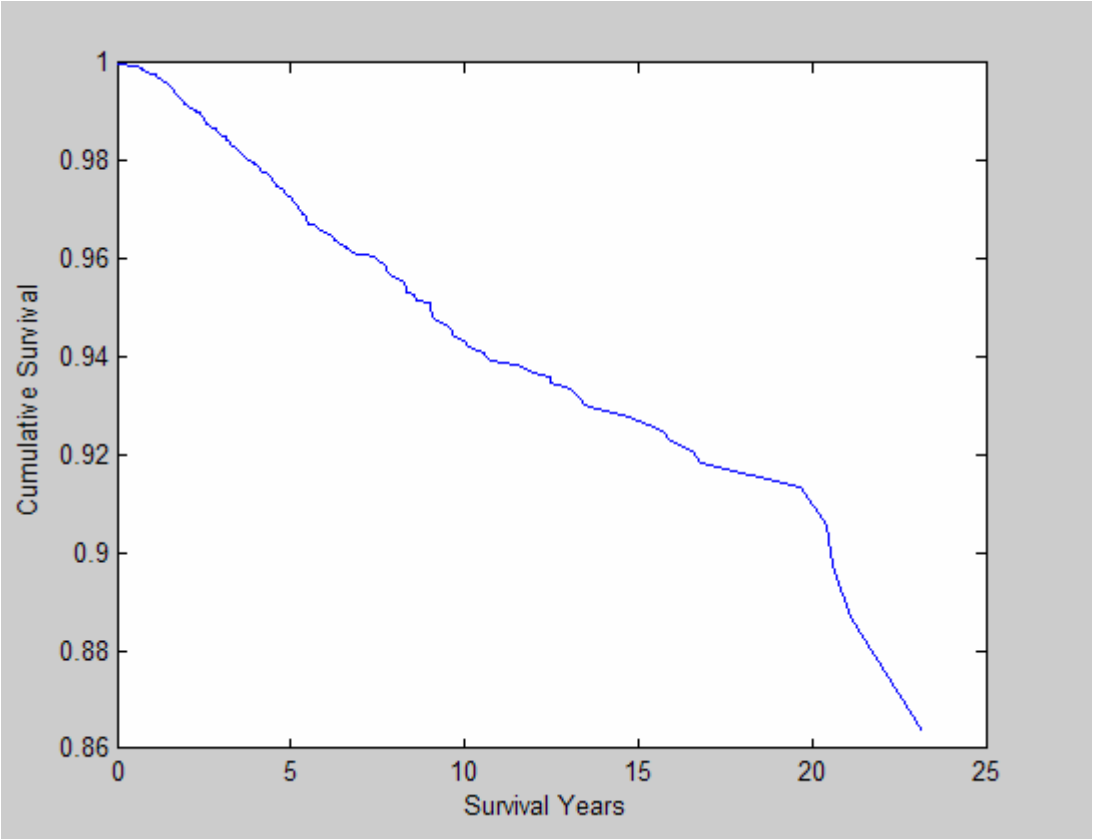


Figure 13. The Cumulative survival function from the Breslow estimator. This plot cannot be considered probabilistic and is much different from the Kaplan Meier survival curve.

Having obtained the baseline hazard function and the regression parameter, the hazard function can be plotted for each patient. Each patient has a specific covariate vector, which is used in the exponential non parametric part of the Cox model in a multiplicative way. In this way, every patient is expected to raise the hazard function on the y axis at an amount proportional to this individual's covariates.

In practice, it is not the hazard function plot that is needed to make conclusions. The Cox model is not just useful in estimating and plotting hazard plots. The main interest is not the hazard function of a certain individual but the form of the baseline hazard function and the values of the regression parameters.

From the plot of the density hazard function one can see how the hazard changes as years pass. It is clear that there is a hazard peak at the 5^{th} year of the disease. This conclusion is justified by the doctors too. It is generally believed that the 5^{th} year of the ocular melanoma disease is critical concerning the survivability. In other words, a patient is more likely to experience the event before the 5^{th} year rather than after it.

Figure 12 is the plot of the baseline hazard function. The dependent variable is the survival of patients which is counted on years. This plot can not be interpreted as a probabilistic plot. Its peak does not show death (Every patient's hazard function will have the exact same form, but this doesn't mean that all patients experienced the event at the 5^{th} year).

There is large distance between the Survival curve of the Breslow estimator and the Kaplan Meier estimator that was analyzed in earlier in this chapter. The risk set in the Kaplan Meier estimator included patients as units. On the contrary the risk set in the Breslow estimator includes the exponential of patient's covariates. In addition, the figure above can not be interpreted as probability of dying, or the proportion of people surviving due to the exact same reason. The cumulative survival or cumulative hazard curves that are obtained form the Breslow estimators are only used to obtain the baseline hazard function and not to make consumptions.

The Kaplan Meier estimator is also an estimator of a cumulative survival function in the absence of covariates. This survivorship function shows how the probability of surviving changes through time, living aside the existence of any kind of covariates. Thus, there is a theoretical relationship of the Kaplan Meier's estimator for the survival function, with the baseline hazard function.

The formula of the Kaplan Meier estimator is $S(t) = \prod_{i=1}^{N} (1 - \frac{d(i)}{n(i)})$. (Eq. 10)

Here d (i) is the number of events that take place at the same time with the event of the ith patient. n (i) is the number of patients who are still at risk when the ith patient experiences the event. In other words, n (i) indicate the number of patients that have more survival years than the ith patient and are not censored.

Of course we can not cancel out completely the fact there are some covariates that affect the survivorship. The main idea of the second step for this implementation is that the Kaplan Meier survival function can be transformed to include somehow the effects of the covariates.

This implementation is completely analogue to the Nelson Aalen and Breslow estimators in chapter 6.1.2. Therefore, an analog of the Kaplan Meier estimator can be derived for S (t) by thinking if h_0 (t) in terms of a discrete hazard having mass $\frac{d(i)}{\sum_{l \in R(t_i)} \exp(z_l^T \beta)}$ at each failure time. This could be interpreted as the probability that m individuals with covariates x=0 fail at t_i conditional on the sets of covariates observed individuals at risk at t_i. Therefore the analogue of the Kaplan- Meier estimator is:

$$S_0(t) = \prod_{t(i) \le t} \left[1 - \frac{d(i)}{\sum_{l \in R(t_i)} \exp(z_l^T \beta)} \right] \quad (11)$$

This estimator is actually the same with the Kaplan Meier (Eq. 10) for β=0 which theoretically means that no covariates affect the survivorship of an individual.

Obviously, the extended Kaplan Meier is an estimator and not a function used to obtain separate estimations of the cumulative survival at each death time. In order to obtain the distribution from this estimator‘s output the MATLAB environment is used to perform the fitting. To make sure that there is only one distribution and a standard selection of its parameters, data is divided into three different sets at random. Each set is tested to examine the quality of fitness to a certain distribution. The main idea is that once the same distribution fits properly three different sets, then it will certainly fit their union.

To examine the quality of fitness some goodness parameters were estimated to indicate a good or bad fit. Those parameters were: **Sum of squares due to error, R-square, Adjusted R square and Root mean square error.**

The analogue to the Kaplan Meier estimator is performed with the help of MATLAB environment (APPENDIX A). Having obtained the values of the survival function from the analog to the Kaplan Meier estimator, they are fitted in a distribution in a way that the quality of fit is accepted, to obtain the survival function. MATLAB environment is used to carry out this process (APPENDIX A). The method used was the non linear least squares fitting. In addition, to make sure there were only one appropriate distribution and a standard selection of parameters, data is divided again into three random sets, which are checked individually for their fitness to the same distribution.

The distribution that turns out to fit the data best is $S_0(t) = \frac{1}{1 + a \cdot t^{\gamma}}$ (7) and the parameters α and γ are estimated to be $S_0(t) = \frac{1}{1 + 0.005169 \cdot t^{1.037}}$. It can be proved that equation 7 is the *log logistic Cumulative Survival.*

$$S(t) = \exp[-H(t)] \quad \Rightarrow \quad H(t) = -\log S(t) \quad \Rightarrow \quad H(t) = -\log \frac{1}{1 + a \cdot t^{\gamma}} \quad \Rightarrow \quad H(t) = \log(1 + at^{\gamma})$$

$$H(t) = \int_0^t h(x)dx \qquad \Rightarrow h(t) = \frac{dH(t)}{dt} = \frac{\alpha\gamma t^{\gamma-1}}{1 + \alpha t^{\gamma}} \quad (8)$$

From the definitions, the relationship of the hazard function to the cumulative survival function is: $h(t) = \frac{f(t)}{S(t)}$ (9) where f (t) is the probability density function (in the case of the log logistic distribution the probability density function is known to be: $f(t) = \frac{\alpha\gamma t^{\gamma-1}}{(1 + \alpha t^{\gamma})^2}$

Equations 7 and 8 are replaced in equation 9.

$$h(t) = \frac{f(t)}{S(t)} \qquad \Rightarrow f(t) = h(t) \cdot S(t)$$

$$f(t) = \frac{\alpha\gamma t^{\gamma-1}}{1 + \alpha t^{\gamma}} \cdot \frac{1}{1 + \alpha t^{\gamma}}$$

$$f(t) = \frac{\alpha\gamma t^{\gamma-1}}{(1 + \alpha t^{\gamma})^2}$$

This is indeed the probability density function of the well known log logistic distribution.

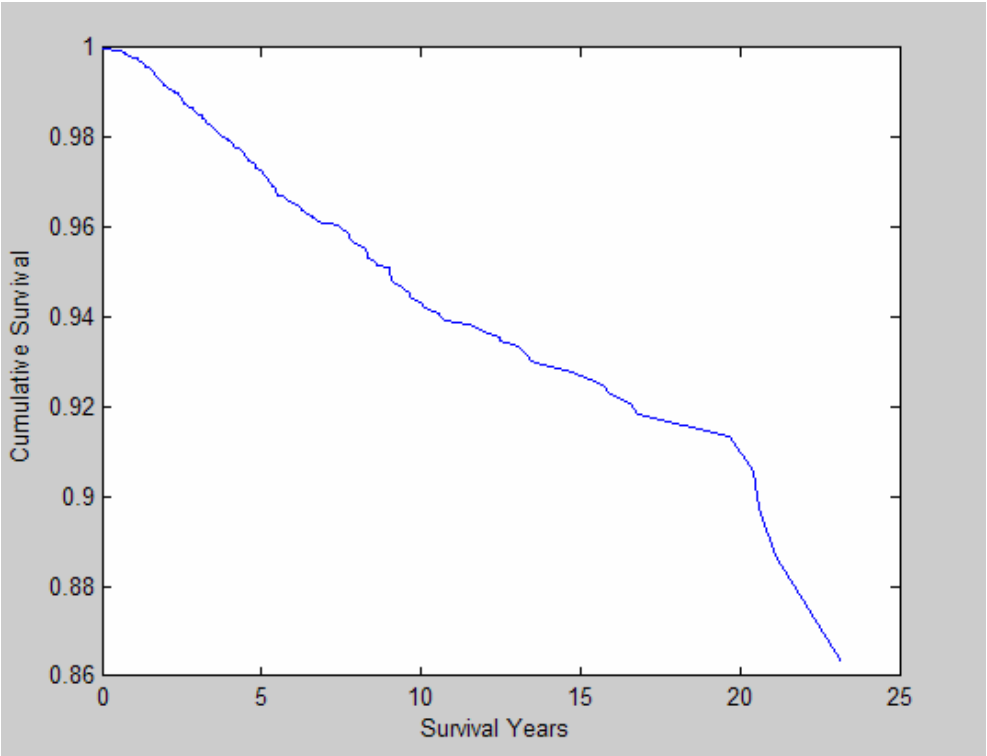


Figure 14. The baseline cumulative survival function from the extended Kaplan Meier estimator. This plot is very close to the one obtained form the Breslow estimator in figure 13.

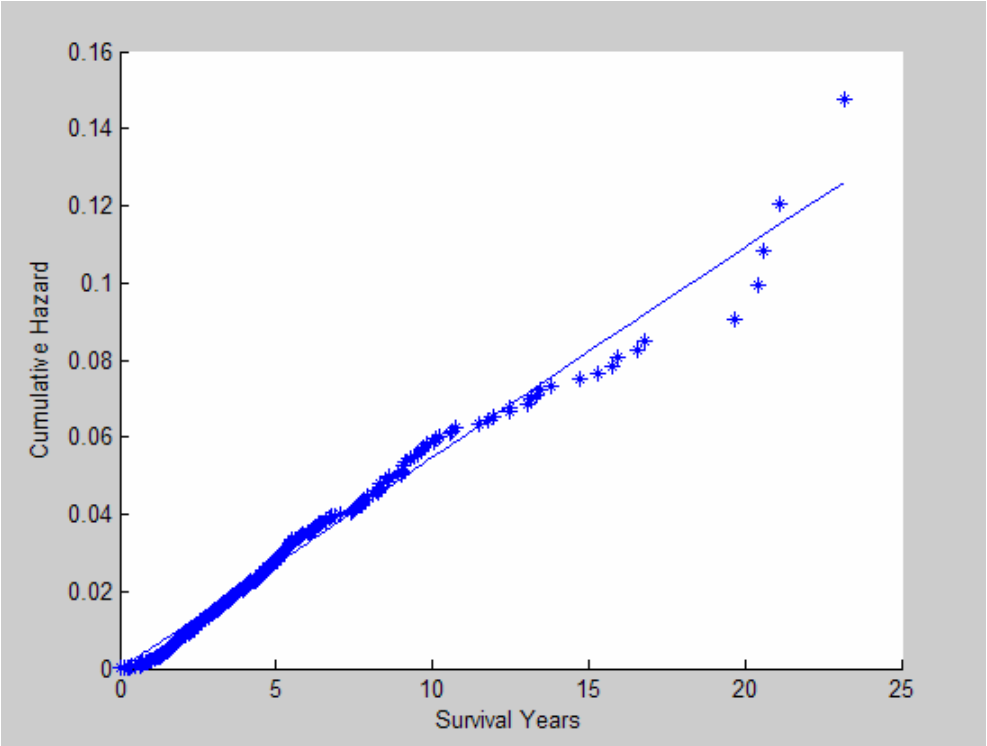


Figure 15. The baseline cumulative hazard function obtained from the survival function of figure 14. This function derives from the extended Kaplan Meier estimator. Figure shows its fit to the log logistic distribution.

The quality of the fitness is good. The parameters checked for the quality are:

Measure	Value	Good Fit values
Sum of squares due to error	0,0022926	Close to 0
R-square	0,98334	Close to 1
Adjusted R- square	0,98327	Close to 1
Root mean square error	0,0031037	Close to 0

We continue with the exact same process with the first implementation. The density hazard function can be mathematically obtained from the cumulative hazard function.

$$H(t) = \int_0^t h(x)dx \qquad \Rightarrow h(t) = \frac{dH(t)}{dt} = \frac{\alpha\gamma t^{\gamma-1}}{1+\alpha t^{\gamma}} \quad (8)$$

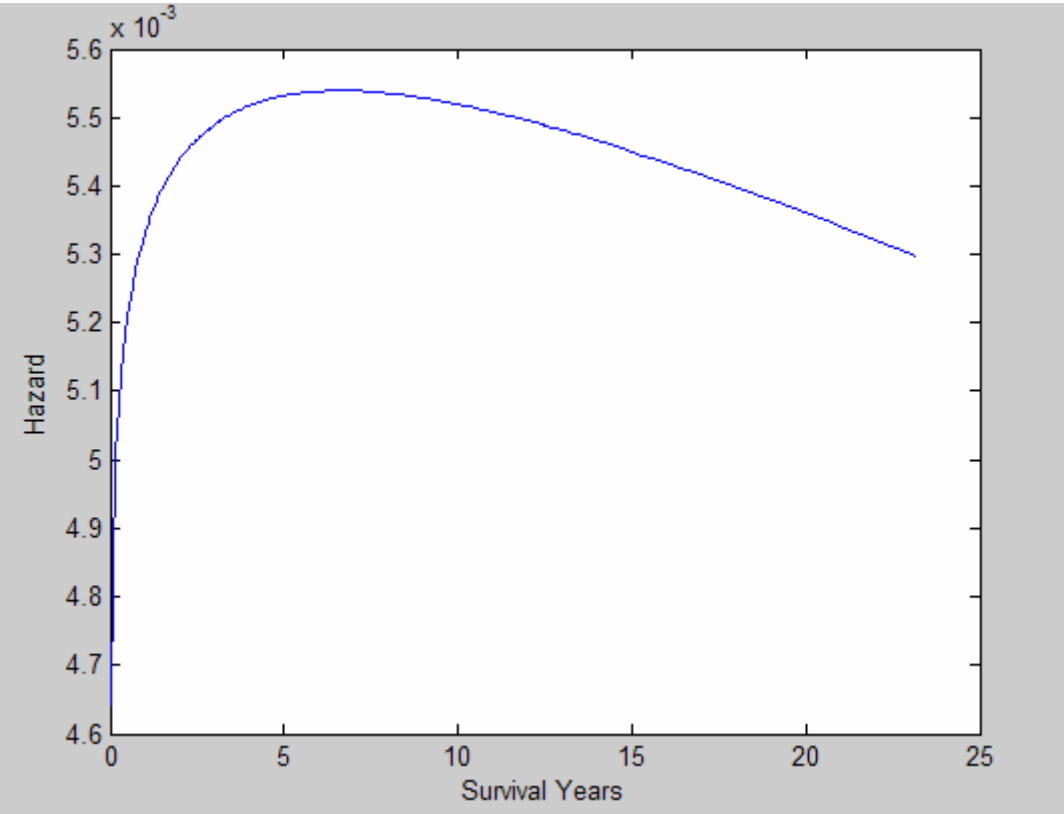


Figure 16. The baseline hazard function derived from the extended Kaplan Meier estimator. This function is very close to the one obtained from the Breslow estimator in figure 12.

From the plot of the density hazard function one can see again how the hazard changes as years pass. It is clear that there is a hazard peak at the 5th year of the disease. Again, there is higher probability that a patient will experience the event before the 5th year rather than after it.

Figure 16 is the plot of the baseline hazard function. The dependent variable is the survival of patients which is counted on years. This plot can not be interpreted as a probabilistic plot. Its peak does not show death (Every patient's hazard function will have the exact same form, but this doesn't mean that all patients experienced the event at the 5^{th} year).

The Breslow estimator and the extended Kaplan Meier estimator have mathematical similarities as explained in chapter 3 and therefore they conclude in close values.

This conclusion is perfectly met in this project since the values of the cumulative hazard, survival and density hazard from the two estimators are very close. This is a reasonable conclusion considering that the Extension of the Kaplan Meier is a derivative of the simple Kaplan Meier estimator. In addition the Breslow estimator is e derivative of the Nelson Aalen estimator. In general the simple Kaplan Meier and the Nelson Aalen estimators conclude in close values for the survival function. It is therefore reasonable that their extensions will act the same since covariates appear multiplicatively in both of them.

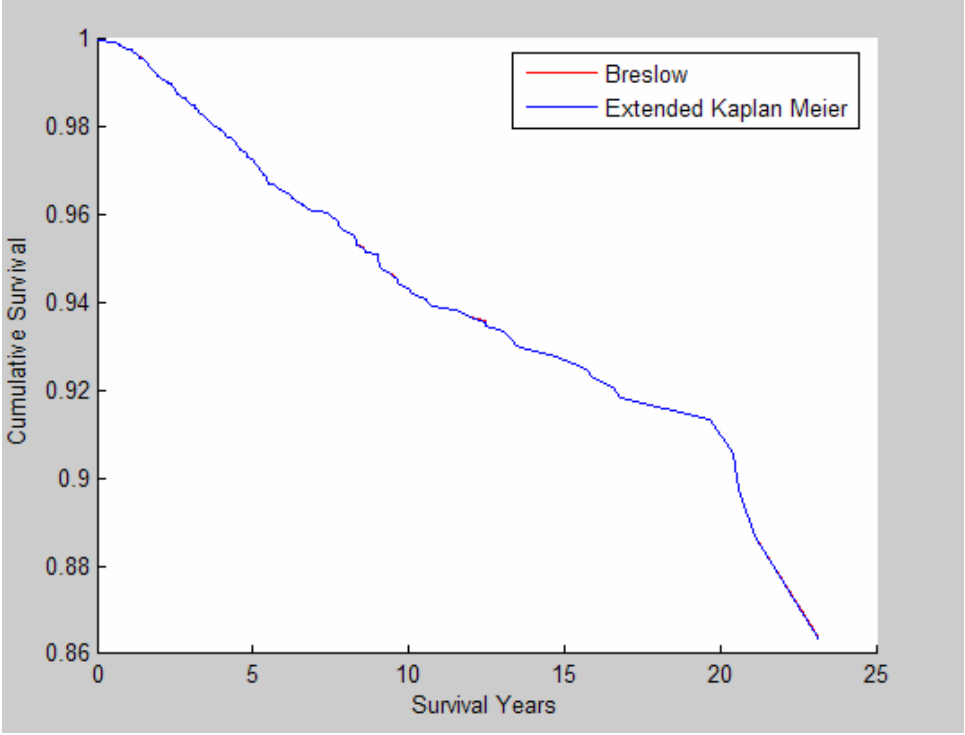


Figure 17. Cumulative survival functions from Breslow and Extended Kaplan Meier estimators.

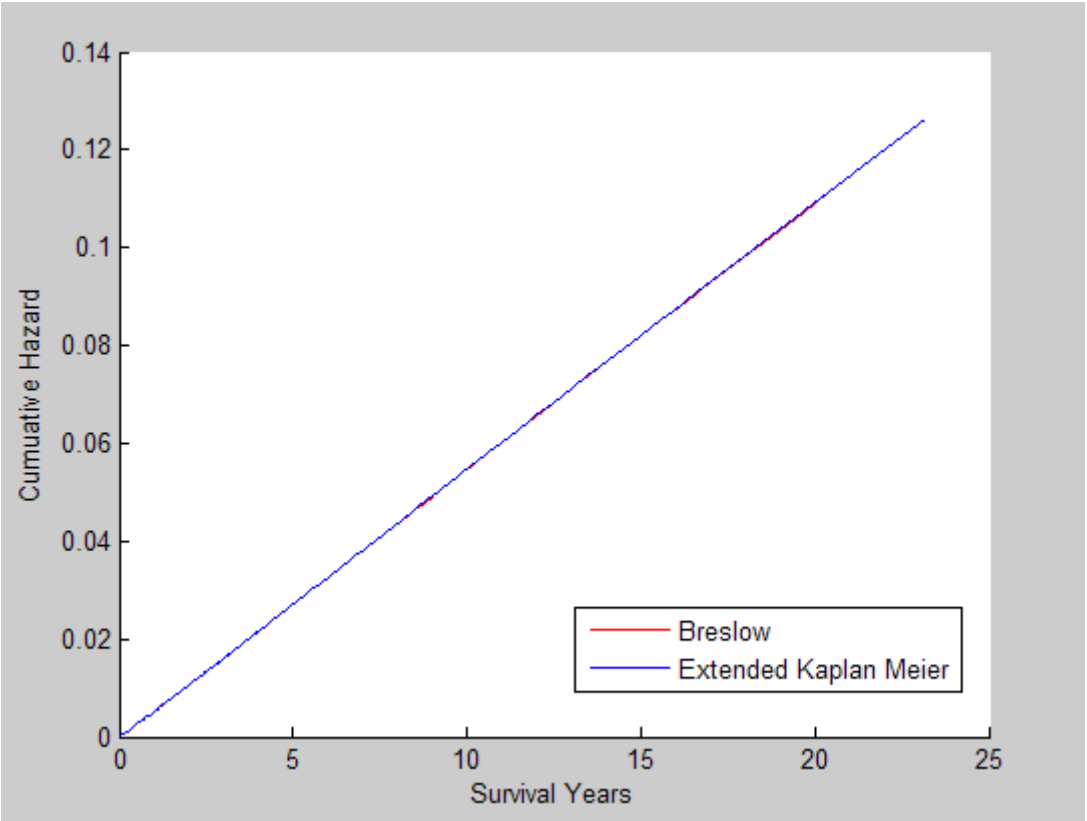


Figure 18. Cumulative hazard functions from the Breslow and the extended Kaplan Meier estimators.

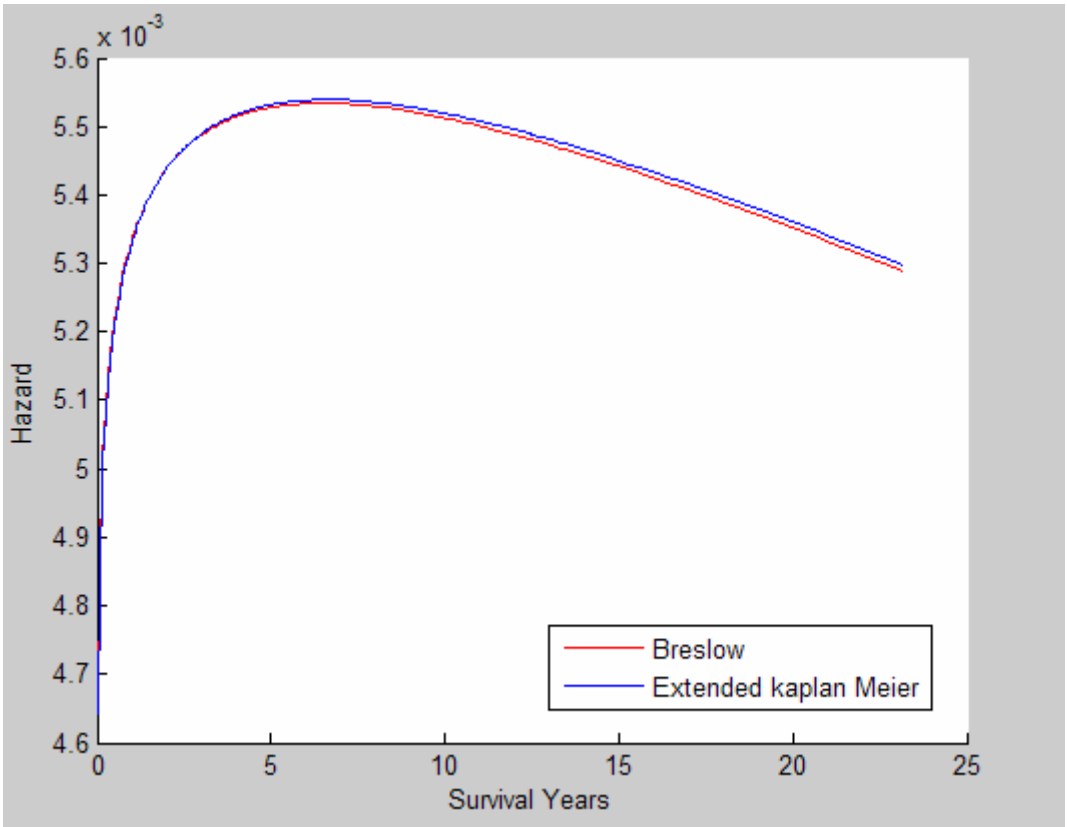


Figure 19. Baseline hazard functions from Breslow and extended Kaplan Meier estimators.

The red line in figures 17, 18 and 19 is for the hazard plot obtained through the Breslow estimator while the blue one is from the analog to Kaplan- Meier estimator. The two plots have the exact same form, which was expected since the same sample was examined in both cases. Furthermore, the implementations were examined in the terms: Regression parameter independent from the baseline hazard distribution and a hazard function that does not cancel out the regression parameter.

Of course there are some differences between these two plots, which happens because the two estimators are not 100% equivalent. Of course a part of these deviations is due to the MATLAB incapacity to approximate perfectly a given set of variables.

At this point it is essential to identify each estimator that was used in the Cox model to obtain the baseline hazard function. The following table summarizes these estimators as well as showing their origin.

The estimator in absence of covariates	***The estimator in presence of covariates***	
Nelson Aalen $H_0(t) = \sum_{t(i)\le t} \frac{d(i)}{n(i)}$	**Breslow** $H_0(t) = \sum_{t(i)\le t} \frac{d(i)}{\sum_{l\in R(t_i)} \exp(z_l^T \beta)}$	Incomparable
Kaplan Meier $S(t) = \prod_{i=1}^{N} (1 - \frac{d(i)}{n(i)})$	**Extended Kaplan Meier** $S_0(t) = \prod_{t(i)\le t} \left[1 - \frac{d(i)}{\sum_{l\in R(t_i)} \exp(z_l^T \beta)} \right]$	Incomparable
Comparable	Comparable	

The extended Kaplan Meier estimator is comparable to the Breslow estimator as proven in figures 17, 18 and 19. The Kaplan Meier is also comparable to the Nelson Aalen estimator. In addition, there can not be any comparison among the Breslow and the Nelson Aalen estimator since only the second one is a probabilistic plot. On the other hand the survival function obtained from the Cox model (using the Breslow estimator) is well compared to the Nelson Aalen and the Kaplan Meier estimator.

The Cox regression with the maximum likelihood estimation was used to obtain the regression parameter β. Analyzing β we have proven that the epithelioid cellularity is the prognostic factor that affects the disease most.

In addition the baseline hazard function was approximated by implementing the Breslow and the extended Kaplan Meier estimator. Therefore all parts of the Cox formula have been estimated and we can use $h(t, z_i) = h_0(t) \cdot \exp(z_i^T \beta)$ to plot the hazard function for each patient. The higher the exponential part of the formula the higher does it displace the baseline hazard function. Therefore, a patient with higher probability of dying has higher values on the y axis of the hazard curve.

Useful conclusions can be made when plotting the hazard functions of patients with epithelioid cellularity versus patients with epi=0. It is expected that patients with epi=1 will have higher curves than those with epi=0. In addition this distinction has to be clear.

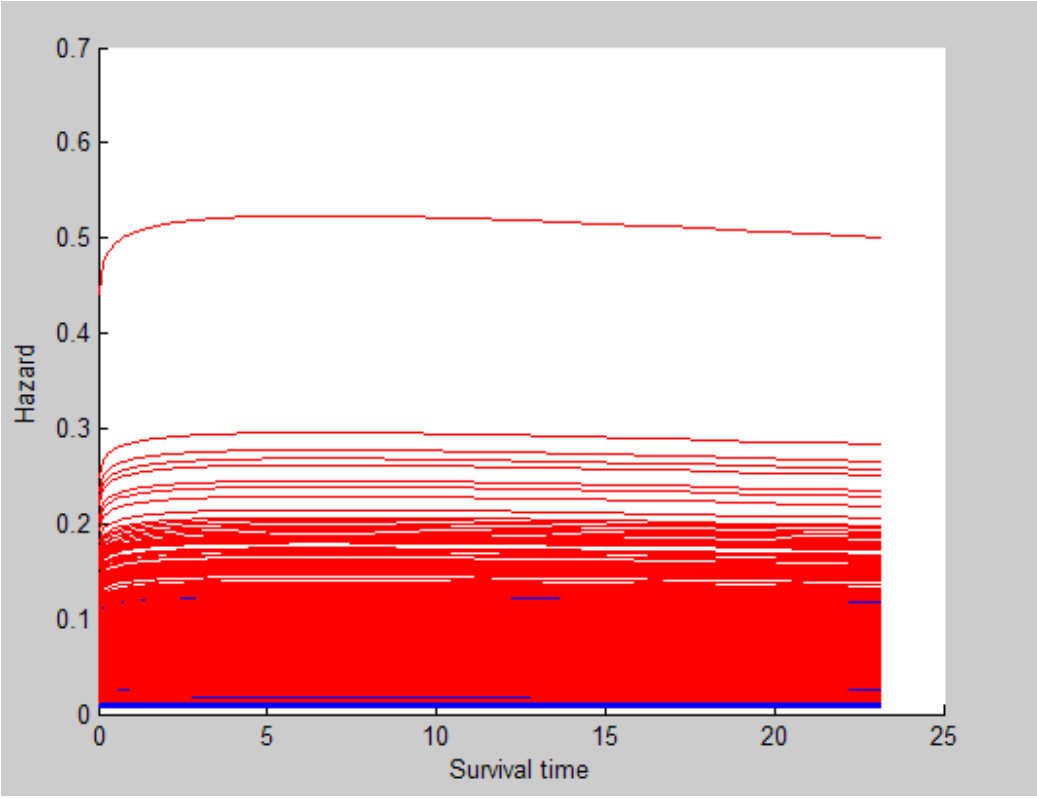


Figure 20. Hazard functions of 743 patients obtained from the Cox model. Red plots are for patients with epithelioid melanoma and blue for patients without it.

In order to obtain the survival plot from the Cox model, the following transformations are performed:

$$h(t, z_i) = h_0(t) \cdot \exp(z_i^T \beta)$$

$$H(t, z_i) = \int_0^t h(t, z_i) \Rightarrow H(t, z_i) = \int_0^t h_0(t) \cdot \exp(z_i^T \beta)$$

$$H(t, z_i) = \exp(z_i^T \beta) \cdot \int_0^t h_0(t) \Rightarrow H(t, z_i) = H_0(t) \cdot \exp(z_i^T \beta)$$

$$S(t, z_i) = \exp[-H(t, z_i)] \Rightarrow S(t, z_i) = \exp[-H_0(t) \cdot \exp(z_i^T \beta)]$$

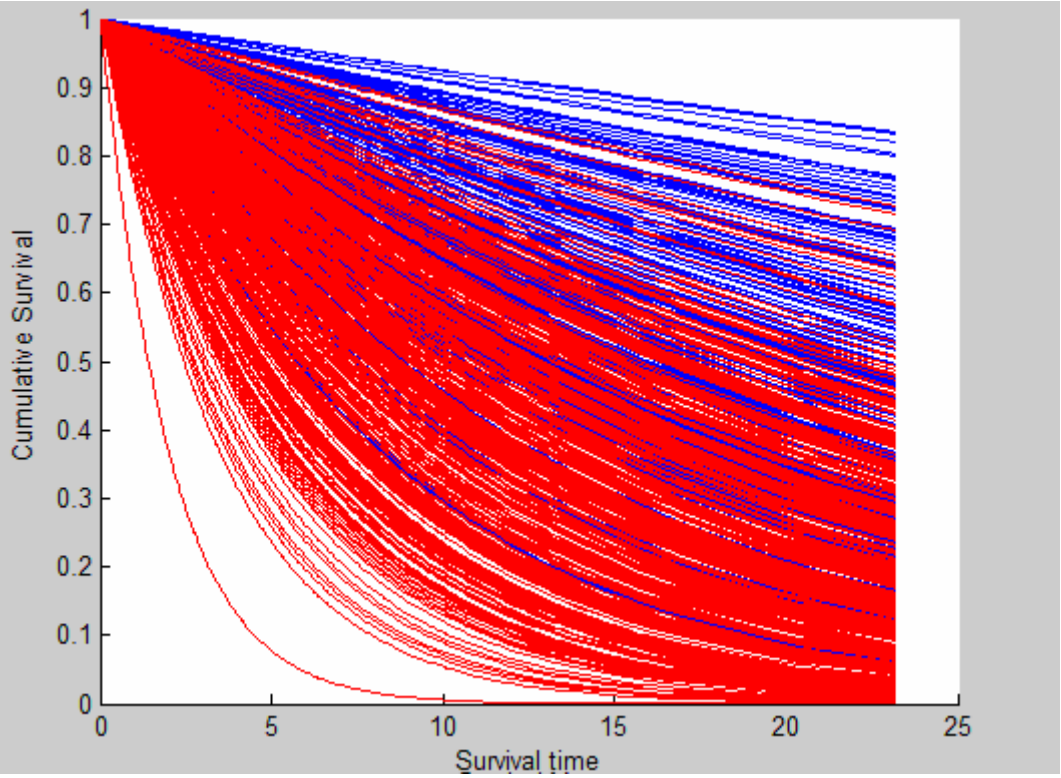


Figure 21. Survival functions for 743 patients from the Cox model. Red plots present patients with epithelioid melanoma and who have less average survivability than patients without it who appear in blue.

To make clearer the influence of epithelioid melanoma on ocular melanoma patients, the original dataset is divided into two sets (with the epi variable) and hazard functions from the Cox model are plotted for each group in their mean covariate value.

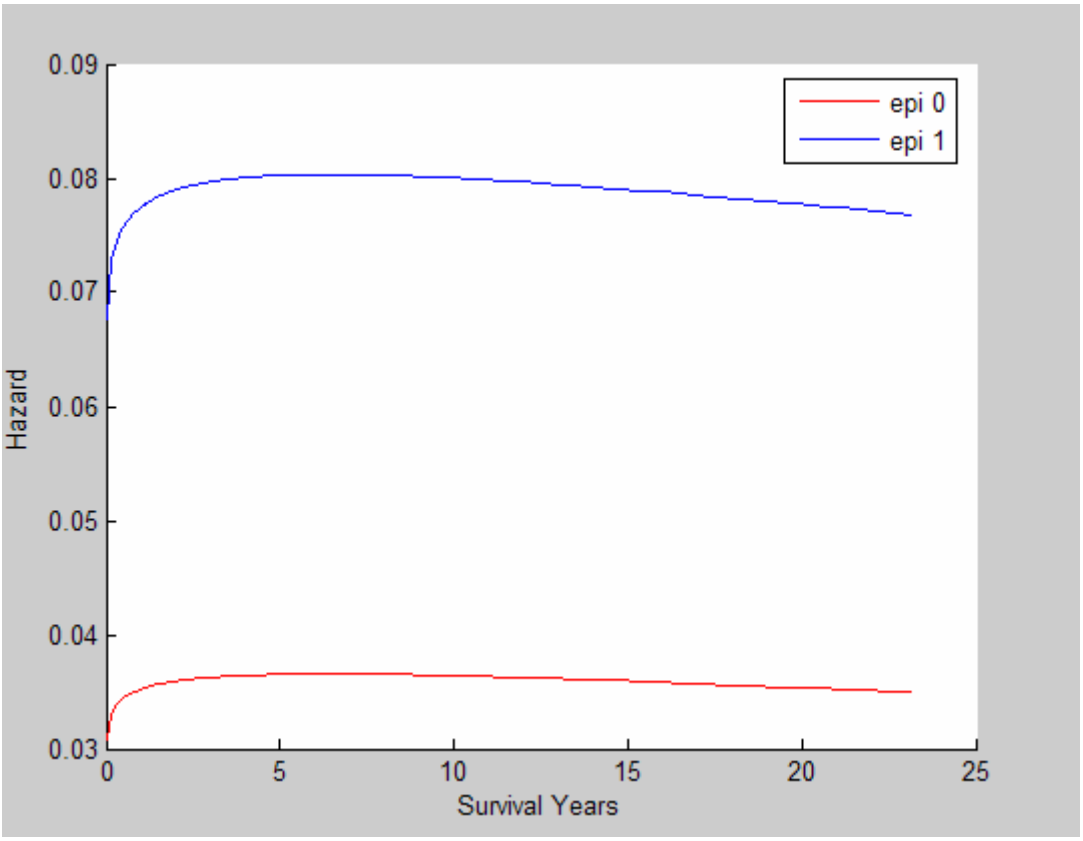
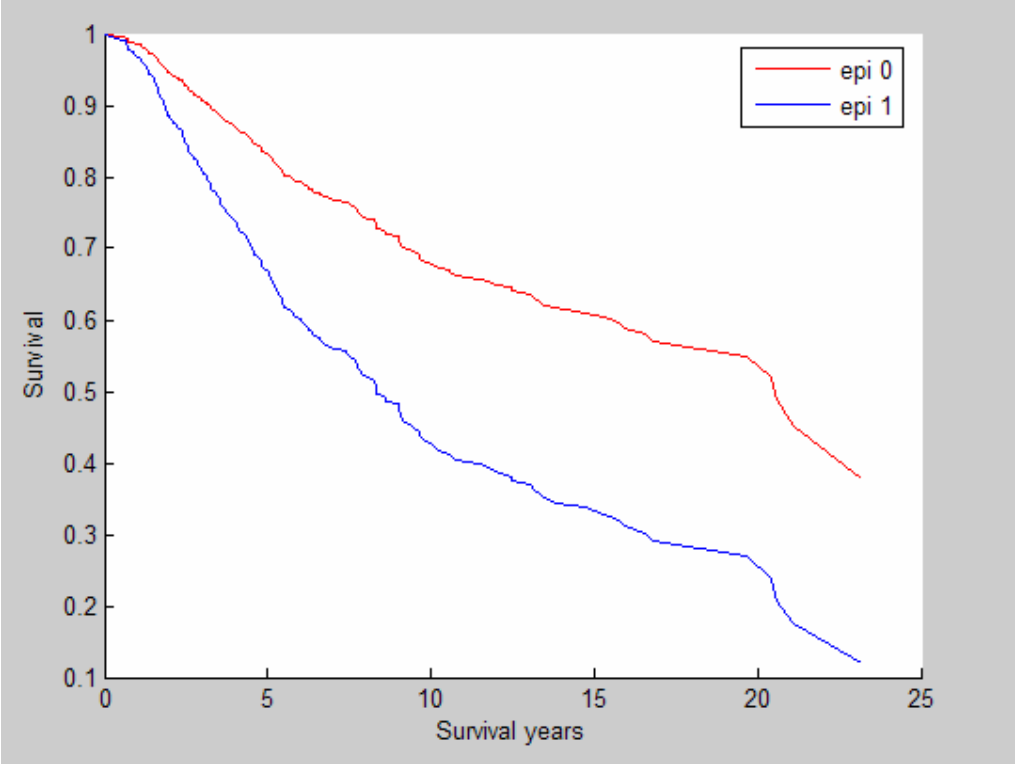


Figure 22. Hazard functions at the mean covariate values of patients with epithelioid melanoma (blue) and patients without it (red). Patients with epi =1 have higher hazard values.

Figure 22 shows that patients with epithelioid cellularity (epi=1, with blue) have higher hazard values than patients with without. This conclusion totally agrees with the one made from the regression results. In addition, figure 23 shows that patients with epi =1 have less survivability than patients without it.

a)



b)

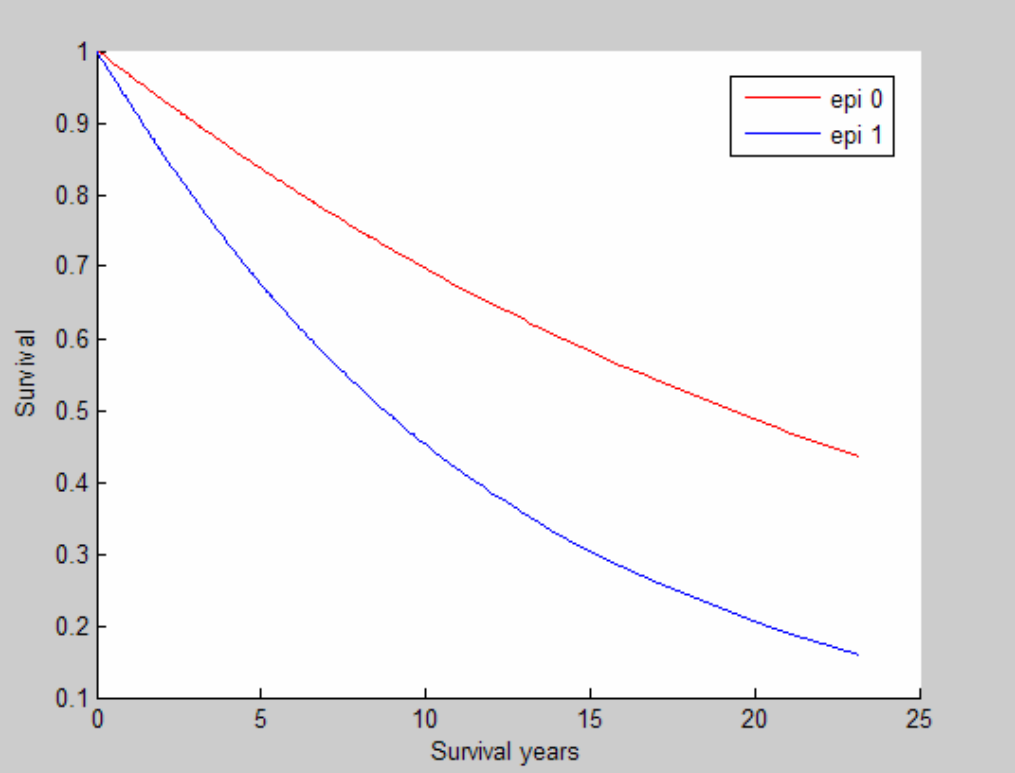


Figure 23. **a)** Survival functions of two groups at their mean covariates values from the Cox model. Blue plot is for patients with epithelioid cellular and red for patients with epi =0. **b)** The same plots fitted in distributions

Following the same steps, the survival plots for patients with high and low ludb can be obtained by dividing the dataset into two groups. The median value of the ludb covariate is 14.9. The plots of 743 patients are shown in figure 24.

The same conclusion can be made about the ultrasound height, the third covariate under examination. The median value is 10. In figure 25, red is for patients with uh larger than 10 and blue lower than 10.

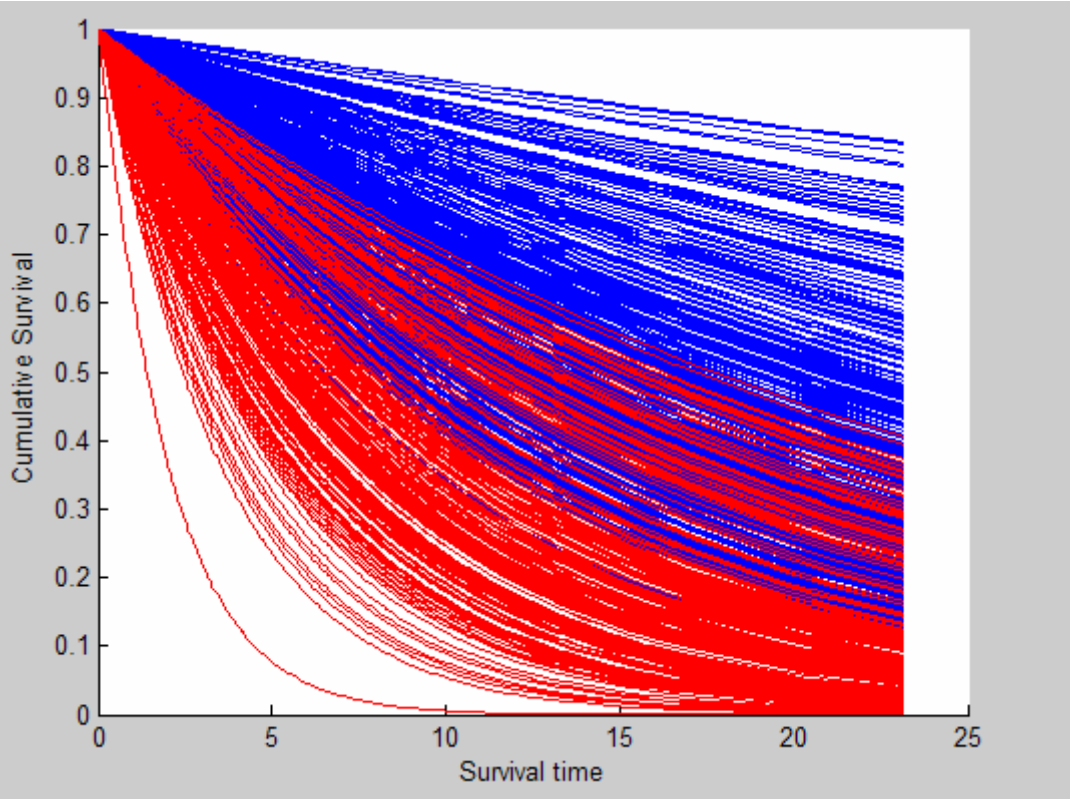


Figure 24. Survival curves of 743 ocular melanoma patients from the Cox model. Red is for ludb > 14.9 and blue for ludb lower than the median value. It is obvious that patients with large values of ludb have less survivability.

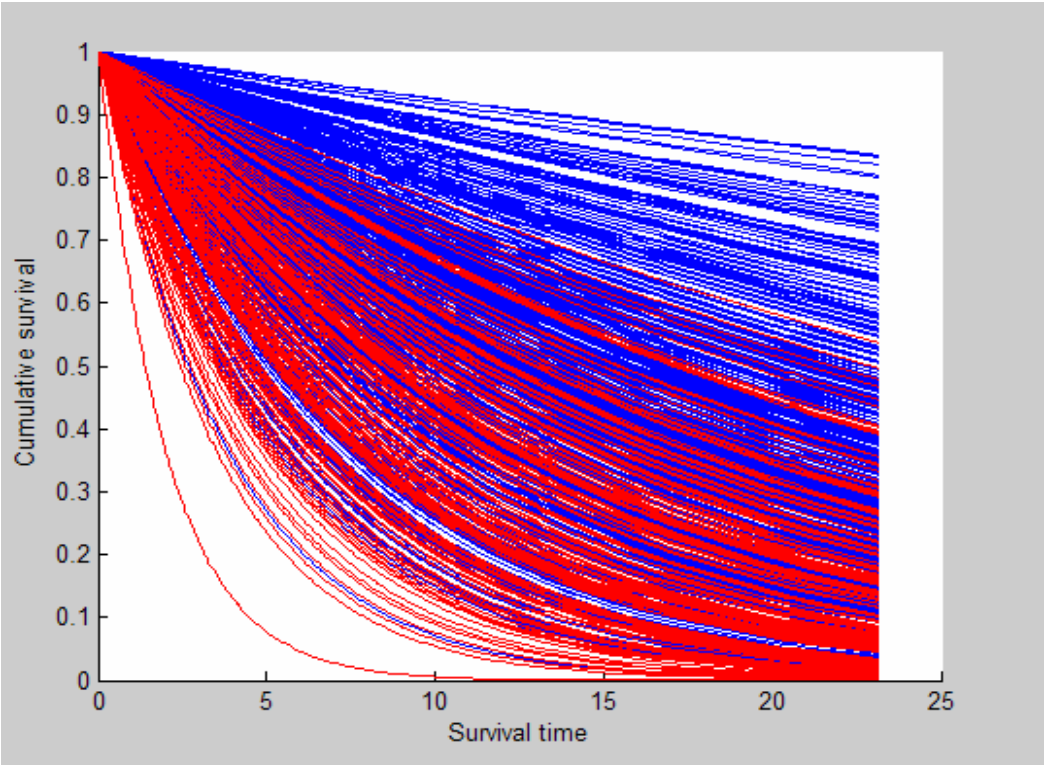


Figure 25. Survival curves for patients with uh< 10 (blue) and uh>10 (red)

Figure 24 shows that atients with ludb higher than 14.9 have 20% probability of surviving after 23 years. At the same time patients with low ludb have 50% probability of surviving. In addition in figure 25 it is obvious that patients with large uh (>=10) have around 20% probability of surviving after 23 years, while patients with low uh 40%.

Figure 26 proves that epithelioid cellularity affects the disease more than the longest ultrasound dimension (ludb). For this reason, patients with large ludb (bigger than the mean value 14.9) and epi 0 or 1 are plotted. The conclusion is that even if a patient has large ultrasound basal dimension he will be less likely to die if he has epi=0. In other words, when epithelioid cellularity appears, the hazard increases despite the values of the rest of the covariates.

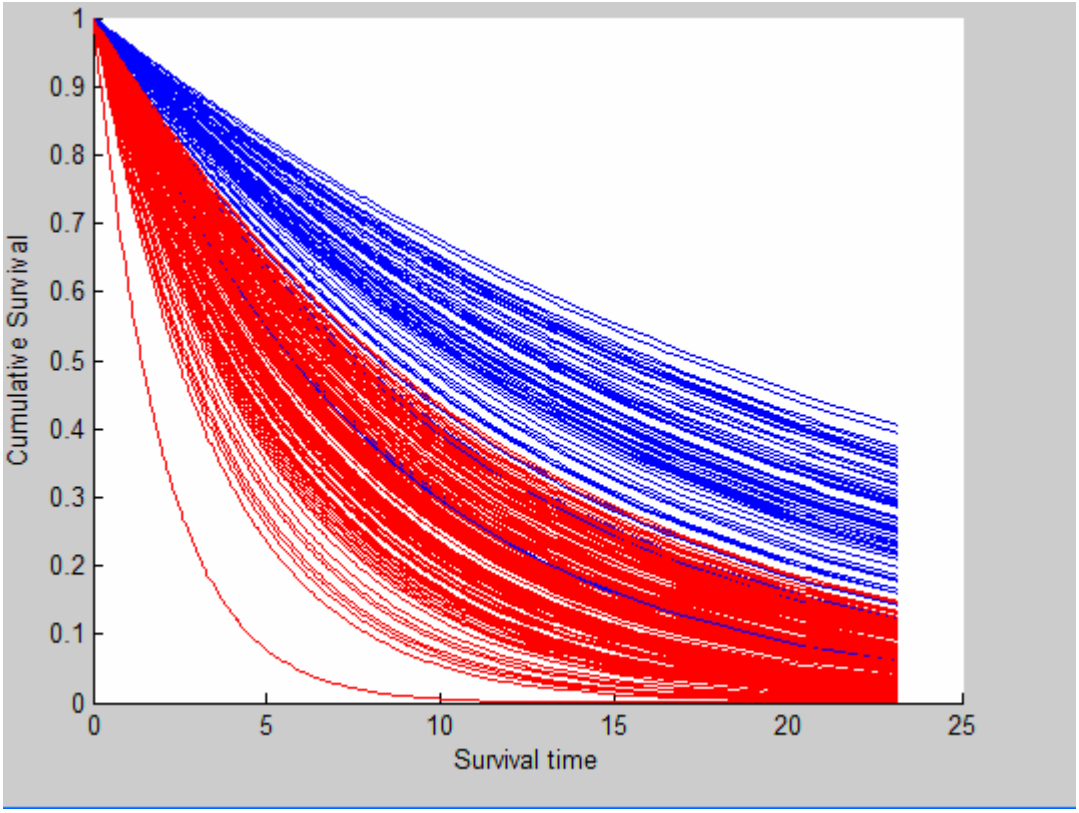


Figure 26. Survival curves from the Cox model for two sets of patients. Red is for those who have large ludb and epithelioid melanoma and blue for large lubd and no epithelioid cellular. It is clear that even if a patient has large value in the ludb covariate, it is still the epi factor that affects survivability.

In figure 26 the red curves are the survival plots of patients with epithelioid cellularity and ludb higher that 14.9. The blue curves are patients with epi= 0 and high ludb. After 23 years, patients who have epithelioid cellularity and ludb bigger that 14.9, have a median probability of 10% to survive. On the contrary, patients who do not appear to have epithelioid cellularity and ludb bigger that 14.9, have median probability of 35% to survive.

5.3 *Fully parametric*: The log logistic regression model

As explained in Chapter 3, the log logistic function suitability should be checked. There is a linear representation of this function, beginning with the cumulative density function and ending with the log odds:

$$F(t) = 1 - \frac{1}{1 + a \cdot t^{\gamma}} \Rightarrow F(t) = \frac{1 + a \cdot t^{\gamma} - 1}{1 + a \cdot t^{\gamma}} \Rightarrow F(t) = \frac{a \cdot t^{\gamma}}{1 + a \cdot t^{\gamma}} \Rightarrow$$

$$\Rightarrow F(t) = a \cdot t^{\gamma} \cdot S(t) \Rightarrow \frac{F(t)}{S(t)} = a \cdot t^{\gamma} \Rightarrow \frac{F(t)}{1 - F(t)} = a \cdot t^{\gamma} \Rightarrow$$

$$\Rightarrow \log[\frac{F(t)}{1 - F(t)}] = \log[a \cdot t^{\gamma}] \Rightarrow$$

$$\Rightarrow \log[\frac{F(t)}{1 - F(t)}] = \log(\alpha) + \gamma \log(t)$$

The log odds can provide a linear mathematical representation of the cumulative density function versus log t as it is shown in the equation above. If a plot of the log odds ($\log[\frac{F(t)}{1 - F(t)}]$) versus the logarithm of time is approximately linear, then the log logistic distribution is appropriate to describe the survivability of the specific dataset. The cumulative density function is approximated non parametrically, and therefore the Kaplan Meier estimator is used to obtain an estimate of S (t). Then, using the formula F (t) = 1- S (t) the cumulative density function derives and the log odds can be obtained. The estimator was created in MATLAB environment (available in APPENDIX A) and the log odds were plotted against the logarithmic time.

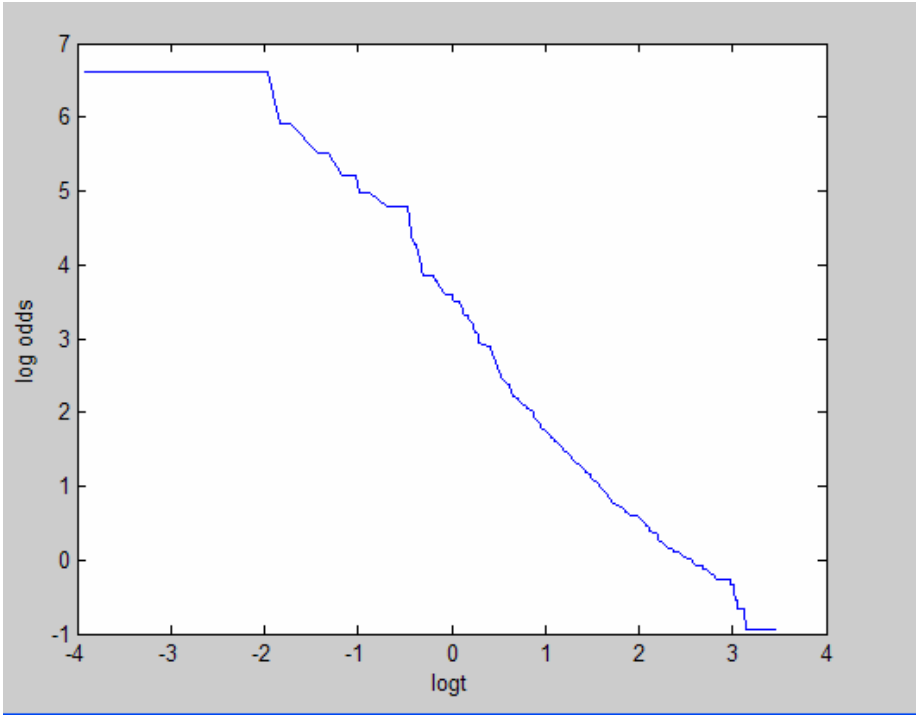


Figure 27. A plot of the log odds versus the logarithm of time, showing approximately a straight line. The odds are obtained in a non parametric way (product limit estimate).

Figure 27 proves that a plot of the log odds versus logT is approximately a straight line (when fitting it in a linear distribution the goodness of fit is accepted). Therefore, the log logistic distribution is suitable for our data. The slope can be calculated to be $\gamma = 1.1054$ and the parameter $\alpha = 0.060904$. With these parameters the survival function is defined and so is the density function.

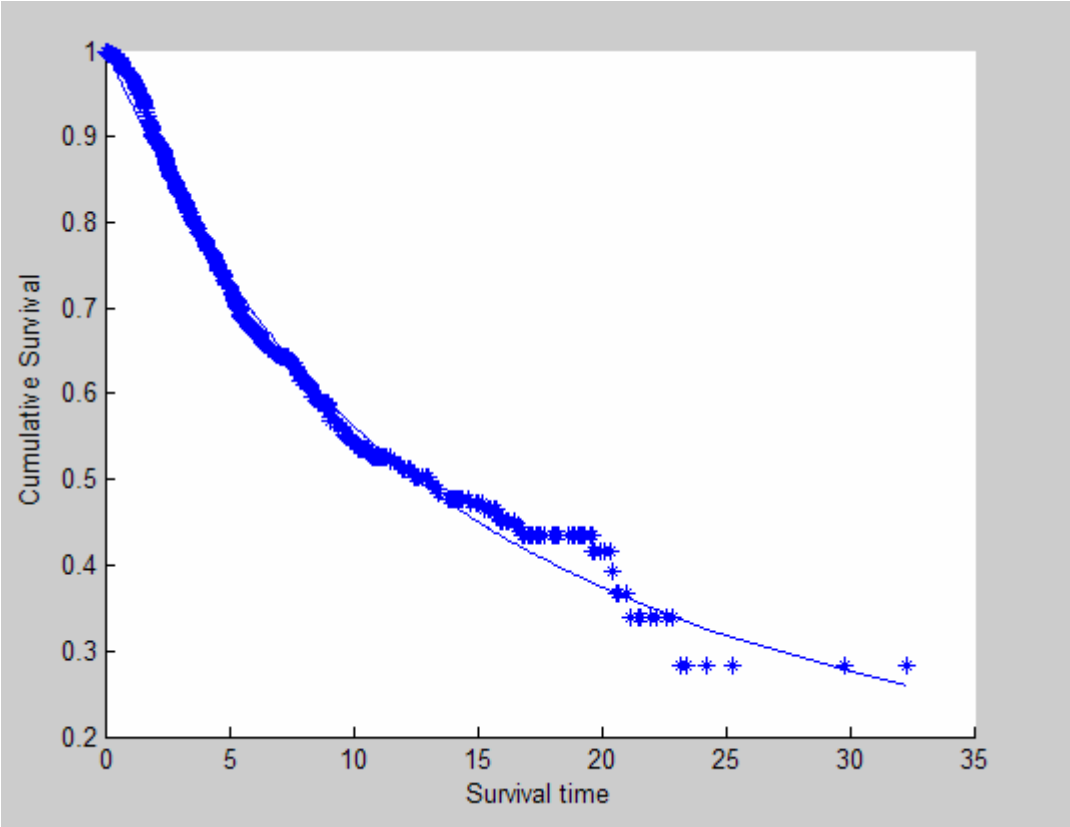


Figure 28. The parameters α and γ obtained from figure 27 are used to plot the survival function and fit it to the Kaplan Meier estimates.

In the log logistic regression model the hazard function comes from the formula:

$$h(t,z) = \frac{\alpha\gamma t^{\gamma-1}\exp(z^T\beta)}{1+\alpha t^{\gamma}\exp(z^T\beta)}$$

The regression process to obtain the regression parameter β follows a specific methodology which was described in chapter 3. It is important to notice that this procedure has to be performed for a homogeneous population. In other words, patients with the same pattern of covariates will be plotted together. If the two patterns are different in one covariate only say, z_1 all the others be the same, the distance between the two lines gives a rough estimate of β_1. Two covariates will be used for the log logistic regression model, the epithelioid cellularity and the ultrasound height. The population was divided into to two sets of patients, to those who had epithelioid melanoma (epi=1) and ultrasound height bigger than 10 and those who have epithelioid melanoma and uh<10.

The separation was performed with the help of MATLAB environment. After the classifications two homogeneous populations derived. The characteristics of those patients where applied in the Kaplan Meier estimator, then used to obtain the cumulative density estimates (F (t) = 1- S (t)) and the log odd ($\log[\frac{F(t)}{1-F(t)}]$) results were plotted against the log t.

If the plots of the two groups show departure from linearity another model from the class of proportional odds models could be used. If linearity is satisfied but the distance between the lines is not constant a model assuming proportional odds model is not appropriate.

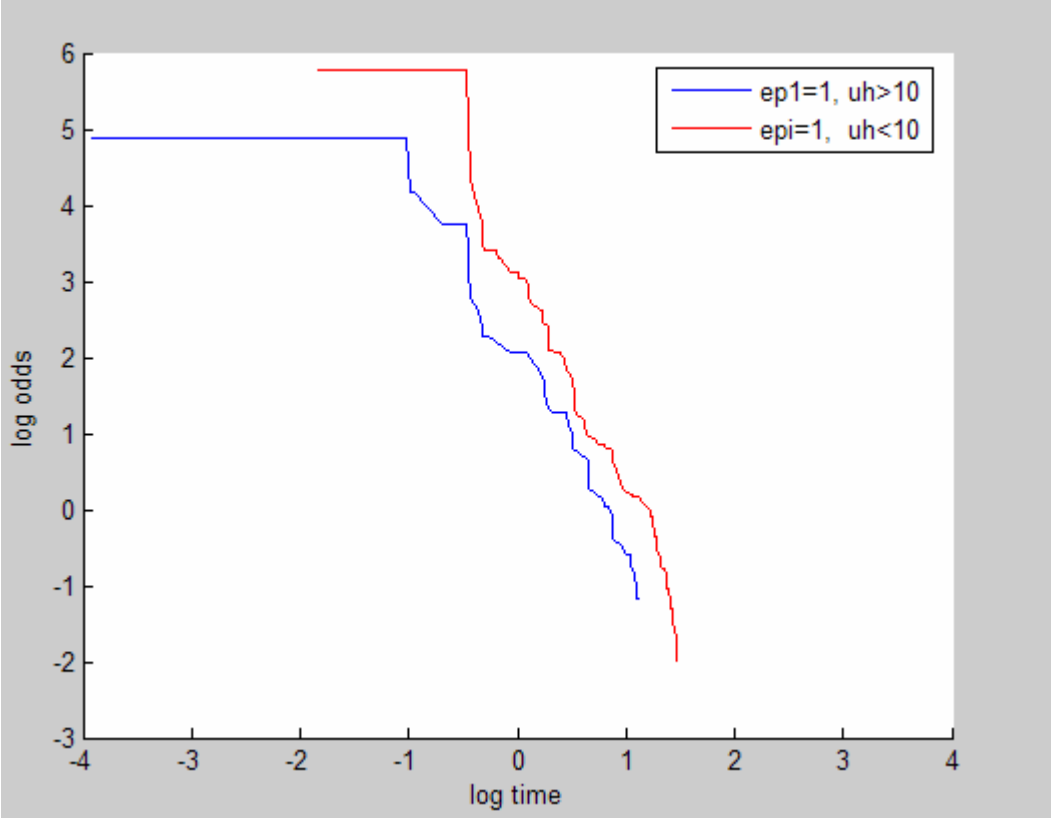


Figure 29. The log odds from the two populations are plotted against the logarithm of time.

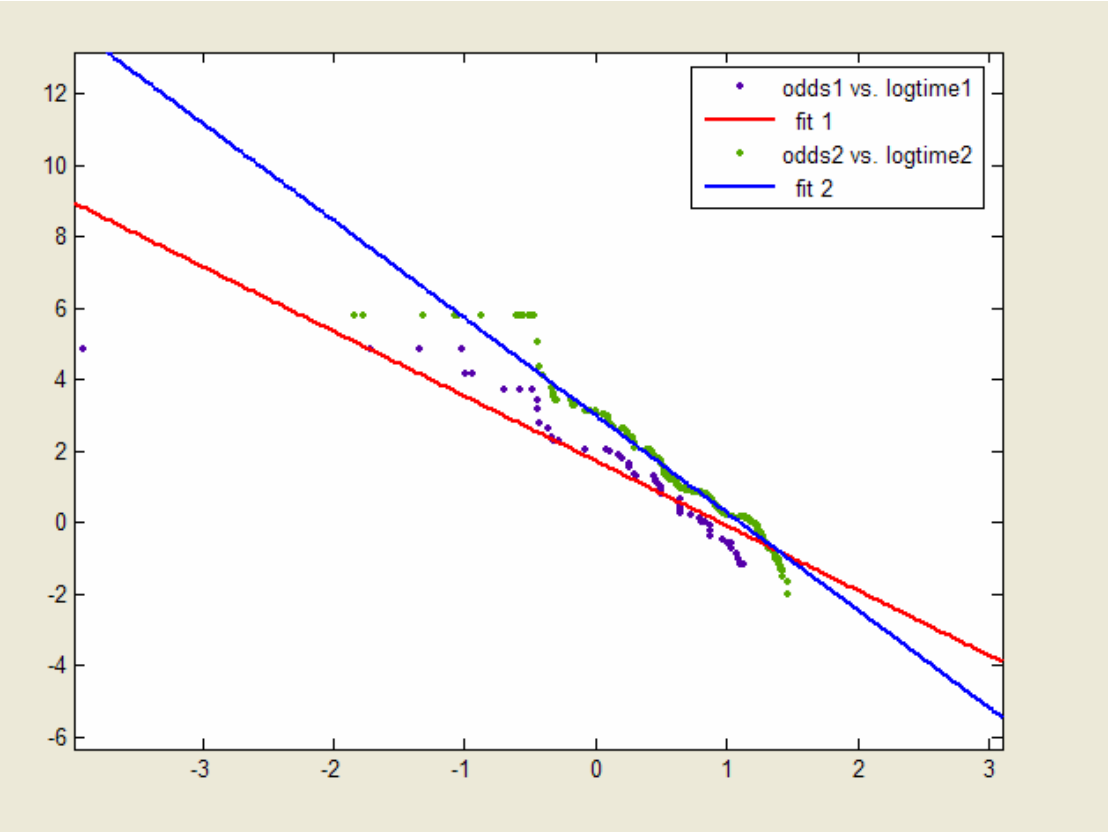


Figure 30. The curve fitting tool of MATLAB environment is used to fit the two plots from figure 29 in order to check linearity better.

The two plots have acceptable linear fits, but they don't have constant distance. Therefore the log logistic regression model cannot be applied to this specific dataset, since it is not possible to count the regression parameter β.

5.4 Neural Networks results

The application of neural networks on specific dataset was introduced widely in chapter 4. The Cox proportional hazards model is implemented with the use of the Backpropagation multilayer neural networks. Two different nets are modeled, one to simulate the baseline hazard function and another one to simulate the covariates (the exponential part of the model) Taking into consideration the general formula of the Cox model the two networks are:

$$h(t, z_i) = h_0(t) \cdot \exp(z_i^T \beta)$$

Baseline network

Covariates network

A code in the MATLAB environment is created to unify the two networks in one in order for the implementation to be friendlier to user.

The baseline function approximation was performed in this chapter in two different ways, with the Breslow and the extended Kaplan Meier estimator. Since these two are mathematically close, the Breslow estimator is used in the network.

The baseline network has two inputs, the number of deaths ant each time and the sum of exponential covariates for the risk patients ($\exp(z_l^T \beta)$) at the same time. The output of the network is obviously one, the values of the cumulative hazard of the baseline hazard function. The input vector has length 240, since there are 240 discrete times when deaths take place among the 743 patients. The output is of the same length. In addition, 6 hidden layers were applied for the outcome to be reasonable since the input is vector is large. The first layer has 2 nodes, the next two hidden layers have 5 nodes each, the fourth 20 and the following two 5 nodes each. The transfer function applied to the input and the 5 hidden layers is the tansig while the output layer has a purelin transfer function from the MATLAB environment. The newff function of MATLAB created a feed forward Backpropagation net and the trainlm training function was applied for fast training. The main idea is that in the future one could simulate this network by supplying the inputs (number of deaths and the sum exponential for patients at risk) to easily obtain the values of the cumulative baseline hazard function.

The input dataset is divided into two sets, one for simulating and one for training The training set and the simulation set must have same deviations and since the original set was divided with a rule that supports both big and small values in the two sets. This is more obvious in the source code which provided in APPENDIX A.

Finally, in order to check the simulation results the postreg function of MATLAB environment is applied. The network training set is post processed by performing a linear regression between one element of the network response and the corresponding target. This way the correlation is checked. This function produces a plot where all outputs are plotted, and were a best possible linear fit is applied. Then this fit is compared to the ideal correspondence of network outputs and targets and the correlation is calculated. Correlation is normalized measure of linear relationship strength between variables. The closer the correlation coefficient to one, the better the correspondence of outputs and targets. The correlation for this network is 0.999:

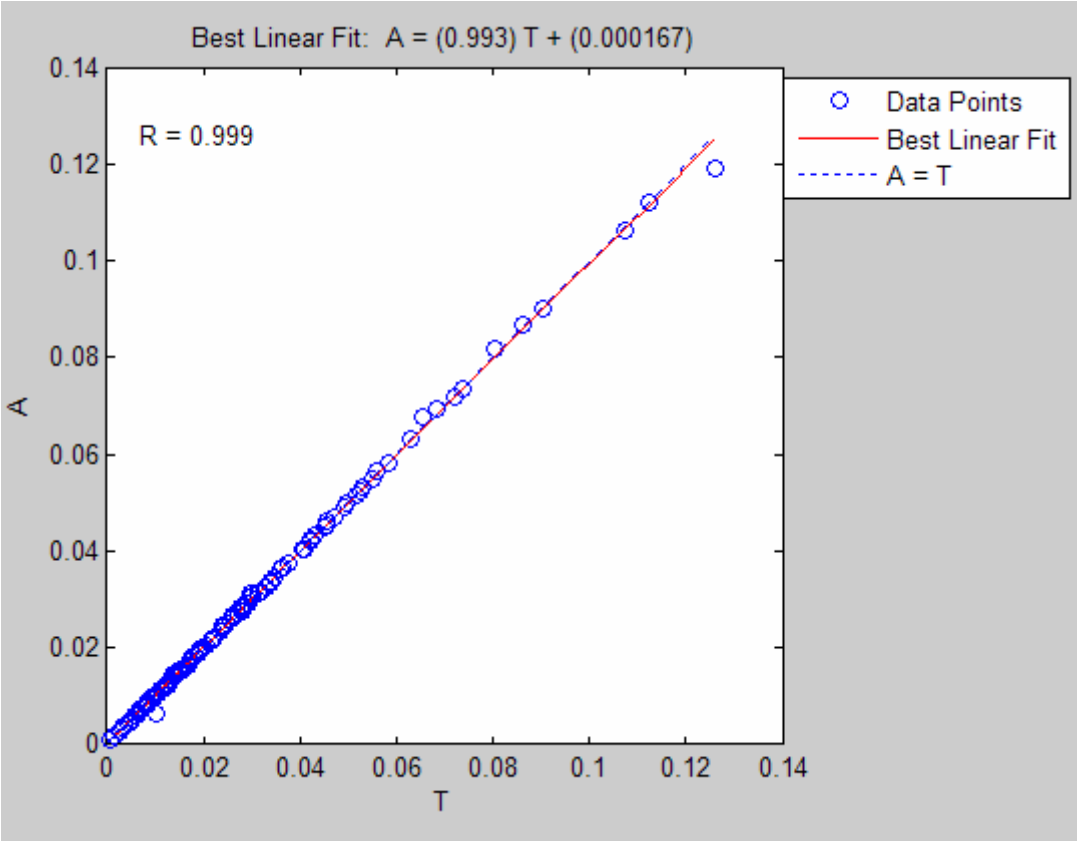


Figure 31. Correlation of network's outputs and targets of the baseline neural Network.

The correlation in figure 31 proves that the network output is the desirable as the output values are very close to their targets and the performance is acceptable.

The Covariates Network is referred to the exponential part of the Cox proportional hazards model. The main idea is to create a network that provides the values of the exponential part for each patient. This network is actually performing a regression since the regression coefficient can be easily obtained from the network's output. In the future any one would be able to obtain the regression parameter by entering the covariate values.

The inputs of the neural network are three. These will be the three covariates under examination, longest ultrasound basal dimension, ultrasound height and the epithelioid cellularity. The input values can either be numerical or binary. The regression parameter is not an input to the neural network so it won't be necessary in later studies to approximate it in order to simulate the network. The output of the network is the value of the exponential part for each patient.

In order to achieve a good performance our data is divided into three sets that will be used for training, testing and validation of the network correspondingly. The Backpropagation Neural Net was chosen as it is one of the most powerful neural net types. Generally this network has the same structure as the Multi Layer Perceptron and it uses the Backpropagation learning algorithm.

Using the newff MATLAB function, the network's number of hidden layers and their corresponding transfer functions can be selected. There is no rule that can guide this selection but many alternatives have been studied and checked their results. In conclusion, the network would be more properly trained if it had one hidden layer with 5 neurons, 1 input layer and (one neuron) and one output layer (one neuron).

Finally, in order to check the simulation results the postreg function of MATLAB environment was applied again. The network training set is post processed by performing a linear regression between one element of the network response and the corresponding target. This way the correlation is checked. The correlation for this network is 0.999

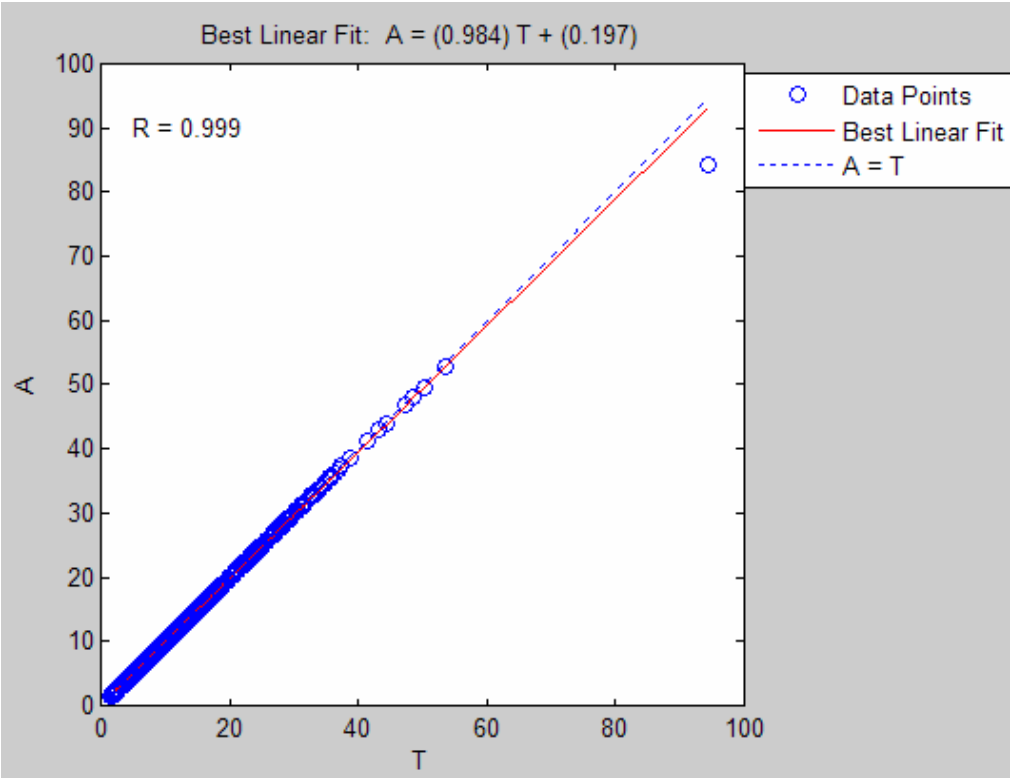


Figure 32. Correlation of network's outputs and targets for the covariates network.

Chapter 6
Comparisons with relevant studies

In the beginning of the chapter 6 some comparisons are made along the models implemented in this study. In particular, The Kaplan Meier results are compared to the Cox model results. Then, there are important comparisons of our results with the results of other studies (that are introduced here) even if they come from a different dataset.

6.1 Comparisons between the methods implemented in this project

The Kaplan Meier estimator is a non parametric method in survival analysis. It has been widely analyzed and implemented in chapters 3 and 5. With this estimator a survival curve for the whole population of the study derived. This curve is very important in medical statistics as it is widely used in prognosis.

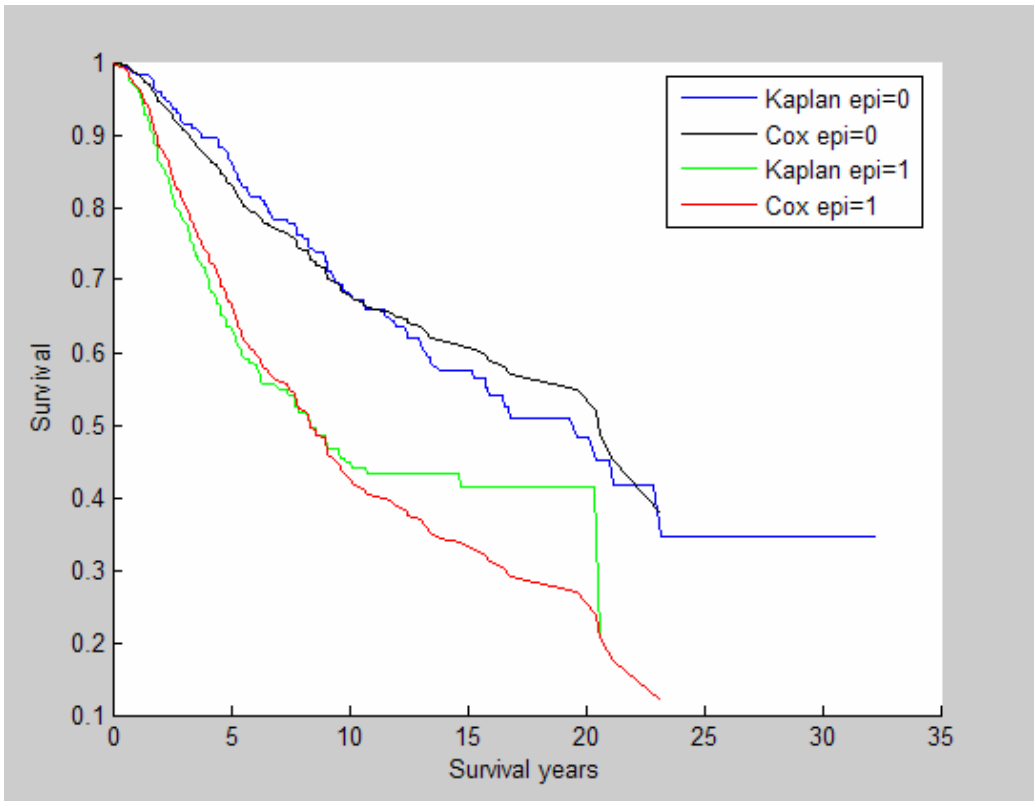
In addition, in chapter 5, the original population is divided into two sets. The first set includes patient that appear to be influenced from epithelioid melanoma and the second one involves patients that are do not have epithelioid cellular. The distinction is performed according to a variable called epi, which is an indicator of epithelioid cellular. Therefore, the first set has epi=1 and the second 0.

There are 279 patients with epithelioid melanoma and 464 without it. Survival curves are plotted for the two populations, using the Kaplan Meier estimator. The results proved that patients with epi =1 have less survival probability that patients without it at any time.

In addition, in chapters 3 and 5 the Cox proportional hazards model was introduced. The influence of three covariates is examined to check whether they are prognostic factors. All three of them (epi, ludb and uh) affect the ocular melanoma disease in a way that the bigger their values the worse the survival probability and the greater the hazard. The covariate that affects more the disease is the epithelioid melanoma.

Again, the population is divided into two sets in respect with the epi variable. Since the Cox model offers only curves for each patient and not for the whole population, the average values of the covariates of the two populations are computed. These mean values are used in the Cox model to obtain hazard and eventually survival functions for the two populations.

a)



b)

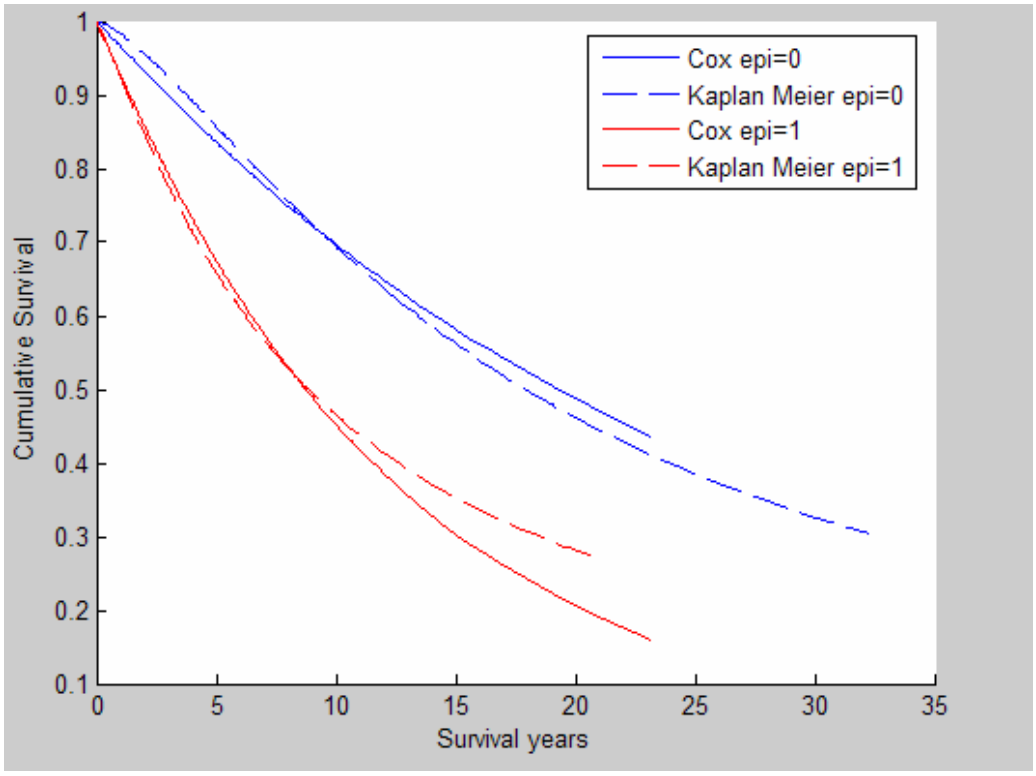


Figure 33. a) Survival plots for two populations (epi=1 and epi=0) with the Cox model and the Kaplan Meier estimator. **b)** The same plots fitted to distributions.

Figure 33 shows the survival curves for the two groups of patients as they derived from the non parametric Kaplan Meier and the Semi parametric Cox model. Red lines are for patients with epithelioid melanoma. The two models are close to each other. The Kaplan Meier estimator is performed for every single time spot of the dataset no matter of it is a death or censoring time. The epi= 0 population is smaller in size than the epi =1 and this is the reason why the red dashed line ends up earlier than the blue one.

On the contrary, in the Cox model there is baseline hazard function that is the same for all patients. This function is approximated from the Breslow estimator which uses only the death times. Since the baseline hazard function is the same for all patients, the baseline survival function is too. This is the reason why in figure 33 the Cox model survival curves have the same form.

An important observation from figure 33 is that the Cox model curve for epi=1, has lower values than the Kaplan Meier for the same population. One could detect here the effect of the epi covariate since it has been proven to be a prognostic variable. A curve that is covariate dependent is expected to show less survival probability than a covariate independent curve. The same conclusion can be made for the epi=0 population .

6.2 Kaplan Meier estimator for the original dataset.

In chapter 5, in order to implement the Cox proportional hazards model, a model where covariates are included to check their effectiveness, the original dataset of the ocular melanoma disease had to be reduced to a subset. In this subset, all the data that contained missing values in the covariates field were removed. As a result the new subset contained 743 elements while the original dataset included 1743 patients.

At this point, it is useful to prove that no information was lost due to this reduction if we want ti generalize the results for the ocular melanoma disease. In order to come up with this conclusion the results from another study are used. The survival curve for the original population was performed by Kourkouta Anna-Maria [45].A comparison of the survival curves from both the original and the reduced dataset could lead to such a conclusion. For the two curves to be analogue the same estimator had to be used to obtain their values.

The Kaplan Meier estimator has been implemented in both cases. The same assumptions had to be made for the censoring type and therefore, in both datasets, a patient is considered to be censored if the death indicator is zero.

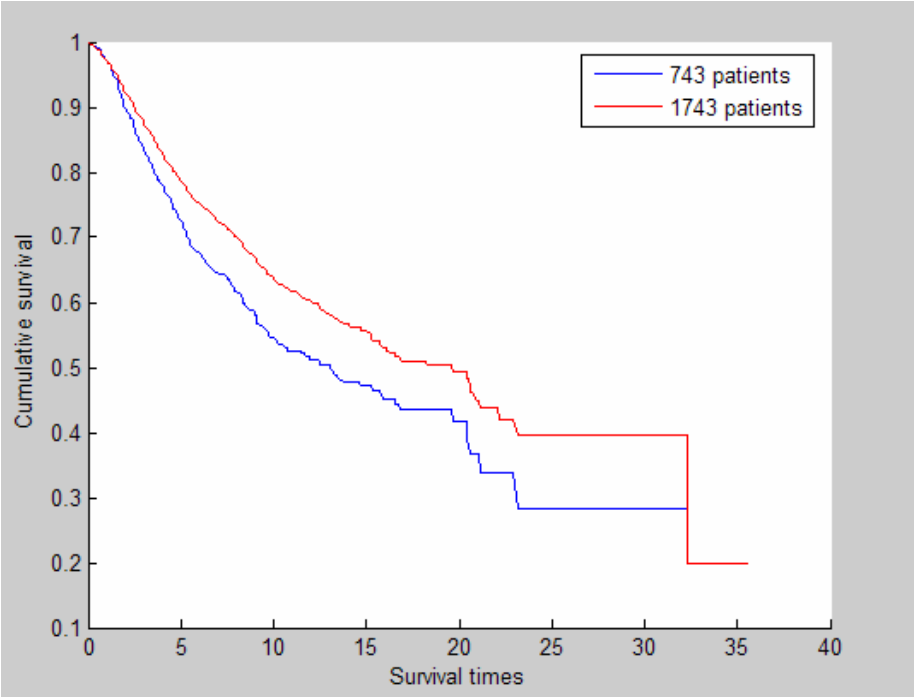


Figure 34. Survival curves form the Kaplan Meier estimator. The red curve is for the original dataset while the blue one is for the reduced dataset.

The two survival curves are acceptably close to each other. In addition they seem to have the same shape. If the reduction had altered the shape of the survival curve, no generalizations could have been made about the ocular melanoma disease concerning the prognostic factors.

6.3 Regression comparison with the Aalen‘s additive model.

The Cox proportional hazards model, which was implemented in chapter 5, and the Aalen’s additive model include their implementation covariates. These covariates affect the survivability and both of the above models have as purpose to show how much and in what way they do it. The Aalen’s additive model that was performed by Kourkouta Anna- Maria includes in its implementation the same dataset of the 743 patient as the one which is included in the Cox model.

The additive hazard model was suggested for the influence of the covariates on the hazard function. This method results in plots that may give information on the change over time in the influence of covariates. It is an alternative to the Cox model which does not condition on constant proportional hazard. In this model the covariates are modeled as additive risks to a baseline hazard.

The basic equation may be formulated as follows:

$$h(t,Z) = \beta_0(t) + \sum_{j=1}^{k} \beta_j(t) Z_j(t)$$

The hazard at any time is thus a sum of a baseline hazard β_0 (t) and a linear combination of the covariate values, Z_j. In the following equations n is the number of subjects and r is the number of covariates.

So, the vector of intensities $h_i(t), i = 1,2,....,r$ is formulated by the linear model: $h(t) = Y(t)\beta(t)$ .The matrix Y is of size $n \times (r+1)$ and is constructed as follows:

- If the ith individual is a member of the risk set at time t then the ith row of **Y** (t) is the vector $Z^i(t) = (1, Z_1^i(t), Z_2^i(t), ...Z_k^i(t))$, where $Z_j^i(t), j = 1,...r$ are the covariate values.
- If the individual is not at risk at time t, meaning that the event of interest has already occurred or the individual has been censored, then the corresponding row in Y (t) contains only zeros.

The first element of the vector **β** (t) is a baseline parameter and the remaining elements $\beta_i(t), i = 1,2,....,r$ are called regression functions and estimate the influence of the covariates. These regression functions are the equivalents to the regression parameters in the Cox regression model. But in contrast to the Cox model, where the regression parameters are constant, the regression functions may vary with time.

Since the regression functions may vary with time, statistical analysis of them may reveal changes in the influence of the covariates over time. This is one of the main advantages of this method.

It is unpractical and difficult to estimate the individual regression functions and instead the cumulative regression functions are estimated. The elements $\mathrm{B}_j(t), j = 0,1,....,k$ of the column vector **B** (t) are the cumulative regression functions and are defined as: $\mathrm{B}_j(t) = \int_0^t \beta_j(t) ds$

The cumulative regression functions are plotted against time and give a description of how the covariates influence the survival over time. It is therefore the change in the cumulative functions, the slope that is of primary interest.

If $T_1 < T_2 < \ldots$ are the ordered event times, meaning the times when an actual event, not censoring, occurs, an estimator of **B** (t) is given by:

$$B^*(t) = \sum_{Tk \le t} X(T_k) I_k$$

where $\mathbf{I}_k$ is a vector of zeros except for a one corresponding to the individual experiencing an event at time $\mathbf{T}_k$. **X** (t) is a generalized inverse of **Y** (t):

$$X(t) = [Y(t)' Y(t)]^{-1} Y(t)' \text{ . [12]}$$

Figure 35 shows the distribution of the covariates that were used in the analysis of the additive model. The brown fit-line represents the epithelioid cellularity (epi), the red one the longest ultrasound basal dimension (lubd), and lastly the blue line represents the ultrasound height (uh). As it can be seen, the covariate that affects more the disease id the epi, since it has the greater slope. The covariate lubd affects the disease less and lastly the covariate uh has almost no affect in the disease.[]

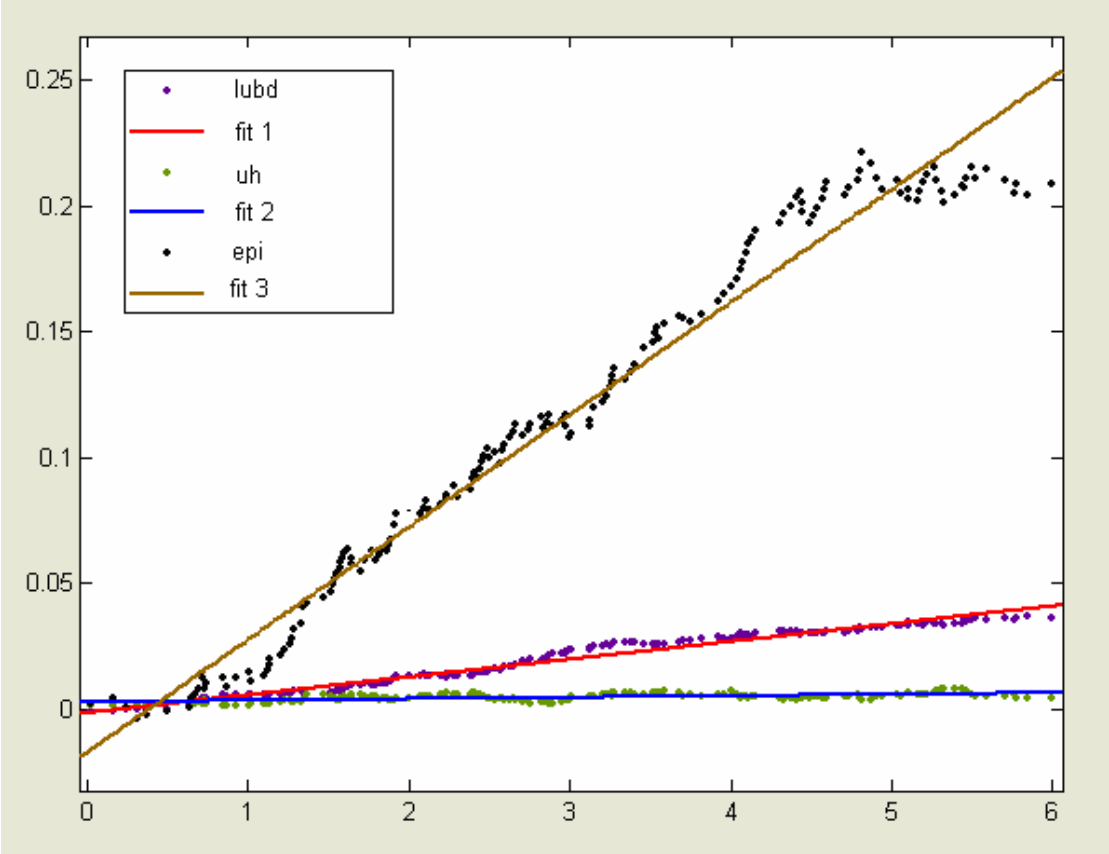


Figure 35. The regression functions obtained form Aalen's model, fitted to linear distributions to show their rank.

The results of the Cox proportional Hazards model are analogue to those from the additive model. Through the regression (maximum likelihood estimation) the regression parameters were calculated to be:

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0.127 \\ 0.033 \\ 0.664 \end{pmatrix}$$ These coefficients indicate the magnitude of the effects of their corresponding covariates. The interpretation of β's sign is:

- If β=0 the covariate has no effect on survival
- If β<0 the covariate affects the survival inversely. This means that the higher the value of the examined covariate the lower the hazard.
- If β>0 the covariate affects the survival. A high value of the covariate would mean high hazard.

Therefore, the factor that affects the disease more is the presence of epithelioid cellular. In addition the lubd affects more than uh. The same results were obtained from analysis of the additive model. The only difference is that the regression parameters are not counted numerically but represented graphically.

Therefore, even though different models were applied to the same dataset and even though each one had each own assumptions, the results are the same. Furthermore this conclusion has real medical interest for both the doctors and the patients since for now one a patient that appears to have epithelioid cellular will have less probability to survive.

6.4 Comparison with relevant study using artificial neural networks and Bayes theorem.

The use of neural networks in survival analysis has been evaluated in earlier chapters. Their ability to generalize made them a powerful tool in clinical survival analysis. There have been various methodologies of neural networks applications.

A relevant study with this thesis was performed by Taktak, Fisher and Damato [42] in 2004. They modeled the survival after treatment of intraocular melanoma patients. This specific study introduces a new idea of how the neural networks can be combined with Bayes theorem. It describes the development of an artificial intelligence system for survival prediction from intraocular melanoma. The network's targets are compared to the Kaplan Meier estimator and the Cox model.

The database originates from patients treated in Glasgow and Liverpool between 1969 and 2001. The measurements that are analyzed to check their influence on the disease are the coronal and sagittal tumor location , anterior tumor margin largest basal tumor diameter and cell type. These covariates along with time consist the inputs of the neural network. The output of the system is the survival probability. The issue is to examine tumor specific survival, that is how the survival is affected by certain characteristics of the tumor.

The dataset includes 2331 patients after excluding records with missing values. The final observation in the database is 15 years. This time period is divided into five time intervals each one trying to contain the same amount of events. There is one neural network for each time interval, estimating the survivability at this time. Censoring patients are considered only in the time intervals when they are active, that is still in the study. Uncensored patients are also considered until the time interval that they die.

Therefore, for each time interval a three layer feed forward network is constructed and trained by back propagation. The output layer contains one node which generates an output value from 0 to 1, where 0 representing high chance of survival and 1 low chance for that time interval. The records in each time interval are divided to training and test sets with a specific methodology in order to eliminate bias.

The main idea of this study is that the ANN output is transformed to a survival function using Bayes theorem. If the ANN output at a time interval is above a certain cut off level (Γi) this would indicate a low chance of survival. An output lower than the cut off would indicate high chance of survival and this record is presented to the

subsequent network for the next time interval. The probability figures are calculated in a way described below.

The probability of death at the end of a time interval [I,i+1) is: $P(D)_{i+1} = 1 - \frac{d_i}{n_{i+1} - c_i - d_i}$ where d (i) is number of deaths at the interval, n (i) is the total number of patients and c (i) is the number of censoring cases. Some definitions are:

T/d_i	The number of patients who died from the tumor and had score value >= Γi
t/d_i	Number of patients who died from the tumor and had score value < Γi
T/n_i	All patients with score >= Γi
t/n_i	All patients who had score < Γi

In this particular study the syndrome S is represented by a high ANN score (>=Γi). The probability of death given the presence and non presence of the syndrome S can be defined with the help pf Bayes theorem.

$$P(D/S)_i = \frac{P(D)_i \;\; x \;\; P(S/D)_i}{P(S)_i} \;,\; P(D/\overline{S})_i = \frac{P(D)_i \;\; x \;\; P(\overline{S}/D)_i}{P(\overline{S})_i}$$

The parameters of the theorem are computed as follows:

$$P(S/D)_i = \frac{patients \;\; died \;\; with \;\; score \;\; \geq \Gamma i}{total \;\; deaths} = \frac{T/d_i}{d_i}$$

$$P(S)_i = \frac{patients \;\; with \;\; score \;\; \geq \Gamma i}{tatal \;\; patients} = \frac{T/n_i}{n_i}$$

$$P(\overline{S}/D)_i = \frac{patients \;\; died \;\; with \;\; score \;\; < \Gamma i}{total \;\; deaths} = \frac{t/d_i}{d_i}$$

$$P(\overline{S})_i = \frac{patients \;\; with \;\; score \;\; < \Gamma i}{tatal \;\; patients} = \frac{t/n_i}{n_i}$$

The survival function SF is represented by the following sequence:

$$SF_t = \begin{cases} 1 & for \;\; t = 0 \\ 1 - P(D/S)_t & for \;\; t \geq i-1 \;\; if \;\; score \;\; for \;\; [i-1,i) \geq \Gamma i \\ 1 - P(D/\overline{S})_i & for \;\; t \geq i \;\; if \;\; score \;\; for \;\; [i-1,i) < \Gamma i \end{cases}$$

The results of the ANN in comparison to the Kaplan Meier and the Cox model, prove that the AI system was on average 41% lower than the survival probability predicted by the Cox model, and 37% lower than the Kaplan Meier. However, when the samples increase the difference is lower than 20%. The Kaplan Meier survival curve is shown in figure 36.

These results can hardly be compared to the ones from this study that are presented on chapter 5. The reason is that the two implementations use different datasets even though our dataset is a subset of Taktak's dataset. In addition, Taktak reduced the dataset n order to treat missing values according to five parameters (covariates). In this study the same methodology was followed for three variables. These eliminations obviously affected the statistical result.

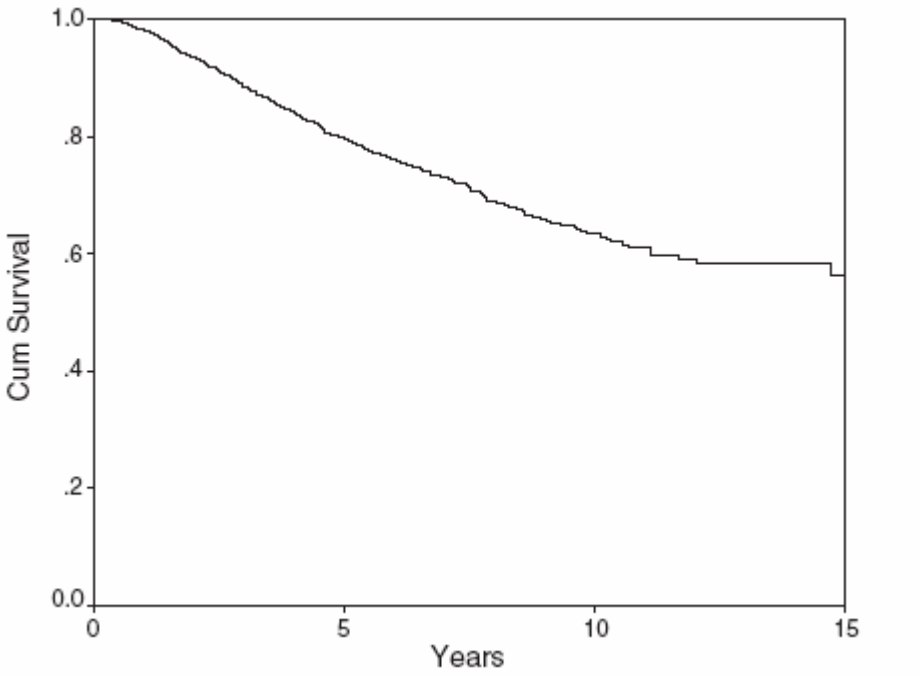


Figure 36. Kaplan Meier survival curve for 2331 patients of intraocular melanoma [42].

6.5 Comparison with a study on prognostic factors for survival after enucleation.

There is a relevant study on prognostic factors from Isager, Ehlers and Overgaard (2003) [43]. The purpose of this study is to evaluate prognostic factors for the survival of patients treated by enucleation for choroidal and Ciliary body melanomas. The non parametric Kaplan Meier and the semi parametric Cox model are used to plot survival curves and discover influential covariates.

The study includes 293 patients treated by primary enucleation for choroidal and Ciliary body melanoma. Kaplan Meier analysis is performed for death from melanoma and death from all causes. In melanoma specific survival, patients that died from other causes are considered censored. Prognostic factors are estimated by univariate Cox proportional Hazards (Cox examining one covariate only) model and by Kaplan Meier with the log rank test.

Parameters that are known from literature to be prognostic or with a log rank p value less than 0.2 (meaning that populations are not similarly affected by the covariate) are include in the multivariate Cox proportional hazards analysis. The covariates that are examined are tumor location, largest basal diameter, epithelioid cellularity and extrascleral extension. The cause of death was melanoma in 56% of the population and non melanoma in 44%. The melanoma related deaths were verified by autopsy. The authors provide the following survival probabilities in the cumulative melanoma specific survival function:

Time	**Survival**
5 years	70%
10 years	53%
15 years	47%
20 years	45%
25 years	41%

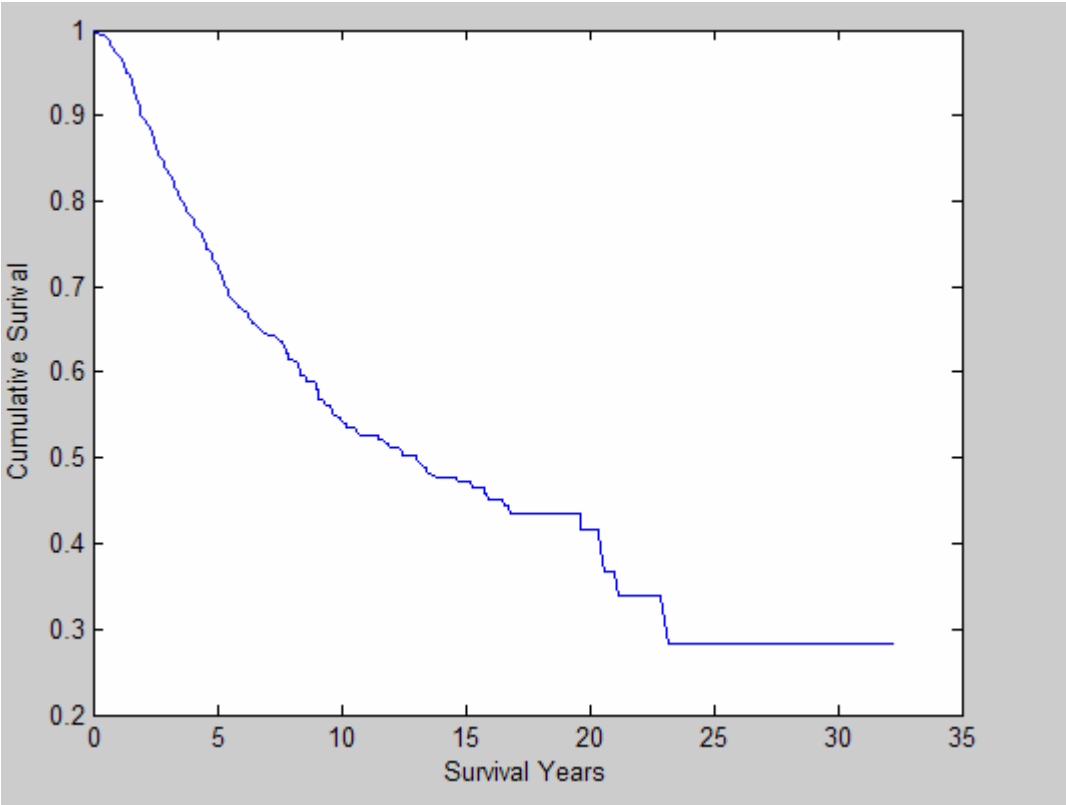


Figure 37. The Kaplan Meier survival curve for the ocular melanoma dataset of this study.

Figure 37 shows the Kaplan Meier survival curve of the present study of the ocular melanoma disease. Using this figure a similar matrix can be constructed.

Time	**Survival**
5 years	73%
10 years	55%
15 years	48%
20 years	45%
25 years	30%

It is obvious that the survival probabilities are close which is reasonable since the same disease is under examination. Any difference could be due to the fact that in Isager's study there are many deaths due to non melanoma causes. On the contrary, in the ocular melanoma dataset censored patients are considered alive.

The results in Isager's study with the Cox proportional hazards model proves that the epithelioid cellularity is associated with the worst prognosis, as expected by the authors. The prognosis has been found to deteriorate with an increasing number of epithelioid cells. The population was divided into two groups the first one containing patients with spindle cell type (non epithelioid) and the second one with epithelioid patients. The survival curves as obtained from the Cox model are shown in figure 38 where both of the populations are shown.

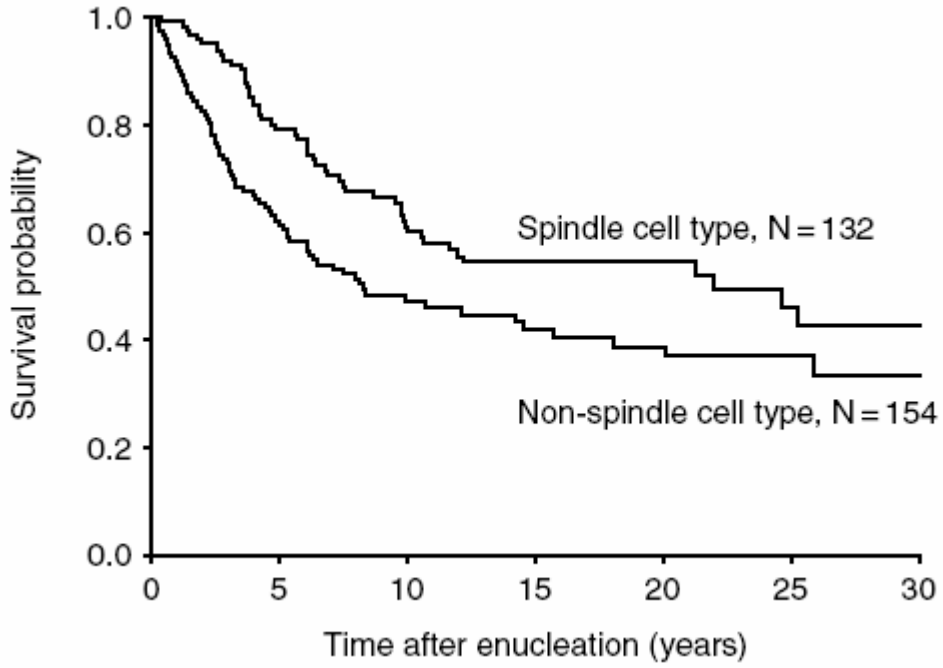
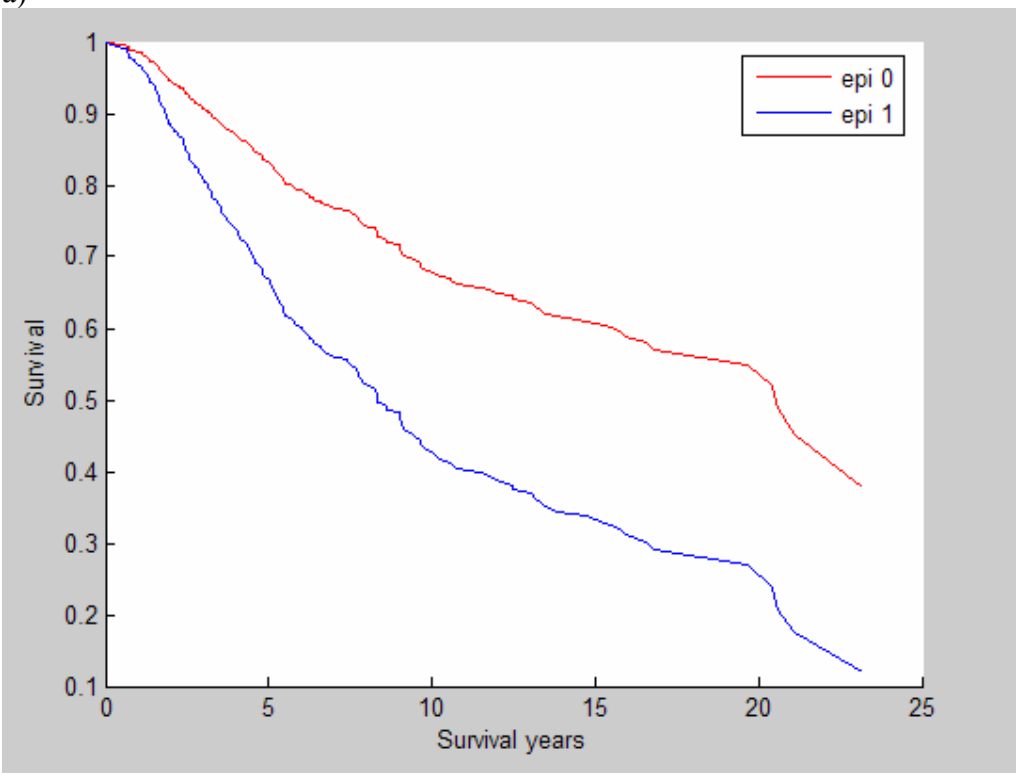


Figure 38. Survival curves for patients with epithelioid and non epithelioid cells as they derived from the Cox model. [43]

a)



b)

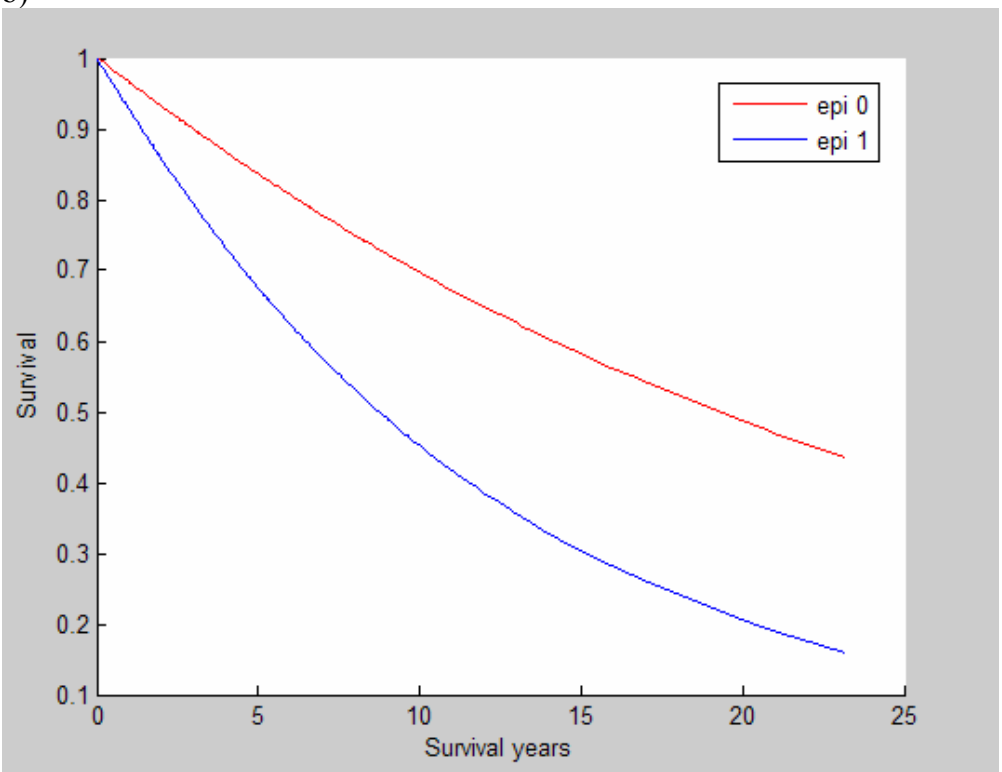


Figure 39. a) Survival curves from the Cox model of this study as it was presented in chapter 5. There are two populations, one with epithelioid cells and one without them. **b)** The same plots fitted to distributions.

Figures 38 and 39.b are comparable. Figure 28 shows the survival curves from Isager's study and figure 39 from this study. Both of them are created with the Cox model. The survivability for the population with epithelioid cells is close in the two studies. On the contrary there is much difference on the epithelioid population. An important mark is that Isager's study is survival analysis after a treatment (enucleation) and the present study does not involve any treatments. Nevertheless, an important conclusion is that Cox regression in both studies showed that epithelioid cellularity the most important factor, the presence of which influences the disease enormously.

6.6 Comparison with a relevant study from the University of Helsinki on prognostic indicators in choroidal and ciliary body melanoma.

There is a relevant study from Teemu Makitie [44] on survival analysis of intraocular melanoma. This study is for the purposes of an Ophthalmology department and therefore there are more clinical and medical than statistical results.

The study implements both the non parametric Kaplan Meier method and the semi parametric Cox model. The product limit estimator is used to obtain melanoma specific survival curves. In the dataset used, there are patients that died from other non melanoma causes and therefore they are considered as censored. The Cox proportional hazards regression is used to adjust the survival for the effect of previously identified independent predictors.

Five covariates are used in the Cox model, the microvascular density, the largest basal tumor diameter. Epithelioid cells , microvascular patterns and Ezrin. The covariate that turned out from Cox regression to influence the disease most is the presence of epithelioid cells. The same conclusion was made in the present study too which is available in chapter 5. In addition, the regression parameters for the epithelioid variable are very close. The regression coefficient of Makitie is 0.612 while our regression coefficient is 0.664. The regression coefficient for the largest basal diameter is 0.128 and our regression coefficient is 0.127.

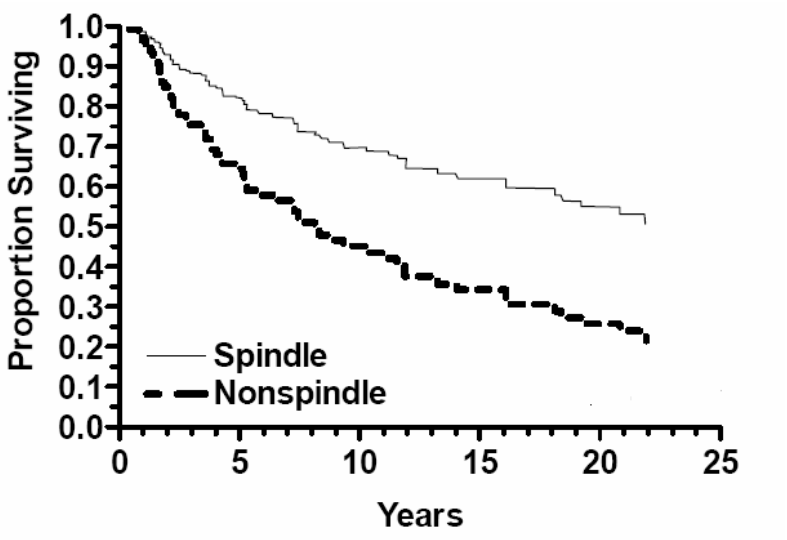


Figure 39. Melanoma specific survival curves generated by Cox regression analysis for two populations, epithelioid and non epithelioid.

The population is divided into two sets considering the existence or not of epithelioid cells. The melanoma specific survival curves (because the author presents non melanoma survival curves including patients that did not die from melanoma disease) is shown in figure 39. Even though the present study and Makitie 's study use different datasets of the same disease, the survival curves are very close.

At the 23rd year of the disease, patients with epithelioid melanomas have around 20% probability of surviving in both cases. In addition, on the 23rd year of the disease non epithelioid patients have 40% probability of surviving. There are of course some differences which are mostly due to the different dataset and to the fact that Makitie's dataset involves records with non melanoma deaths.

Chapter 7
Conclusions and future work

The purpose of this study is to analyze the survivability of patients suffering from the ocular melanoma disease. Survival analysis offers three kinds of models. The non parametric models make no assumption about the form of the survival function, while they do not include covariates. They are the simplest models to implement in order to plot patients' hazard and survival functions. The second class of models is the semi parametric models. Like non parametric, the semi parametric make no assumption about the distribution of the survival function but what makes them unique is their ability to investigate the effect of covariates on the evolution of the disease. Covariates are measurements taken from a group of patients while they were studied. Semi parametric models allow identifying, through regression procedures, covariates that can be considered as prognostic factors as well as making a comparison among them. The third class of models consists of fully parametric models. A study of the survivability of patients can lead biostatisticians to have an a priori knowledge of the distribution of the survival function. Therefore there is a fully parametric model for each well known distribution. A regression procedure can also be applied to identify which covariates are prognostic factors and explore the amount of their influence on the evolution of the disease.

From the class of non parametric models the Product limit model (Kaplan Meier estimator) is implemented to plot the survival function of the population. This function proves that the median survival, which is 50% survivability, is at the 17^{th} year after the disease is diagnosed. In other words, an ocular melanoma patient has 50% probability of surviving after the 17^{th} year conditional that he has reached this year. In addition, the original population is divided into two sets, the first one including patients with epithelioid cellular and the second patients without it. Two separate survival curves are plotted showing the patients with epithelioid cellularity have at any time point less survival probabilities. The significant difference of these curves is also checked with the log rank test.

The Cox proportional hazards model is implemented from the class of semi parametric models. In this model there is baseline hazard function that depicts the hazard ratio or the survival probability of the whole population. This function is flexed up or down when a single patient is considered. The baseline function is approximated through the Breslow and the extended Kaplan Meier estimator. In addition three covariates are examined in this model, the ultrasound height, the longest ultrasound basal dimension and the epithelioid cellularity. Regression is performed through maximum likelihood estimation and the effect of each covariate derives. The covariate that affects the disease most is the epithelioid cellular. This is confirmed from the results gained from the non parametric analysis. In addition the longest ultrasound basal dimension affects the melanoma more than the ultrasound height. To explain further, a patient that has big ultrasound height is more likely to experience to experience the event (for this study the event under observation is death) if he has a large value of the longest ultrasound basal dimension covariate. All three covariates are proved to have direct and positive effect on the disease, that is the bigger their values the bigger the effect.

From the class of fully parametric model, the log logistic model is implemented. The reason why this model is used is because after observation of the survival function the log logistic distribution fitted well. First a test is applied on the specific dataset to check if logistic distribution is appropriate for the data. This test checks the linearity of certain expression of the log logistic distribution against the logarithm of time. The time variable in this study is the survival years variable from the dataset. The linearity is satisfied and therefore the log logistic distribution fits the specific dataset. Furthermore, regression is applied in a graphical way that differs a lot form the Cox model. Unfortunately, log logistic regression is proven that it can not be performed for the specific dataset.

Neural networks are also used in this study. Their purpose in the Cox model is to approximate the baseline hazard function of a certain population and estimate the regression parameters. Neural networks, with their advantage to generalize, have been tested with unseen patients and the results showed high correlation of the inputs and targets.

The conclusions of this study are of high interest in the medical field. The experience of a certain group of subjects can be generalized for the ocular melanoma disease. Doctors often need a priori knowledge of the survivability of each patient. When doctors take measurements from a newly diagnosed patient, they can use the knowledge of the prognostic factors of this study (that are confirmed with other relevant studies) to predict survival.

This study can be extended in the future in two ways. The first is to continue working on these specific models by exploring their variations. The Cox model can be altered to include stratification. Stratum is a group of patients with a specific characteristic. Given a covariate the original population can be divided into two or more sets including patients with specific numerical values of this covariate. The epithelioid cellular can be used to divide the population in to two sets as it has been widely performed in this study. The difference would be that the baseline hazard approximation and the regression would be performed separately for each stratum. The results of such a model can justify more the serious effect of epithelioid cellularity on survivability.

In addition the "age" can be used in the Cox model to examine whether there is a dependence of age and survivability due to the melanoma. In this case, the age should be considered as a time dependent covariate as it changes throughout the study period. The date of birth of each patient is known and age can be extracted. To perform such an analysis the simple version of the Cox model is non applicable. An extension of it with time dependent covariates is more appropriate. In addition, a test of the proportional hazards assumption should be performed (Test based on defined time dependent covariates, Gill and Schumacher test, O' Quigley and Pessione test).

A second way to extend this study is to apply on the ocular melanoma dataset new models. Several other fully parametric models can be applied on the dataset (Weibull, Gamma, Exponential regression models). In addition there are many semi parametric models that can be used. Apart from the Aalen's additive model, a combination is the Cox Aalen model where covariates are accepted to have multiplicative or additive effect on hazard.

In the Cox proportional hazards model the baseline hazard is conditioned out and only the impact of the covariates are estimated by maximizing the partial likelihood. No functional form of hazard has to be specified which make the Cox model very flexible. On the other hand the parametric models have to specify the functional form of the hazard function. However when the hazard function is of interest it is usually estimated with the Breslow estimator which lacks the ability to test hypothesis about the shape of the hazard function. The Piecewise Exponential model can extend this study in the future, as it is a model that is between two extremes. In this model time is divided into intervals. The hazard in each interval is assumed to be constant but can vary across intervals. It has the flexibility of the Cox model and the ability to statistically check the hazard function.

Neural networks can also be used in a different way than they are in this study. They are not only functional in approximating a specific function as they are in this study. Neural networks are lately used to predict survival themselves. Even though this field is still under examination, several types and structures can be used to predict survival for the ocular melanoma dataset (Ohno- Machado (1996), Ravdin and Clark.(1992), by Biganzoli et al. (1998), Lapuerta et al.(1995) , Faraggi (1995), Street (1998) , Mani (1999)).

The event under examination in this study is death. Time to event is computed with the models analyzed earlier. The survival analysis field has been extended to measure not only time to death but also to other events, for example recovery after surgery. This study can be extended by examining metastatic death as the event under investigation. Metastasis is the primary reason of death for ocular melanoma patients. It is of high medical interest to examine the effect of covariates on metastasis as well as predicting time to metastasis. This variable can be used in all three models examined in this thesis (Kaplan Meier, Cox proportional hazards and log logistic regression) and offer interesting results.

APPENDIX A

This Appendix includes all programming codes that have been programmed for the purposes of Kaplan Meier, Cox regression, Log logistic regression, log rank test and neural networks. The environments used are MATLAB, R language and SPSS.

The MATLAB code for the Kaplan Meier estimator:

```
function kaplan_meier

%with the kaplan.m workspace
load kaplan.mat
kaplan = [];
mul=1;
j=1;
k=1;
temp1=0;
risk=[];

    for k=1:560
          risk(k)=743-temp1;
          temp1=temp1+m(k,2)+s(k,2);
    end

      for i=1:560
          temp2(i)=1-m(i,2)/risk(i);
          time2(i)=m(i,1);
      end

kaplan(1)=temp2(1);
      for  j = 2:560
       kaplan(j)=kaplan(j-1)*temp2(j);
      end
 time2=transpose(time2);
figure(1)
 plot(time2,kaplan);
 xlabel('Survival Years');
 ylabel('Cumulative Surival');
```

The SPSS code to perform the Kaplan Meier:

```
KM
  V5  /STATUS=V4(1)
  /PRINT TABLE MEAN
  /PLOT SURVIVAL .
```

Where V5 stands for the survival years and V4 the death (or censoring) indicator.

The MATLAB code for the logrank test:

```
function log_rank
load bres.mat
k=0;
n=0;
o=0;
t=0;
for i=1:743
    if data(i,3)==1
        k=k+1;
        if death(i)==1
            o=o+1;
        end
    end
    if data(i,3)==0
        n=n+1;
        if death(i)==1
            t=t+1;
        end
    end
end

c=1;
f=1;
prob1=[];
prob0=[];
for i=1:743
    if death(i)==0
        if data(i,3)==1
            k=k-1;
        end
         if data(i,3)==0
            n=n-1;
         end
    end

    if death(i)==1
      prob1(c)=k*(1/(k+n));
      prob0(f)=n*(1/(k+n));
      c=c+1;
      f=f+1;
        if data(i,3)==1
            k=k-1;
        end
        if data(i,3)==0
            n=n-1;
        end
    end

end

k=184;
n=89;
exp_death1=sum(prob1);
exp_death0=sum(prob0);

chi=(((n-exp_death0)^2)/exp_death0) + (((k-
exp_death1)^2)/exp_death1);
```

The SPSS code to perform the Cox Regression is:

Here the variable V5 is time, V4 shows whether an individual has experienced the event or not (0= alive (censored), 1=dead (uncensored)) and variables V1, V2, V3 are the regression's covariates. It is also required a baseline and survival plot.

```
COXREG
  V5  /STATUS=V4(1)
  /METHOD=ENTER V1 V2 V3
  /PLOT SURVIVAL HAZARD
  /PRINT=CI(95) CORR BASELINE
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20)
```

R code for the Breslow estimator based on basehaz.gbm

```
t.unique <- sort(unique(t[delta==1]))
   nominator <- length(t.unique)
   for(i in 1:length(t.unique))
   {
      nominator[i] <- sum(t[delta==1]==t.unique[i])
}
denominator <- length(t.unique)
   for(i in 1:length(t.unique))
   {
      denominator[i] <- sum(exp(f.x[t>=t.unique[i]]))

}

print(nominator)
print(denominator)

breslow <- length(t.unique)
   for(i in 1:length(t.unique))
   {
      breslow[i] <- nominator/denominator;
   }

print (breslow)
```

Alternatively, a MATLAB code to compute the Breslow estimator and fit to the log logistic cumulative hazard is:

```
function breslow
load bres.mat
temp3=0;
time2=[];
t_unique=[];
k=1;
m2=[];

for i=1:560
    if m(i,2)>0
        t_unique(k)=m(i,1);
        m2(k)=m(i,2);
        k=k+1;
    end
end
sum=[];


for i=1:240
    for k=1:743
        if (time(k)==t_unique(i))
           for j=k:743
            sum(i)=temp3+exp(data(j,:)*b(:,1));
            temp3=sum(i);
           end
         temp3=0;
         break;
        end
    end
end

for i=1:240
    breslow(i)=m2(i)/sum(i);
end
breslow_est=cumsum(breslow);
figure(2)
plot(t_unique,breslow_est)

%---------------------------------------
t_unique=t_unique';
breslow_est=breslow_est';
g = fittype('log(1+a*(x)^c)');
F=FITOPTIONS('METHOD','NonLinearLeastSquares','StartPoint',[0.1,0.1])
;
[FITTEDMODEL,GOODNESS,OUTPUT]=fit(t_unique,breslow_est,g,F);

a=FITTEDMODEL.a;
c=FITTEDMODEL.c;
y=[];
for i=1:240
    y(i)= log(1+a*(t_unique(i))^c);
end

figure(1)
 hold on
 plot(t_unique,y);
 xlabel('Survival Years');
```

```
 ylabel('Cumulative Hazard');
 plot(t_unique,breslow_est,'*');
 hold off;

 survival=[];

 for i=1:240
%      survival(i)=1/(1+a*((t_unique(i))^c));
survival(i)=exp(-breslow_est(i));
 end

 figure(2)
 plot(t_unique,survival);
 xlabel('Survival Years ');
 ylabel('Cumulative Survival');

 h=[];
for i=1:240
     h(i)=(a*c*(t_unique(i)^(c-1)))/(1+a*(t_unique(i)^c))
end

figure(3)
plot(t_unique,h);
xlabel('Survival Years');
ylabel('Hazard');
```

The MATLAB code of the analogue to the Kaplan Meier estimator and its fitting to the log logistic distribution

```
function kaplan_extended

load bres.mat
breslow;
t_unique2=[];
m3=[];
k=1;


for i=1:560
    if m(i,2)>0
        t_unique2(k)=m(i,1);
        m3(k)=m(i,2);
        k=k+1;
    end
end

temp3=0;
sum2=[];
for i=1:240
    for k=1:743
        if (time(k)==t_unique2(i))
           for j=k:743
            sum2(i)=temp3+exp(data(j,:)*b(:,1));
            temp3=sum2(i);
           end
         temp3=0;
         break;
        end
    end
```

```
end
kaplan=[];
for i=1:240
    kaplan(i)=1-m3(i)/sum2(i);
end

kaplan_est=[];
kaplan_est(1)=kaplan(1);
for i=2:240
    kaplan_est(i)=kaplan_est(i-1)*kaplan(i);
end

kaplan_hazard=[];
for i=1:240
    kaplan_hazard(i)=-log(kaplan_est(i));
end

%------------------------
 t_unique2=t_unique2';
kaplan_hazard=kaplan_hazard';
g = fittype('log(1+a*(x)^c)');
F=FITOPTIONS('METHOD','NonLinearLeastSquares','StartPoint',[0.1,0.1])
;
[FITTEDMODEL,GOODNESS,OUTPUT]=fit(t_unique2,kaplan_hazard,g,F);

a=FITTEDMODEL.a;
c=FITTEDMODEL.c;
y2=[];
for i=1:240
    y2(i)= log(1+a*(t_unique2(i))^c);
end

figure(4)
 hold on
 plot(t_unique2,y2);
 xlabel('Survival Years');
 ylabel('Cumulative Hazard');
 plot(t_unique2,kaplan_hazard,'*');
 hold off;

 survival2=[];


 figure(5)
 plot(t_unique2,kaplan_est);
 xlabel('Survival Years ');
 ylabel('Cumulative Survival');

 h2=[];
for i=1:240
     h2(i)=(a*c*(t_unique2(i)^(c-1)))/(1+a*(t_unique2(i)^c))
end

figure(5)
plot(t_unique2,h2);
xlabel('Survival Years');
ylabel('Hazard');
ylabel('Hazard');
```

The MATLAB code to plot the cumulative survival function from the cox model for patients with epi=0 and epi=1

```
function survival_epi01

load matlab2.mat

b=[0.127;0.033;0.664];
k=1;
m=1;
exponential0=[];
exponential1=[];

for i=1:743
    if data(i,3)==1
        exponential1(k)=exp(data(i,:)*b(:,1));
        k=k+1;
    end
    if data(i,3)==0
        exponential0(m)=exp(data(i,:)*b(:,1));
        m=m+1;
    end
end

s0=[];
s1=[];

for i=1:279
    for j=1:240
        s0(j,i)=exp(-y(j)*exponential0(i));
    end
end

for i=1:464
    for j=1:240
        s1(j,i)=exp(-y(j)*exponential1(i));
    end
end

figure (1)
hold on;

  plot(t_unique,s0(:,8));
  plot(t_unique,s0(:,12));
  plot(t_unique,s0(:,16));
  plot(t_unique,s0(:,25));
  plot(t_unique,s0(:,34));

for i=1:5
  plot(t_unique,s1(:,i),'r');
end
xlabel('Survival time');
ylabel('Hazard')
hold off;
```

The MATLAB code to plot the density hazard function from the cox model for patients with epi=0 and epi=1

```
function hazard_epi01

load matlab.mat
b=[0.127;0.033;0.664];
k=1;
m=1;
exponential0=[];
exponential1=[];

for i=1:743
    if data(i,3)==1
        exponential1(k)=exp(data(i,:)*b(:,1));
        k=k+1;
    end
    if data(i,3)==0
        exponential0(m)=exp(data(i,:)*b(:,1));
        m=m+1;
    end
end
h0=[];
h1=[];
for i=1:279
    for j=1:240
    h0(j,i)=h(j)*exponential0(i);
    end
end


for i=1:464
    for j=1:240
    h1(j,i)=h(j)*exponential1(i);
    end
end


figure (1)
hold on;

  plot(t_unique,h0(:,8));
  plot(t_unique,h0(:,12));
  plot(t_unique,h0(:,16));
  plot(t_unique,h0(:,25));
  plot(t_unique,h0(:,34));

for i=1:5
  plot(t_unique,h1(:,i),'r');
end
xlabel('Survival time');
ylabel('Hazard')
hold off;
```

MATLAB code to plot the survival function o patients with epi0 and epi1

```
function survival_epi01

load matlab2.mat

b=[0.127;0.033;0.664];
k=1;
m=1;
exponential0=[];
exponential1=[];

for i=1:743
    if data(i,3)==1
        exponential1(k)=exp(data(i,:)*b(:,1));
        k=k+1;
    end
    if data(i,3)==0
        exponential0(m)=exp(data(i,:)*b(:,1));
        m=m+1;
    end
end

s0=[];
s1=[];

for i=1:279
    for j=1:240
        s0(j,i)=exp(-y(j)*exponential0(i));
    end
end

for i=1:464
    for j=1:240
        s1(j,i)=exp(-y(j)*exponential1(i));
    end
end

figure (1)
hold on;

  for i=1:279
  plot(t_unique,s0(:,i));
end

for i=1:464
  plot(t_unique,s1(:,i),'r');
end
xlabel('Survival time');
ylabel('Hazard')
hold off;
```

MATLAB code to plot the survival function o patients with high and low ludb

```
function survival_ludb

load matlab2.mat

b=[0.127;0.033;0.664];
k=1;
m=1;
ludb0=[];
ludb1=[];

for i=1:743
    if data(i,1)>=14.9
      ludb1(k)=exp(data(i,:)*b(:,1));
        k=k+1;
    end
    if data(i,1)<14.9
        ludb0(m)=exp(data(i,:)*b(:,1));
        m=m+1;
    end
end

s0=[];
s1=[];

for i=1:463
    for j=1:240
        s0(j,i)=exp(-y(j)*ludb0(i));
    end
end

for i=1:280
    for j=1:240
        s1(j,i)=exp(-y(j)*ludb1(i));
    end
end

figure (1)
hold on;

for i=1:463
  plot(t_unique,s0(:,i));
end

for i=1:280
  plot(t_unique,s1(:,i),'r');
end
xlabel('Survival time');
ylabel('Cumulative Survival')
hold off;
```

MATLAB code to plot the survival function o patients with high and low uh

```
function survival_uh01
load matlab2.mat

b=[0.127;0.033;0.664];
k=1;
m=1;
uh0=[];
uh1=[];

for i=1:743
    if data(i,2)>=10
       uh1(k)=exp(data(i,:)*b(:,1));
        k=k+1;
    end
    if data(i,2)<10
        uh0(m)=exp(data(i,:)*b(:,1));
        m=m+1;
    end
end

s0=[];
s1=[];

for i=1:549
    for j=1:240
        s0(j,i)=exp(-y(j)*uh0(i));
    end
end

for i=1:194
    for j=1:240
        s1(j,i)=exp(-y(j)*uh1(i));
    end
end

figure (2)
hold on;

  for i=1:549
  plot(t_unique,s0(:,i));
end

for i=1:194
  plot(t_unique,s1(:,i),'r');
end
xlabel('Survival time');
ylabel('Cumulative survival')
hold off;
```

MATLAB code to plot the survival function o patients with epi1 and high ludb and patients with epi0 and high ludb

```
funtion survival_epi_ludb01
load matlab2.mat

b=[0.127;0.033;0.664];
k=1;
m=1;
exponential0=[];
exponential1=[];

for i=1:743
    if (data(i,3)==1 && data(i,1)>=14.9)
        exponential1(k)=exp(data(i,:)*b(:,1));
        k=k+1;
    end
    if (data(i,3)==0 && data(i,1)>=14.9)
        exponential0(m)=exp(data(i,:)*b(:,1));
        m=m+1;
    end
end

s0=[];
s1=[];

for i=1:87
    for j=1:240
        s0(j,i)=exp(-y(j)*exponential0(i));
    end
end

for i=1:193
    for j=1:240
        s1(j,i)=exp(-y(j)*exponential1(i));
    end
end

figure (1)
hold on;

  for i=1:87
  plot(t_unique,s0(:,i));
end

for i=1:193
  plot(t_unique,s1(:,i),'r');
end
xlabel('Survival time');
ylabel('cumulative survival')
hold off;
```

The MATLAB code to check the suitability of the log logistic distribution

```
function logistic_suitability
load log_suit.mat

s=[];
f=[];
odds=[];
logtime=[];

for i=1:560
    f(i)=1-kaplan(i);
end

for i=1:560
    odds(i)=log(kaplan(i)/f(i));
end
for i=1:560
    logtime(i)=log(time2(i));
end
plot(logtime,odds);
xlabel('logt')
ylabel('log odds')
```

The MATLAB code for the log logistic regression (subsets: 1) epi=1, uh<10 2)epi=0, uh>10)

```
function kaplan_logistic
load kaplan_log.mat
    for k=1:132
          risk1(k)=132-temp1;
          temp1=temp1+m1(k)+s1(k);
    end
      for i=1:132
          temp2(i)=1-m1(i)/risk1(i);
      end
kaplan1(1)=temp2(1);
      for  j = 2:132
       kaplan1(j)=kaplan1(j-1)*temp2(j);
      end
 %---------------------

    for k=1:332
          risk2(k)=332-temp1;
          temp1=temp1+m2(k)+s2(k);
    end

      for i=1:332
          temp2(i)=1-m2(i)/risk2(i);
      end

kaplan2(1)=temp2(1);
      for  j = 2:332
       kaplan2(j)=kaplan2(j-1)*temp2(j);
```

```
      end


function logistic_regression
load logistic_regr.mat

for i=1:132
    f1(i)=1-kaplan1(i);
end

for i=1:132
    odds1(i)=log(kaplan1(i)/f1(i));
end

for i=1:132
    logtime1(i)=log(time1(i));
end
%----------------------------
for i=1:332
    f2(i)=1-kaplan2(i);
end

for i=1:332
    odds2(i)=log(kaplan2(i)/f2(i));
end

for i=1:332
    logtime2(i)=log(time2(i));
end

figure(1)
hold on;
plot(logtime1,odds1);
plot(logtime2,odds2,'r');
xlabel('log time');
ylabel('log odds')
hold off;
```

The MATLAB code that creates and trains the baseline network

```
function [Y]=network_breslow
load net_bre;

train_p=[];
train_t=[];
sim_p=[];
sim_t=[];

k=1;
m=1;
index=[];
for i=1:240
    index(i)=0;
end
for i=1:2:240
    index(i)=1;

end
for i=1:240
    if index(i)==0
       sim_p(:,k)=p(:,i);
       sim_t(k)=t(i);
       k=k+1;
    end
    if index(i)==1;
       train_p(:,m)=p(:,i);
       train_t(m)=t(i);
       m=m+1;
    end

end

net=newff(minmax(train_p),[2 5 5 20 5 5
1],{'tansig','tansig','tansig','tansig','tansig','tansig','purelin'},
'trainlm','learngdm','mse');
net.trainParam.epochs=2000;
net.trainParam.goal=0;
net=train(net,train_p,train_t);

Y=[];
a_sim=[];
t_sim=[];


k=1;
    for i=1:120
    Y(k)=sim(net,[sim_p(1,i);sim_p(2,i);]);
    a_sim(k)=Y(k);
    t_sim(k)=sim_t(k);
   k=k+1;
    end
figure(2)
[m(2),b(2),r(2)]=postreg(a_sim,t_sim);
```

The MATLAB code that creates and trains the covariates network

```
function [Y]=train_net_3(NTR,NTST,PATIENT);
%NTR=# training Data
%NTST=# testing Data
load p_t_3; %load input output vectors

[pn,meanp,stdp,tn,meant,stdt] = prestd(p,t);
%[ptrans,transMat] = prepca(pn,0.02);
%[m,n] = size(ptrans);

% iitest=2:round(n/NTST):n;
% iival=4:round(n/NVAL):n;
% iitrn=[1:round(n/NTR):743 3:round(n/NTR):n];
[m,n]=size(p);
available_index(n)=0;
k=1; %number of indices for training data
while 1==1,
    j=round(rand(1)*n);
    if j>0
        if available_index(j)==0
            available_index(j)=1;
            iitrn(k)=j;
            k=k+1;
        end
        if k==NTR
            break;
        end
    end
end


k=1; %number of indices for testing data
while 1==1,
    j=round(rand(1)*n);
    if j>0
        if available_index(j)==0
            available_index(j)=1;
            iitest(k)=j;
            k=k+1;
        end
        if k==NTST
            break;
        end
    end
end

k=1; %number of indices for validation data
while 1==1,
    j=round(rand(1)*n);
    if j>0
        if available_index(j)==0
            available_index(j)=1;
            iival(k)=j;
            k=k+1;
        end
        if k==(n-NTR-NTST)
            break;
        end
    end
end
```

```
val.P=p(:,iival); %validation Input Data
val.T=tn(:,iival);%validation Output Data

%test.P=ptrans(:,iitest); %test Input Data
test.P=p(:,iitest); %test Input Data
test.T=tn(:,iitest);%test Output Data

%ptr=ptrans(:,iitrn); %Training Input Data
ptr=p(:,iitrn); %Training Input Data
ttr=tn(:,iitrn);%Training Output Data

net=newff(minmax(ptr),[5 1],{'tansig' 'purelin'},'trainlm');
net.trainParam.epochs = 500;
net.Trainparam.show=1;
[net,tr]=train(net,ptr,ttr,[],[],val,test);

an=sim(net,p);
a=poststd(an,meant,stdt);

a_sim=[];
t_sim=[];
    SIM_VECTOR=[p(1,PATIENT);p(2,PATIENT);p(3,PATIENT)];
    Yn=sim(net,SIM_VECTOR);
    Y=poststd(Yn,meant,stdt);
    disp('-----------------------');
    disp('Item Selected:');
    disp(PATIENT);
    disp('->Neural Network output:');
    disp(Y);
    disp('->Target Vector:');
    disp([t(PATIENT)]);
    disp('->Input Vector:');
    disp(SIM_VECTOR);

    a_sim=[a_sim Y];
    t_sim=[t_sim t(PATIENT)];

figure(1)
[m(1),b(1),r(1)]=postreg(a,t);

figure(2)
[m(2),b(2),r(2)]=postreg(a_sim,t_sim);
```

APPENDIX B

In this Appendix additional results are presented for the Kaplan Meier and Cox models. SPSS results, MATLAB fitting results and MATLAB neural network results are available here.

The results from the SPSS Kapan Meier

Kaplan-Meier

Case Processing Summary

Total N	N of Events	Censored	
		N	Percent
743	273	470	63,3%

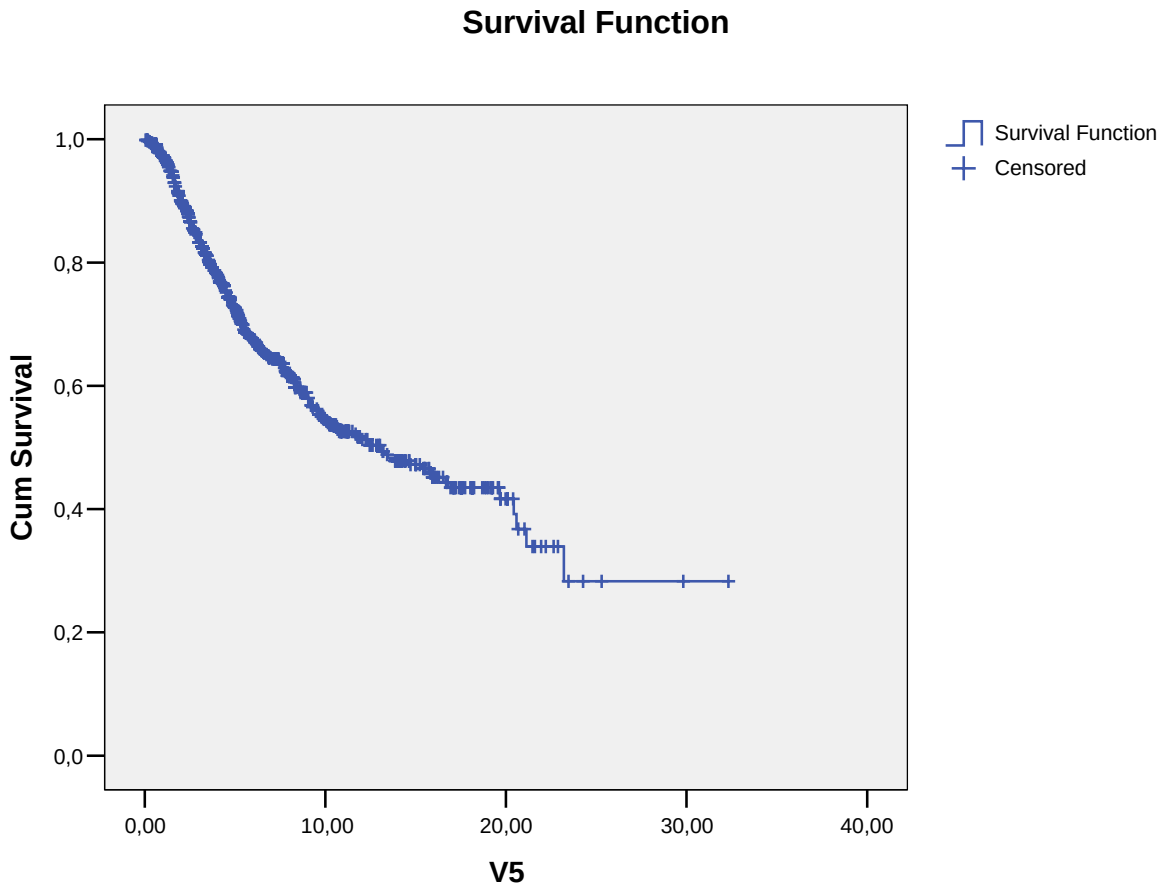
Survival Table

	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
			Estimate	Std. Error		
1	,020	1	,999	,001	1	742
2	,060	0	.	.	1	741
3	,070	0	.	.	1	740
4	,140	0	.	.	1	739
5	,140	0	.	.	1	738
6	,160	1	,997	,002	2	737
7	,170	0	.	.	2	736
8	,180	0	.	.	2	735
9	,240	1	,996	,002	3	734
10	,260	0	.	.	3	733
11	,270	0	.	.	3	732
12	,310	1	,995	,003	4	731
13	,340	0	.	.	4	730
14	,350	0	.	.	4	729
…	…	…	…	…	….	…
736	22,640	0	.	.	272	7
737	22,880	0	.	.	272	6
738	23,210	1	,283	,066	273	5
739	23,460	0	.	.	273	4
740	24,270	0	.	.	273	3
741	25,300	0	.	.	273	2
742	29,820	0	.	.	273	1
743	32,320	0	.	.	273	0

Means and Medians for Survival Time

Mean(a)				Median			
		95% Confidence Interval				95% Confidence Interval	
Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
15,969	,861	14,282	17,655	13,040	1,566	9,971	16,109

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



The results from the SPSS Kaplan Meier for epi 1 and 0 as well as the logrank test

Kaplan-Meier

Case Processing Summary

epi	Total N	N of Events	Censored	
			N	Percent
0	279	89	190	68,1%
1	464	184	280	60,3%
Overall	743	273	470	63,3%

Means and Medians for Survival Time

epi	Mean(a)				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
0	18,844	1,139	16,610	21,077	19,660	1,952	15,833	23,487
1	11,299	,491	10,337	12,261	8,320	,827	6,700	9,940
Overall	15,969	,861	14,282	17,655	13,040	1,566	9,971	16,109

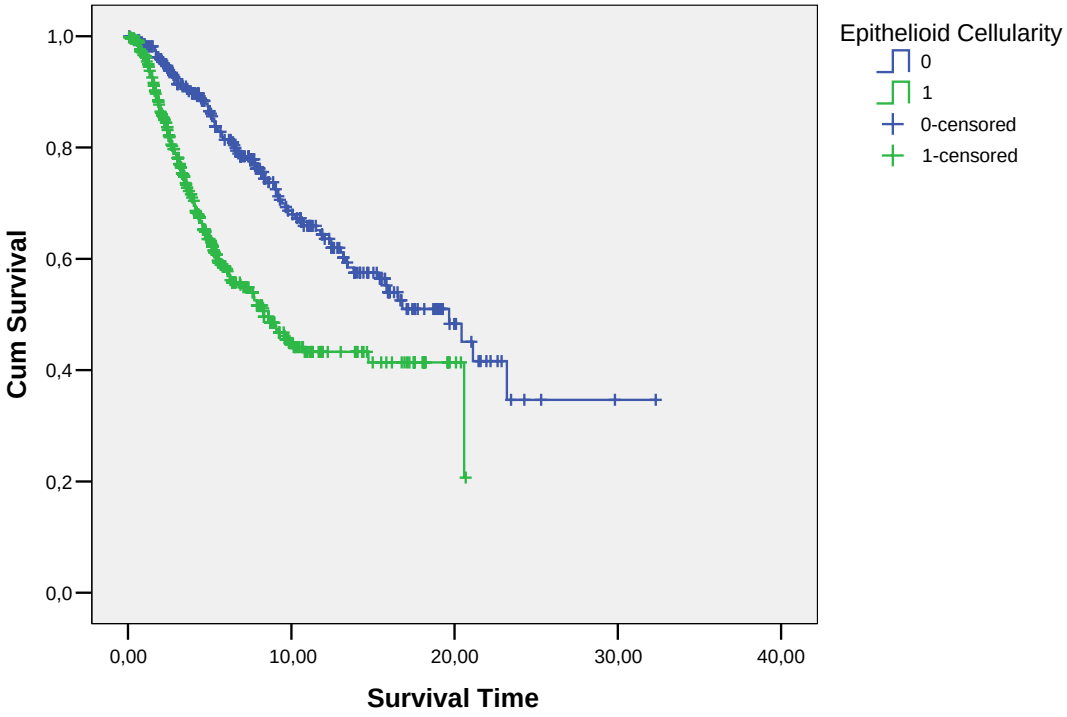
a Estimation is limited to the largest survival time if it is censored.

Overall Comparisons

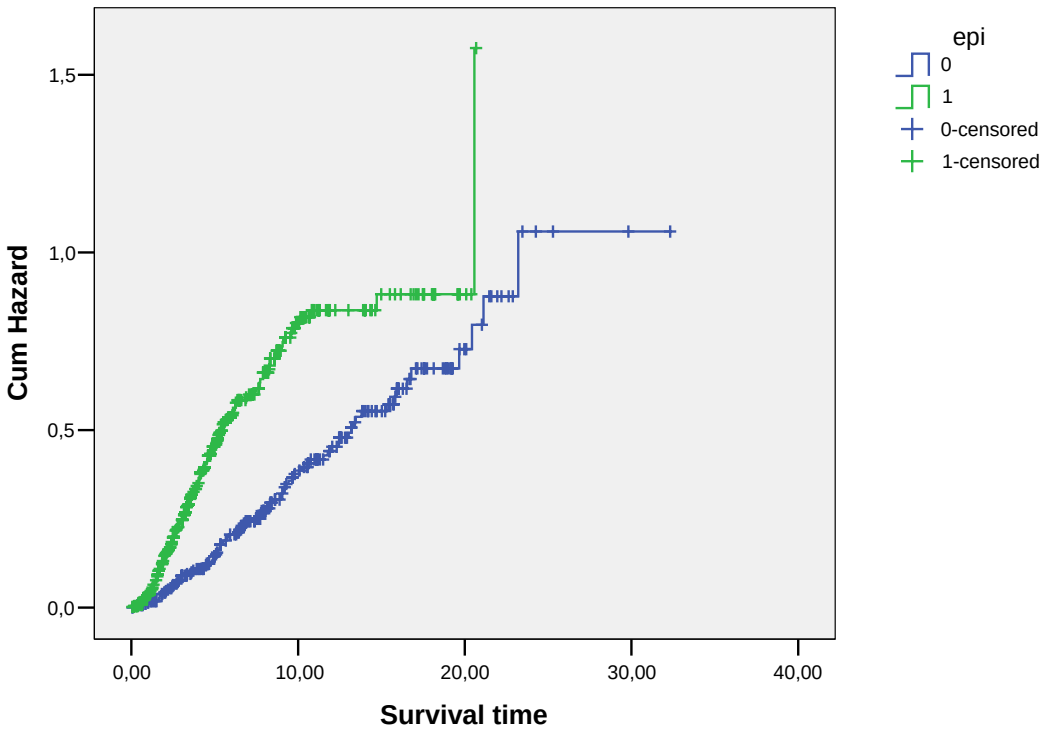
	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	30,525	1	,000

Test of equality of survival distributions for the different levels of V3.

Survival Functions



Hazard Function



The results from the SPSS Cox regression

Case Processing Summary

		N	Percent
Cases available in analysis	Event(a)	273	36,7%
	Censored	470	63,3%
	Total	743	100,0%
Cases dropped	Cases with missing values	0	,0%
	Cases with negative time	0	,0%
	Censored cases before the earliest event in a stratum	0	,0%
	Total	0	,0%
Total		743	100,0%

a Dependent Variable: V5

Block 0: Beginning Block

Omnibus Tests of Model Coefficients

-2 Log Likelihood
3225,927

Block 1: Method = Enter

Omnibus Tests of Model Coefficients(a,b)

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
3130,945	92,791	3	,000	94,982	3	,000	94,982	3	,000

a Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 3225,927
b Beginning Block Number 1. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
V1	,127	,019	44,466	1	,000	1,136
V2	,033	,021	2,490	1	,115	1,034
V3	,664	,133	24,855	1	,000	1,943

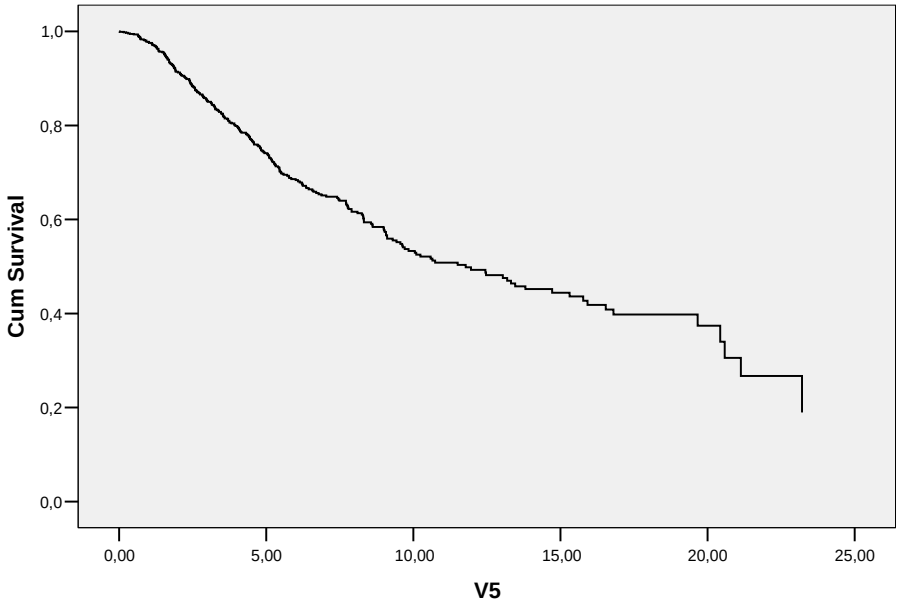
Survival Table

Time	Baseline Cum Hazard	At mean of covariates		
		Survival	SE	Cum Hazard
,02	,000	,999	,001	,001
,16	,000	,998	,002	,002
,24	,000	,997	,002	,003
,31	,000	,996	,002	,004
,37	,001	,994	,002	,006
,50	,001	,993	,003	,007
,64	,001	,992	,003	,008
,65	,001	,990	,003	,010
,67	,001	,989	,004	,011
,70	,001	,987	,004	,013
,72	,001	,985	,004	,015
….	……	…..	…..	…..
16,54	,082	,408	,032	,896
16,80	,085	,398	,033	,921
19,66	,090	,374	,038	,983
20,43	,099	,340	,047	1,079
20,58	,109	,306	,052	1,186
21,13	,153	,267	,057	1,320
23,21	,153	,190	,069	1,663

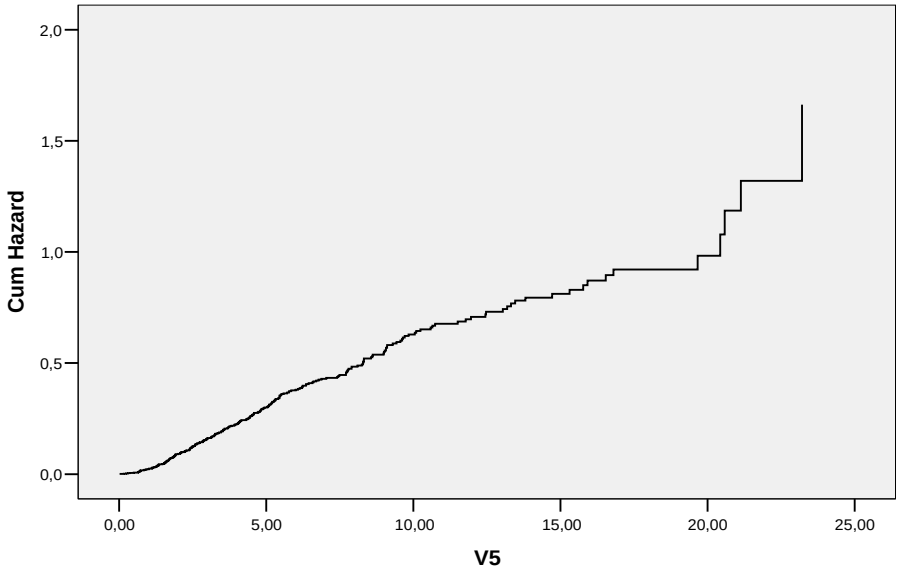
Covariate Means

	Mean
V1	13,538
V2	7,447
V3	,624

Survival Function at mean of covariates



Hazard Function at mean of covariates



The results of the R code for the Breslow estimator

```
  basehaz.gbm(t, delta, f.x,
                  t.eval = NULL,
                  smooth = FALSE,
                  cumulative = TRUE)
 [1] 0.0001017569 0.0001310564 0.0001383812 0.0001896552 0.0001896552
 [6] 0.0002043050 0.0002171620 0.0002300190 0.0003071611 0.0003367694
[11] 0.0003515735 0.0004107899 0.0004628876 0.0004802535 0.0004976194
[16] 0.0005149853 0.0005149853 0.0005311369 0.0005472885 0.0005553642
[21] 0.0006199705 0.0006582002 0.0006658462 0.0006658462 0.0006811381
[26] 0.0006887840 0.0006964299 0.0007117218 0.0007117218 0.0007117218
[31] 0.0007117218 0.0007193678 0.0007270137 0.0007270137 0.0007270137
[36] 0.0009442709 0.0009442709 0.0010537475 0.0010903537 0.0011635659
[41] 0.0013835023 0.0013835023 0.0014940618 0.0014940618 0.0014940618
[46] 0.0014940618 0.0014940618 0.0016056479 0.0016169248 0.0016169248
[51] 0.0016282018 0.0017071404 0.0017184173 0.0017562368 0.0017562368
[56] 0.0018318758 0.0020588912 0.0020588912 0.0020732906 0.0021020894
[61] 0.0021020894 0.0021452876 0.0021596870 0.0021596870 0.0021740864
[66] 0.0021740864 0.0022897423 0.0023482604 0.0023482604 0.0023628900
[71] 0.0023775195 0.0024067786 0.0024067786 0.0024067786 0.0025247743
[76] 0.0026429452 0.0027613462 0.0027812699 0.0027812699 0.0028211172
[81] 0.0028609645 0.0028808882 0.0030007768 0.0030007768 0.0030812418
[86] 0.0031214743 0.0032423680 0.0032423680 0.0032423680 0.0034854365
[91] 0.0034854365 0.0035467956 0.0036081547 0.0036081547 0.0039777124
[96] 0.0039777124 0.0039777124 0.0041018326 0.0041267635 0.0041890910
```

The MATLAB fitting results of Breslow in the Cox Model

FITTEDMODEL =

General model:
FITTEDMODEL(x) = log(1+a*(x)^c)
Coefficients (with 95% confidence bounds):
a = 0.005176 (0.004935, 0.005417)
c = 1.036 (1.016, 1.055)

GOODNESS =
sse: 0.0022721
rsquare: 0.98346
dfe: 238
adjrsquare: 0.98339
rmse: 0.0030897

OUTPUT =
numobs: 240
numparam: 2
residuals: [240x1double]
Jacobian: [240x2double]
exitflag: 1
iterations: 18
funcCount: 55
firstorderopt: 9.9587e-007
algorithm: 'Trust-Region Reflective Newton'

The MATLAB fitting results of the extension of the Kaplan Meier for the Cox Model

FITTEDMODEL =

General model:
FITTEDMODEL(x) = log(1+a*(x)^c)
Coefficients (with 95% confidence bounds):
a = 0.005169 (0.004927, 0.005411)
c = 1.037 (1.017, 1.056)

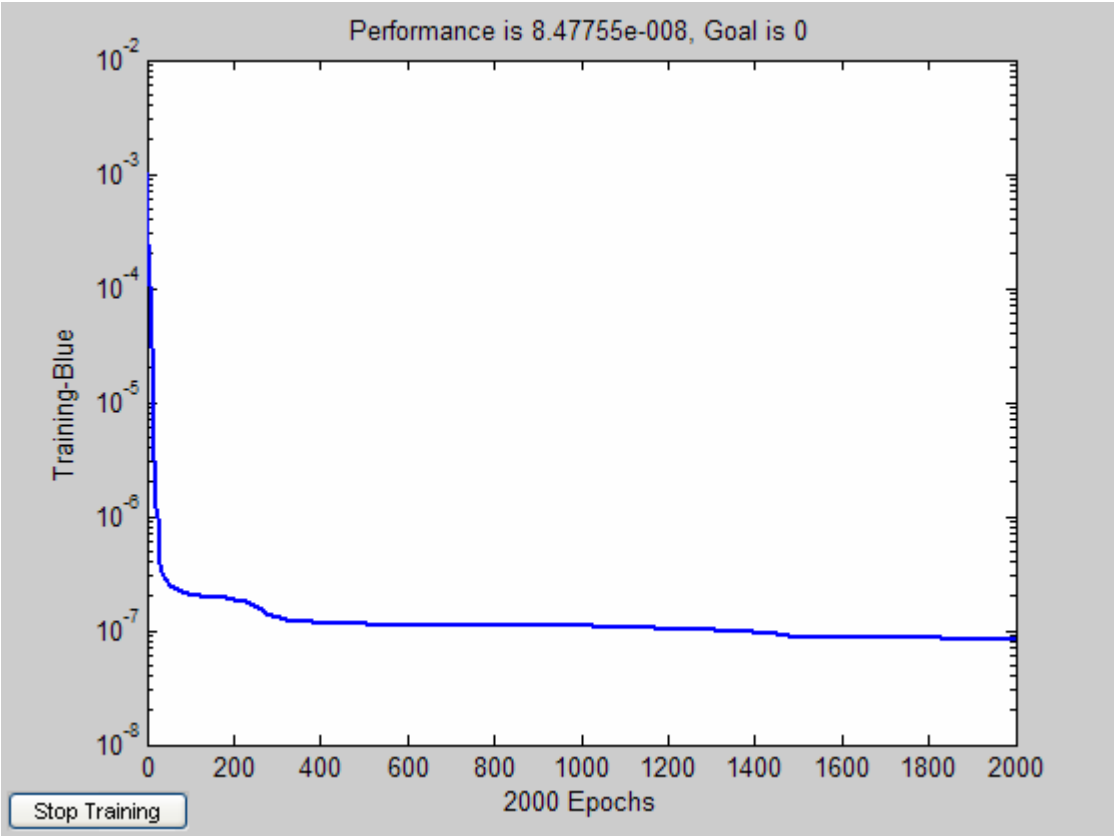
GOODNESS =
sse: 0,0022926
rsquare: 0,98334
dfe: 238
adjrsquare: 0,9832
rmse: 0,0031037

OUTPUT =
numobs: 240
numparam: 2
residuals: [240x1 double]
Jacobian: [240x2 double]
exitflag: 1
iterations: 28
funcCount: 55
firstorderopt: 9,7583e-007
algorithm: 'Trust-Region Reflective Newton'

Baseline network training for 2000 epochs

```
TRAINLM, Epoch 1500/2000, MSE 8.73935e-008/0, Gradient 0.049681/1e-010
TRAINLM, Epoch 1525/2000, MSE 8.68863e-008/0, Gradient 0.0263874/1e-010
TRAINLM, Epoch 1550/2000, MSE 8.65758e-008/0, Gradient 0.0213411/1e-010
TRAINLM, Epoch 1575/2000, MSE 8.63367e-008/0, Gradient 0.0202124/1e-010
TRAINLM, Epoch 1600/2000, MSE 8.61356e-008/0, Gradient 0.0195313/1e-010
TRAINLM, Epoch 1625/2000, MSE 8.59593e-008/0, Gradient 0.0188001/1e-010
TRAINLM, Epoch 1650/2000, MSE 8.58019e-008/0, Gradient 0.0179413/1e-010
TRAINLM, Epoch 1675/2000, MSE 8.56609e-008/0, Gradient 0.0169455/1e-010
TRAINLM, Epoch 1700/2000, MSE 8.55349e-008/0, Gradient 0.0158403/1e-010
TRAINLM, Epoch 1725/2000, MSE 8.54231e-008/0, Gradient 0.0146763/1e-010
TRAINLM, Epoch 1750/2000, MSE 8.53243e-008/0, Gradient 0.0135071/1e-010
TRAINLM, Epoch 1775/2000, MSE 8.52375e-008/0, Gradient 0.0123773/1e-010
TRAINLM, Epoch 1800/2000, MSE 8.51614e-008/0, Gradient 0.0113197/1e-010
TRAINLM, Epoch 1825/2000, MSE 8.50948e-008/0, Gradient 0.010356/1e-010
TRAINLM, Epoch 1850/2000, MSE 8.50364e-008/0, Gradient 0.00949962/1e-010
TRAINLM, Epoch 1875/2000, MSE 8.49849e-008/0, Gradient 0.00875988/1e-010
TRAINLM, Epoch 1900/2000, MSE 8.49392e-008/0, Gradient 0.00814915/1e-010
TRAINLM, Epoch 1925/2000, MSE 8.48979e-008/0, Gradient 0.0076999/1e-010
TRAINLM, Epoch 1950/2000, MSE 8.48599e-008/0, Gradient 0.00752505/1e-010
TRAINLM, Epoch 1975/2000, MSE 8.48226e-008/0, Gradient 0.00819814/1e-010
TRAINLM, Epoch 2000/2000, MSE 8.47755e-008/0, Gradient 0.0184105/1e-010
TRAINLM, Maximum epoch reached, performance goal was not met.
```

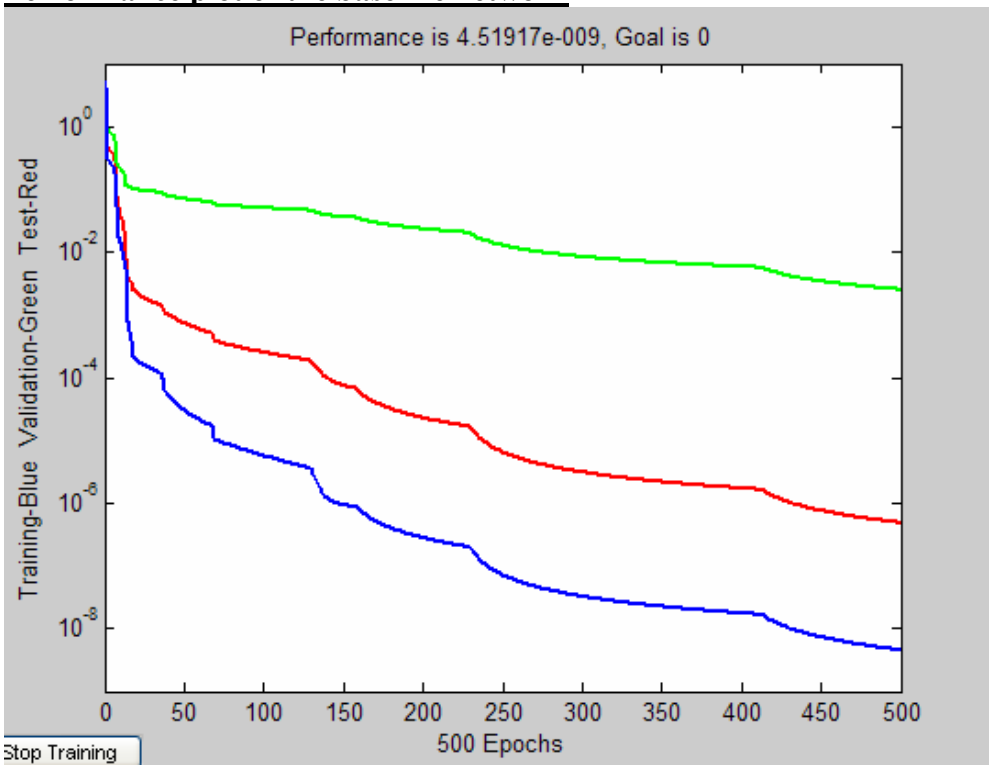
Performance plot of the baseline network



Covariates network training for 500 epochs

```
TRAINLM, Epoch 477/500, MSE 5.43268e-009/0, Gradient 0.00409138/1e-010
TRAINLM, Epoch 478/500, MSE 5.38407e-009/0, Gradient 0.00399926/1e-010
TRAINLM, Epoch 479/500, MSE 5.33645e-009/0, Gradient 0.00391022/1e-010
TRAINLM, Epoch 480/500, MSE 5.28978e-009/0, Gradient 0.00382413/1e-010
TRAINLM, Epoch 481/500, MSE 5.24404e-009/0, Gradient 0.00374086/1e-010
TRAINLM, Epoch 482/500, MSE 5.19921e-009/0, Gradient 0.00366027/1e-010
TRAINLM, Epoch 483/500, MSE 5.15524e-009/0, Gradient 0.00358226/1e-010
TRAINLM, Epoch 484/500, MSE 5.11212e-009/0, Gradient 0.00350671/1e-010
TRAINLM, Epoch 485/500, MSE 5.06981e-009/0, Gradient 0.00343353/1e-010
TRAINLM, Epoch 486/500, MSE 5.02831e-009/0, Gradient 0.00336261/1e-010
TRAINLM, Epoch 487/500, MSE 4.98757e-009/0, Gradient 0.00329387/1e-010
TRAINLM, Epoch 488/500, MSE 4.94759e-009/0, Gradient 0.0032272/1e-010
TRAINLM, Epoch 489/500, MSE 4.90833e-009/0, Gradient 0.00316254/1e-010
TRAINLM, Epoch 490/500, MSE 4.86978e-009/0, Gradient 0.00309979/1e-010
TRAINLM, Epoch 491/500, MSE 4.83192e-009/0, Gradient 0.00303889/1e-010
TRAINLM, Epoch 492/500, MSE 4.79472e-009/0, Gradient 0.00297975/1e-010
TRAINLM, Epoch 493/500, MSE 4.75818e-009/0, Gradient 0.00292232/1e-010
TRAINLM, Epoch 494/500, MSE 4.72227e-009/0, Gradient 0.00286653/1e-010
TRAINLM, Epoch 495/500, MSE 4.68697e-009/0, Gradient 0.00281231/1e-010
TRAINLM, Epoch 496/500, MSE 4.65227e-009/0, Gradient 0.00275961/1e-010
TRAINLM, Epoch 497/500, MSE 4.61816e-009/0, Gradient 0.00270837/1e-010
TRAINLM, Epoch 498/500, MSE 4.58461e-009/0, Gradient 0.00265854/1e-010
TRAINLM, Epoch 499/500, MSE 4.55162e-009/0, Gradient 0.00261006/1e-010
TRAINLM, Epoch 500/500, MSE 4.51917e-009/0, Gradient 0.00256288/1e-010
TRAINLM, Maximum epoch reached, performance goal was not met.
```

Performance plot of the baseline network



Some patients selected to simulate the covariates network and their results

```
------------------------
Item Selected:
   607

->Neural Network output:
    6.7601

->Target Vector:
    6.7598

->Input Vector:
     8
     7
     1
```

```
------------------------
Item Selected:
   123

->Neural Network output:
   27.6755

->Target Vector:
   27.6770

->Input Vector:
   17.8000
   12.0000
    1.0000
```

```
------------------------
Item Selected:
   300

->Neural Network output:
    2.6493

->Target Vector:
    2.5234

->Input Vector:
    1.8000
    1.0000
    1.0000
```

```
------------------------
Item Selected:
   654

->Neural Network output:
   48.2506

->Target Vector:
   48.5212

->Input Vector:
    23
     9
     1
```

References

[1] Sveltana Borovkova (2002) *Analysis of survival data*

[2] Web source: http://www.pitt.edu Yin Bun Cheung, Paul Yip *"Survival Analysis. From Square one to Square two"*

[3]Web source: http://www.cancerbacup.org.uk

[4]Web source http://www.bccancer.bc.ca

[5] Elisa T .Lee, John Wenyu Wang (3rd edition 2003) *Statistical Methods for Survival Data Analysis*. Wiley

[6] Ettore Marubini and Maria Garcia Valsecchi (1995), *Analysing survival datafrom clinical trials and observational studies*, John Wiley and Sons Ltd.

[7] Kevin Gurney (2003),*An introduction to neural networks*"

[8]Howard Demuth, Mark Beale, *Neural Network toolbox, for use with MATLAB*

[9] Web source http://www.cmh.edu/stats

[10] J Martin Bland and Douglas G Altman *The logrank test* (www.bmj.com)

[11] Web source http://members.aol.com/johnp71/prophaz2.html

[12] Kaplan E L. and Meier P. (1958) Non parametric estimation from incomplete Observations. *Journal of the American Statistical Association*

[13] Elandt- Johnson R. C. and Johnson N. L. (1980) *Survival Models and Data Analysis.* Wiley, New York.

[14] Nelson W. (1972) *Theory and Applications of hazard plotting for censored failure data.* Technometrics

[15] Nelson W. (1982) *Applied life data analysis.* Wiley, New York

[16] German Rodriguez 2001, *Non parametric estimation on survival models*. Revised spring 2005

[17] Greenwood J. A. and Durand D. (1960) *Aids for fitting the Gamma Distribution by maximum likelihood*. Technometrics

[18] Greenwood M.(1926) *A report on the natural duration of cancer*. Reports on public Health and Medical Subjects. H. M. Stationery Office London

[19] Aalen O. (1976) *Non parametric inference in connection with multiple decrement models.* Scandinavian Journal of statistics

[20] Aalen O. (1978) *Non parametric inference for a family of counting processes*. The Annals of statistics.

[21] Breslow N. E. (1974) *Covariance analysis of censored survival data*. Biometrics.

[22] Breslow N. E. and Day N. E. (1980) *Statistical methods in cancer research. Vol. 1: The analysis case control studies* . International agency for research on cancer. London.

[23] Breslow N. E. and Day N. E. (1987) *Statistical methods in cancer research. Vol. 2: The design and analysis of cohort studies.* International Agency of research on cancer. Lyon.

[24] Cox D. R. (1972) *Regression models and life tables*. Journal of the Royal statistical society.

[25] Cox D. R. (1975) *Partial likelihood*. Biometrica.

[26] Cox D. R. and Hinkley D. V. (1974) *Theoretical Statistics.* Chapman and Hall. London

[27] Cox D. R. and Oakes D. (1984) *Analysis of survival data*. Chapman and Hall. London

[28] Cox D. R. and Snell E. J. (1968) *A general definition of residuals*. Journal of the Royal statistical society .

[29] Kalbfleisch J. D. and Prentice R. L.(1980) *The statistical analysis of failure time data*. John Wiley and sons. New York

[30] Thomas H. Scheike and Mei-Jie Zhang (2003) *Extensions and applications of the Cox Aalen survival model.* Biometrica

[31] Peter D. Sasieni (1996) *Proportional excess hazards*. Biometrica, Great Britain

[32] Xinghen Yuan (2005). *Survival model and estimation for lung cancer patients*. Thesis presented to the faculty of the department of mathematics. East Tennessee State University.

[33] Bart Baesens, Tony Van Gestel, Maria Stepanova, Dirk Van de Poel. (November 2004) *Neural Network Survival Analysis for Personal Loan Data*. Gent Univarsity, faculty o economy

[34] Ohno Machalo L. (1996). *Sequential use of neural networks for survival prediction in aids.* J Am Med Inf

[35] Ravdin PM and Clark GM (1992) *A practical applications of neural network analysis for predicting the outcome of individual breast cancer patients.* Breast Cancer Res Treat

[36] Biganzoli E. Boracchi P. Mariani L. and Marubini E (1998) *Feed forward neural networks for the analysis of censored survival data: a practical logistic regression approach.* Stat. Med.

[37] Lapuerta P. Azen SP and LaBree L (1995*) Use of neural networks in predicting the risk of coronary artery disease.* Comput. Biomed

[38] Faraggi D. and Simon R (1995*) A neural network model for survival data*. Stat. Med.

[39] Street WN (1998). *A neural network model for prognostic prediction*. In Proceedings of the fifteenth International Conference on Machine Learning , Morgan Kaufmann, Madison, Wisconsin, USA.

[40] Mani DR, Drew J, Betz A and Datta R (1999). *Statistics and data mining techniques for lifetime value modelling*. In Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), ACM, San Diego, CA, USA

[41] Brown SF, Branford A. and Moran W. (1997). On the use of artificial neural networks for the analysis of survival data. IEEE Trans Neural Netw.

[42] Taktak Azzam F. G. Fisher Antony C. and Bertil Damato E. (2004) *Modeling survival after treatment of intraocular melanoma using artificial neural networks and Bayes theorem*. Publishing Ltd, UK.

[43] Isager Peter, Ehlers Niels and Overgaard Jens (2004). *Prognostic factors for survival after enucleation for choroidal and Ciliary body melanomas* Acta Ophthalmologica Scandinavica.

[44] Makitie Teemu (2001) *Prognostic indicators in choroidal and Ciliary body melanoma*. Department of Ophthalmology, University of Helsinki, Finland.

[45] Kourkouta Anna-Maria (2006) *Survival analysis on ocular melanoma patients using non parametric and additive regression models.* Technical University of Crete.