

TECHNICAL UNIVERSITY OF CRETE

**DEPARTMENT OF ELECTRONICS
AND COMPUTER ENGINEERING**



**Survival Analysis on Ocular Melanoma patients using
nonparametric and additive regression models**

UNDERGRADUATE DIPLOMA THESIS

Kourkouta Anna-Maria

Examination Committee:

Zervakis Michalis

Balas Kostantinos

Liavas Athanasios

Chania 2006

Abstract

The primary objective of this work is to perform survival analysis with nonparametric and semiparametric models, as well as with neural networks. Survival analysis is another name for time to event analysis. Usually, death or failure is called an "event" in the survival analysis literature, and so models of death or failure are generically termed time-to-event models. In this thesis, the event that is studied is the death provoked from the ocular melanoma disease. The dataset that contains 1743 patients suffering from ocular melanoma in different parts of the eye, is obtained from the Biopattern group. From the nonparametric methods, the Kaplan-Meier method is used and also the Life-Table and the Nelson-Aalen method. All three of these methods have as primary objective to analyze the data so as to come up with conclusions that concern the survivability of the patients. These conclusions are derived from the survival curves, which are very informative since the probabilities of survival are shown for each time point. The semiparametric model that this study uses is the Aalen's additive model. This model makes use of covariates so as to come up with conclusions about the effect of the covariates in the survivability of the patients. The covariates, or else the prognostic factors, that are used are the longest ultrasound basal dimension, the ultrasound height and the epithelioid cellularity. Except for the information that is given for the effect of the covariates, information about the survivability of the patients can also be given from the analysis of the additive model. Finally neural networks are used in this thesis. They are trained with a known survival distribution, with the Kaplan-Meier one, in order to predict survivability of the patients.

Keywords: Ocular melanoma, Kaplan-Meier, Life-table, Nelson-Aalen, Additive model, Neural networks

*Dedicated to my parents,
who made me who I am with their love*

Acknowledgments

I would like to thank my professor Dr. Michalis Zervakis for his tireless guidance throughout the performance of this thesis. It was really a pleasure working with him because his kindness and his willingness to help me was just I needed during the periods where things were not that good. His scientific advices were very valuable to me and he was always giving me the right motives to continue with my work.

I would like also to thank the other members of the committee for their advice and support during the preparation of this thesis.

I also would like to acknowledge Azzam Taktak and Nicola Lama, even though I did not have the pleasure to meet them from close, for their precious advice and guidance that gave me on matters of survival analysis. Also Biganzoli Elia gave me very useful information at very critical time periods and that is why I would also like to thank him.

I would also like to thank my friend Foteini Agrafioti, who also was a co-worker, for her help and support that offered me throughout this thesis. Her help was very precious to me because she gave it to me at times that I really needed it. Also I would like to thank Kostas Papadakis for his help and ideas.

Table of contents

List of tables.....	8
List of figures.....	10
Chapter 1: Introduction.....	14
1.1 : Information about the ocular melanoma disease.....	14
1.1.1: Description of the ocular melanoma disease.....	14
1.1.2: Causes.....	15
1.1.3: Signs and Symptoms.....	15
1.1.4: How it is diagnosed.....	16
1.1.5: Treatment.....	17
1.1.6: Prevention.....	18
1.2: Overview of survival analysis.....	18
1.2.1 Models that perform survival analysis.....	20
1.2.2: Categorization of the existing models.....	21
1.3: Goal and structure of the thesis.....	30
Chapter 2: Theoretical background.....	32
2.1: Functions of survival time.....	32
2.1.1: Definitions of functions of survival time.....	32
2.1.2: Relationships between the survival functions.....	35
2.2: General description of the models.....	36
2.2.1: Description of the Kaplan-Meier method.....	36
2.2.2: Description of the Life-Table method.....	38
2.2.3: Description of the Nelson-Aalen method.....	38
2.2.4: Description of the additive model.....	39
2.3: Theory about censoring.....	40
2.4: Theory about covariates.....	43
Chapter 3: Model application on specific dataset.....	45
3.1: The Kaplan-Meier method.....	45
3.1.1: Censoring in the ocular melanoma dataset.....	45
3.1.2: Analytical description of the method.....	47
3.1.3: Confidence intervals.....	48
3.1.4: Implementation of the method in the ocular melanoma dataset.....	50

3.2: The Life-Table method.....	52
3.2.1: Analytical description of the method.....	52
3.2.2: Confidence Intervals.....	54
3.3: The Nelson-Aalen method.....	56
3.3.1: Analytical description of the method.....	56
3.3.2: Confidence Intervals.....	56
3.4: The Aalen’s additive model.....	56
3.4.1: Analytical description of the method.....	57
3.4.2: Description of the covariates of the ocular melanoma dataset.....	60
3.4.3: Implementation of the method in the ocular melanoma dataset.....	63
Chapter 4: Artificial Neural Networks.....	67
4.1: Theoretical background of the neural networks.....	67
4.1.1: Transfer functions.....	68
4.1.2: Network Architectures.....	69
4.1.3: General information about networks.....	70
4.1.4: The backpropagation learning rule.....	71
4.2: Implementation of the neural networks in the ocular melanoma dataset.....	74
Chapter 5: Results of the statistical methods and of the neural networks.....	76
5.1: Results from the Kaplan-Meier method.....	76
5.1.1: Fitting procedure of the survival function.....	83
5.2: Results from the Life-Table method.....	88
5.2.1: Fitting procedure of the survival function.....	95
5.3: Results from the Nelson-Aalen method.....	97
5.4: Results from the Aalen’s additive model.....	100
5.4.1: Division of the dataset into two subgroups.....	106
5.4.1.1: Ciliary body.....	108
5.4.1.2: Choroidal.....	113
5.4.2: Survival function.....	118
5.4.2.1: Survivability versus covariates.....	120
5.5: Results from the neural networks.....	125
Chapter 6: Comparisons of the results of this study and from other studies...129	
6.1: Comparisons of the results of all the methods of this study.....	129
6.2: Comparisons with other relevant studies.....	131
6.2.1: Comparisons with studies that use the same dataset.....	131

6.2.1.1: Comparison with study that uses the Cox model and the Kaplan-Meier method.....	131
6.2.1.2: Comparison with study that uses Artificial Neural Networks and Bayes theorem.....	135
6.2.2: Comparisons wit studies that use different data sets.....	139
6.2.2.1: Comparison with a study based on prognostic factors for survival after enucleation.....	139
6.2.2.2: Comparison with a study based on prognostic indicators in choroidal and ciliary body melanoma.....	142
6.2.2.3: Comparison with a study which models survival of uveal malignant melanoma.....	144
Chapter 7: Conclusions and future work.....	148
7.1: Conclusions.....	148
7.2: Future work.....	149
Appendix A.....	151
Appendix B.....	161
References.....	171

List of tables

1. Table 1: Contains the functions of the exponential regression model
2. Table 2: Contains the functions of the Weibull regression model
3. Table 3: Contains the functions of the Log-logistic regression model
4. Table 4: Contains the number of the censored and dead patients
5. Table 5: Contains a sample of four patients of the dataset
6. Table 6: Contains a sample of the Kaplan-Meier analysis
7. Table 7: Contains the format of the Life-table analysis[b-1]
8. Table 8: Contains a system that categorize the dimension of the basal tumor
9. Table 9: Contains the description of the anterior tumor margin covariate
10. Table 10: Contains the two subgroups that can be derived from categorization of the anterior tumor margin covariate
11. Table 11: Contains the number of the patients of the two subgroups
12. Table 12: Contains an example of two patients and their covariates
13. Table 13: Contains the three covariates that will be examined in the Aalen's additive model
14. Table 14: Contains the events that take place in two time point before to be summarized
15. Table 15: Contains the events that take place in two time point after to be summarized
16. Table 16: Contains the characteristics of the neural network that we used in this study
17. Table 17: Contains information from the SPSS about the confidence intervals and the median of the survival function
18. Table 18: Contains the goodness of fit of the exponential distribution in the Kaplan-Meier method
19. Table 19: Contains the goodness of fit of the log-logistic distribution in the Kaplan-Meier method
20. Table 20: Contains the parameters of the log-logistic distribution in the Kaplan-Meier method
21. Table 21: Contains the goodness of fit of the log-logistic distribution in the Life-Table method

22. Table 22: Contains the parameters of the log-logistic distribution in the Life-table method
23. Table 23: Contains the two subgroups that can be derived from categorization of the anterior tumor margin covariate
24. Table 24: Contains the number of the patients of the two subgroups
25. Table 25: Contains the rank of the covariate in influence
26. Table 26: Contains some of the values of the outputs of the neural network and the Kaplan-Meier method
27. Table 27: Contains the basic equations of the Taktak study
28. Table 28: Contains the survival probabilities of the study of Isager

List of figures

1. Figure 1: It shows the type I of censoring[b-1]
2. Figure 2: It shows the type II of censoring[b-1]
3. Figure 3: It shows the type III of censoring[b-1]
4. Figure 4: The anatomy of the eye
5. Figure 5: Diagram that shows the central idea of the neural networks
6. Figure 6: Diagram that shows the hard-limit Transfer function
7. Figure 7 : Diagram that shows the linear transfer function
8. Figure 8: Diagram that shows the sigmoid transfer function
9. Figure 9: Diagram that shows a multilayer feedforward network
10. Figure 10a: The survival curve of the 1743 patients from the Kaplan-Meier method
11. Figure 10b: The survival curve of the 1743 patients from the Kaplan-Meier method without the last death taken into account.
12. Figure 11: The survival curve with the median survival
13. Figure 12a: The cumulative hazard function
14. Figure 12b: The cumulative hazard function without the last death taken into account
15. Figure 13: The survival function from the SPSS program
16. Figure 14: The cumulative hazard function from the SPSS program
17. Figure 15: Survival functions of the male and female patients
18. Figure 16: The fit of the exponential survival distribution in the Kaplan-Meier method
19. Figure 17: The fit of the log-logistic survival distribution in the Kaplan-Meier method
20. Figure 18a: The survival function of the Life-table method from Matlab
21. Figure 18b: The survival function of the Life-table method from Matlab that does not contain the last death.
22. Figure 19: the survival function with the median survival of the Life-table method
23. Figure 20: The survival function of the Life-table method with the SPSS program
24. Figure 21a: The probability density function with Matlab

25. Figure 21b: The probability density function with Matlab without the last death taken into account
26. Figure 22: The probability density function with SPSS
27. Figure 23a: The hazard function with Matlab
28. Figure 23b: The hazard function with Matlab without the last death
29. Figure 24: The hazard function with SPSS
30. Figure 25: The cumulative hazard function of the Life-Table method
31. Figure 26: The fit of the log-logistic survival distribution in the Life-table method
32. Figure 27a: The cumulative hazard function of the Nelson-Aalen method
33. Figure 27b: The cumulative hazard function of the Nelson-Aalen method without the last death taken into account
34. Figure 28: The survival function of the Nelson-Aalen method
35. Figure 29: the survival function with the median survival of the Nelson-Aalen method
36. Figure 30: The plot of the regression function that corresponds to the longest ultrasound basal dimension covariate
37. Figure 31: The plot of the regression function that corresponds to the ultrasound height covariate
38. Figure 32: The plot of the regression function that corresponds to the epithelioid cellularity covariate
39. Figure 33: This diagram gives information about which covariate influences more the disease
40. Figure 34: The plot that contains the survival functions of the two subgroups choroidal and ciliary body.
41. Figure 35: The plot of the regression function that corresponds to the longest ultrasound basal dimension covariate for the ciliary body
42. Figure 36: The plot of the regression function that corresponds to the ultrasound height covariate for the ciliary body
43. Figure 37: The plot of the regression function that corresponds to the epithelioid cellularity for the ciliary body
44. Figure 38: This diagram gives information about which covariate influences more the disease in the ciliary body

45. Figure 39: This figure contains the survival functions that show the survivability of the patients under the presence or absence of epithelioid cells for the subgroup ciliary body
46. Figure 40: The plot of the regression function that corresponds to the longest ultrasound basal dimension covariate for the choroidal
47. Figure 41: The plot of the regression function that corresponds to the ultrasound height covariate for the choroidal
48. Figure 42: The plot of the regression function that corresponds to the epithelioid cellularity for the ciliary
49. Figure 43: This diagram gives information about which covariate influences more the disease in the choroidal
50. Figure 44: This figure contains the survival functions that show the survivability of the patients under the presence or absence of epithelioid cells for the subgroup choroidal
51. Figure 45: This diagram contains the survival functions of fifteen random patients from the additive model
52. Figure 46: The survival function of the entire population of the patients of the additive model
53. Figure 47: Shows how the survival functions are distributed for the covariate epi
54. Figure 48: The survival function for the entire population with the mean value of epi 1 and epi 0
55. Figure 49: Shows the distribution of the survival functions for the covariate “longest ultrasound basal dimension
56. Figure 50: Shows the distribution of the survival functions for the covariate “ultrasound height”
57. Figure 51: Shows the distribution of the survival functions for big lubd - epi 1 and for big lubd - epi 0.
58. Figure 52: Shows the last epochs of the training and the error
59. Figure 53: This diagram shows the epochs of the training and the final performance
60. Figure 54: Shows the correlation coefficient, which is one of the results of the training test

61. Figure 55: Kaplan- Meier estimator for the dataset of ocular melanoma for the 1743 patients and the 743 patients
62. Figure 56: This diagram gives information about which covariate influences more the disease
63. Figure 57: Kaplan Meier survival curve for 2331 patients of intraocular melanoma [p-21].
64. Figure 58: The Kaplan Meier survival curve for the ocular melanoma dataset of our study
65. Figure 59: Survival curves for patients with epithelioid and non epithelioid cells as they derived from the Isager's study. [p-22]
66. Figure 60: The survival function for the entire population with the mean value of epi 1 and epi 0.
67. Figure 61: Diagram that shoes how the survivability is affected by the gender in the study of Makitie [oth-8].
68. Figure 62: Survival functions of the male and female functions
69. Figure 63: Survival curves for patients with epithelioid and non epithelioid Cells, spindle, as they derived from the Makitie's study. [oth-8]
70. Figure 64: Diagram that shoes how the survivability is affected by the gender in the study of Kivela [p-23].
71. Figure 65: Survival curves for patients with presence of epithelioid cells and no presence of epithelioid cells as derived from the Kivela's study. [p-23]

Chapter 1

Introduction

In the beginning of this chapter is given a description of the ocular melanoma disease and after that an overview of the survival analysis and the models that perform survival analysis are presented. Finally the goal and the structure of this thesis are given.

1.1 Information about the ocular melanoma disease

1.1.1 Description of the ocular melanoma disease

The ocular melanoma is a melanoma of the eye and is the most common type of cancer to affect the eye, although, generally, it is still quite rare. [ws-6] In general, melanoma is a very aggressive type of cancer that can spread rapidly and it develops from cells that are called melanocytes. Melanocytes produce the dark-colored pigment melanin, which is responsible for the color of our skin. These cells are found in many places in our body including the skin, the hair and the internal organs. The melanoma of the eye differs in behaviour and management from melanoma arising from skin. [ws-5]

If not treated, it may spread to other parts of the body but rarely this type of cancer starts in another part of the body and spreads to the eye. [ws-4] Between 60% and 80% of patients will survive at least 5 years from the time of diagnosis, if the cancer has not metastasized. If there has been spread outside the eye, the chance of survival is much lower. [ws-8] The incidence of melanoma is highest from ages 50-70 and is almost the same in men and women. Very few cases occur in people under 30 years of age. [ws-2]

The most common type of ocular melanoma is uveal melanoma. This means it occurs along the uveal tract of the eye, which includes the choroid, the ciliary body and the iris. [ws-5] Most common site is the choroid and less frequently ocular melanoma involves the ciliary body, the iris, and the sclera. [ws-7] The behaviour of intraocular melanoma varies with site. [ws-8]

Melanomas of the iris appear to be quite different from melanomas arising elsewhere in the eye or body. This tumour is very slow growing and death rate from it is very low (1 - 4%). [ws-7] Local excision is effective in most of these cases, although some may require enucleation, and metastasis is seen in only 3% of cases. [ws-8] On the contrary, melanoma of the choroid is much more serious because about 60% are alive five years after diagnosis and patients often die from metastases to the liver. [ws-7]

Melanoma can also occur in the thin lining over the white part of the eye (the conjunctiva) or on the eyelid, but this is very rare. [ws-5]

1.1.2 Causes

Ocular melanoma is a rare type of tumor and as for many other forms of cancer the exact cause is unknown. [ws-5] Familial histories of uveal melanoma have been regarded as predisposing factors. A large range of risk factors have been investigated: sex related factors, social class, and socioeconomic factors, lightness of complexion and hair, eye colour, smoking, viruses, chemicals, links with other cancers, but no causative agent has been recognised. Moreover, several studies have addressed the possible importance of genetic factors, following the demonstration of genetically determined predisposition in other cancers. Clear evidence of a genetic component in ocular melanoma is the racial risk with a strong predominance among white people and an absence among dark skinned people. [ws-8] It is also 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 these people are those 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. [ws-5]

1.1.3 Signs and symptoms

Symptoms include: blurred vision, painful, red eyes, flashing lights, loss of side vision in the affected eye, shadows, and misting of the lens of the eye, which is known as cataract. [ws-4,6] Also iris color may change. Often no symptoms are noticed until the tumor is quite large. [ws-5]

1.1.4 How it is diagnosed

- **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. [ws-5]
- **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. [ws-5] This method which may help the doctor to difference a melanoma from another kind of tumor. [ws-4]
- **Slit lamp:** The ophthalmologist can detect a melanoma on the iris using a lighted tool called slit lamp. [ws-4]
- **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 (computerized 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 tumor 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 earplugs or headphones are given 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. [ws-5]

1.1.5 Treatment

A number of different treatments are used for ocular melanoma depending on the size, cell type and position of the tumor, as well as other factors such as the general health, age and level of vision in both eyes. The aim of the treatment is to destroy the cancer cells, stop the cancer coming back, and save as much of the vision as possible.

1. **Radiotherapy:** This type of treatment uses high-energy rays to destroy the cancer cells while doing as little harm as possible to normal cells. Radiotherapy may be given either from outside the body, external radiotherapy, or from within, internal radiotherapy. Radiotherapy may be the only treatment or it may be given after surgery. Recent developments in radiotherapy have made it possible to preserve sight in the eye, either completely or partly.

External radiotherapy: In external radiotherapy a beam of radiation is directed to the area of the tumor. The treatment is normally given as small doses, over a few days. Different types of radiotherapy machines can be used.

Internal radiotherapy: This treatment is given by placing a radioactive source close to the tumor. This normally involves a stay in hospital of up to a week. The radioactive plaque is placed close to the tumor in the eye under general anesthetic. Another operation is carried out to remove it when the treatment is finished

2. **Transpupillary thermotherapy (TTT):** This can be used to treat very small ocular melanomas or as an additional treatment after radiotherapy. The tumor is heated with a special type of laser beam. Cancer cells are more susceptible to heat than normal cells and so will be destroyed. Several treatments are normally needed.
3. **Surgery:** Depending on the size and position of the tumor, it may be possible to remove it without needing to remove the eye. However, if the cancer is growing rapidly, is large or painful, surgery to remove the eyeball may be the most appropriate treatment. This is called enucleation. [ws-5] The enucleated eye cannot be treated or repaired and replaced in the eye socket. Instead, it is

replaced with a ball implant that may be made of plastic or other materials. [ws-3] For many people, this suggestion can be quite shocking and a lot of discussion may be needed with the doctors involved before the decision to go ahead is taken. [ws-5]

1.1.6 Prevention

The most important way to prevent eye melanoma is to avoid excessive exposure to sunlight, especially between 10 a.m and 2 p.m., when the sun's rays are most intense. When exposure cannot be prevented, sunglasses with ultraviolet protection must be used. [ws-9]

1.2 Overview of survival analysis

The origin of survival analysis goes back to mortality tables centuries ago. However, it was not until World War II that a new era of survival analysis came to the surface. This new era was characterized by interest in reliability, or failure time, of military equipment. At the end of the war these newly developed statistical methods became more reliable products, and as the use of survival analysis grew, parametric models gave way to nonparametric and semiparametric models for dealing with the growing field of clinical trials in medical research.

Survival analysis was appropriate for such work because medical studies could start without all patients being present at the beginning of the study and could end before all patients had experienced an event. This was extremely important because even in the best-developed studies, there were patients who chose to quit participating, who moved too far away to follow, or who died from other causes, irrelevant to the event of interest. The researcher was no longer forced to withdraw the patient and all the associating data from the study, due to techniques called censoring. These techniques enabled the researchers to analyze incomplete data. In this way, each patient could contribute all of the information possible to the model for the amount of time the researcher was able to observe the unit. [p-5]

From the above it becomes clear that in survival analysis, censoring is an important concept. A censored observation is an observation for which the exact time to the event is unknown because the event is not observed during the period of observation. There are various kinds of censoring, including, left, right, and interval censoring. If censoring occurs, there is only partial information about a subject's experience of the event under study. A unique feature of survival analysis is that it can make use such incomplete information in the analysis. [ws-24] All the complete information about censoring is given in chapter 2.

It can be said that survival analysis is another name for time to event analysis. Death or failure is called an "event" in the survival analysis literature, and so models of death or failure are generically termed time-to-event models. Time to event analysis has also been used widely in the social sciences where interest is on analyzing time to events such as job changes, marriage, birth of children and so on. [ws-27]

The engineering sciences have also contributed to the development of survival analysis which is called "reliability analysis" or "failure time analysis" in this field, since the main focus is in modelling the time it takes for machines or electronic components to break down. The developments from these diverse fields have for the most part been consolidated into the field of "survival analysis".

Survival analysis attempts to answer questions such as:

- What is the percentage of a population which will survive past a certain time?
- Of those that survive, at what rate will they die or fail?
- Can multiple causes of death or failure be taken into account?
- How do particular circumstances or characteristics increase or decrease the odds of survival? [ws-28]

1.2.1 Models that perform survival analysis

All the above questions are answered by the use of the various models that exist and perform survival analysis. These models are categorized into three categories:

- Nonparametric models: the survival distribution is estimated but without restricting this distribution to be a member of a standard parametric family distribution, possibly because no suitable theoretical distributions are known. Most commonly, the survival function is estimated. [ws-26] These models are usually used to compare similar groups of time-to-event data to determine, for example, whether there is a difference among different treatments. However, nonparametric models cannot be used to estimate the effect of explanatory variables explicitly. They are only applicable to right censored data. If the main objective is to find a model for the data, estimates obtained by nonparametric models and graphs can be helpful in choosing a distribution. [b-1]
- Parametric models: In many clinical studies there are some patients' characteristics, such as prognostic factors and covariates, and it is often difficult to sort out which ones are most closely related to prognosis. These models are needed to prepare a compact summary of the data that can reveal their relationship. [b-1] One way to achieve this purpose is to search for a theoretical distribution, that fits the observed data and identify the most important. Some of the most important parametric models include the exponential distribution, the Weibull, the Gamma, the log normal and the log logistic distribution. So, when covariates are considered, we assume that the survival time, or a function of it, has a relationship with the covariates. Furthermore, when a parametric model is considered, we assume that the survival function follows a given theoretical distribution and has a specific relationship with the covariates.
- Semiparametric models: If the survival distribution is unknown and it is desirable to analyze the associated information, sometimes referred to as covariates, explanatory variables, or independent variables, on survival, then the semiparametric models are the forms of model to use. Such a case is the situation in modelling travel demand for hurricane evacuation. The interest is

in knowing what variables influence the decision to evacuate or not to evacuate and when evacuation will take place, if at all. Such variables may be socio-economic, demographic, or psychological characteristics of the population, or they may be related to the characteristics of the hurricane, or the characteristics of the location of the home of evacuees. [ws-24]

So the main difference between the parametric models and the semiparametric is that in the former the distribution is known, while in the latter the distribution is unknown.

There are various models used in the survival analysis that can be categorised into the above three categories. This is done below and in addition a short description of each one of them is given.

1.2.2 Categorization of the existing models

1. Nonparametric models

- Kaplan-Meier method: it is a method of estimating time-to-event data in the presence of censored cases. It was developed by Kaplan and Meier (1958). Is based on estimating conditional probabilities at each time point when the event occurs. By taking the product limit of those probabilities, the survival rate is estimated at each time. The equation that gives the Kaplan-Meier estimator is: $S(t_i) = \prod_{j=1}^i (1 - \frac{d_j}{n_j})$ where d_j are the dead patients and n_j are the at risk patients.

- Life-Table method: this method is one of the oldest techniques for measuring mortality and describing the survival experience of a population. It looks like the Kaplan-Meier method with the difference that the time is separated into time-intervals. That is why this method requires a large number of observations so that the survival times can be grouped into intervals. The

equation that gives the proportion survival is:
$$S(t_i) = \begin{cases} 1, & i = 0 \\ p_{i-1} \cdot S(t_{i-1}), & i > 0 \end{cases}$$

- Nelson-Aalen method: This method computes directly the cumulative hazard function, in contrast with the Kaplan-Meier method that computes directly the survival function. The equation that gives the Nelson-Aalen estimator is:

$$H(t) = \sum_{t_j < t} \frac{d_j}{n_j} \text{ where } d_j \text{ and } n_j \text{ are as above.}$$

- Relative survival rate or annual survival ratio. The relative survival rate evaluates the survival experience of patients in terms of the general population. Greenwood (1926) first suggested this approach for evaluating the efficacy of cancer treatment. If the average survival time of the patients treated equals that of a random sample of persons of the same age, gender, occupation, and so on, the patients could be considered “cured.” The relative survival rate is defined as the ratio of the survival rate, probability of surviving one year, for a patient under study, observed rate, to someone in the general population of the same age, gender, and race, expected rate, over a specified period of time.[b-1]

2. Parametric models

All the models below follow one of the known distributions that were mentioned above. The survival and the hazard functions of these distributions are extended so as to include the vector z of the covariates. It will be given a table that will contain the survival and the hazard functions of the distribution and of the regression model. In this way a more complete image of the relationship between the distributions and the models will be formed.

- Exponential regression model: this model follows the exponential distribution, which is characterized by a constant hazard rate λ , its only parameter. A high λ value indicates high risk and short survival while a low λ value indicates low risk and long survival. This 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, for example an accident

	Survival function	Hazard function
Distribution	$S(t) = \exp(-\lambda t)$	$h(t) = \lambda$
Regression model	$S(t, z) = \exp(-\lambda t) \cdot \exp(z^T \beta)$	$h(t, z) = \lambda \cdot \exp(z^T \beta)$

Table 1: Contains the functions of the exponential regression model

In the table above as well as in the tables that follow, z is the vector of covariates and β are the regression coefficients [b-1,2]

- Weibull regression model: this model follows the Weibull distribution, which is a generalization of the exponential distribution. The Weibull distribution is characterized by two parameters, λ and γ . The value of λ determines the shape of the distribution curve and the value of γ determines its scaling. Consequently, λ and γ are called the shape and scale parameters, respectively. [b-1,2] This 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.

	Survival function	Hazard function
Distribution	$S(t) = \exp[-(\lambda t)^\gamma]$	$h(t) = \lambda \gamma (\lambda t)^{\gamma-1}$
Regression model	$S(t, z) = \{\exp[-(\lambda t)^\gamma]\}^{\exp(z^T \beta)}$	$h(t, z) = \lambda \gamma (\lambda t)^{\gamma-1} \cdot \exp(z^T \beta)$

Table 2: Contains the functions of the Weibull regression model

- Log-logistic regression model: this model follows the log-logistic distribution, which is characterized by two parameters α , and γ . [b-1,2] This model could be used 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 survived the period of highest risk.

	Survival function	Hazard function
Distribution	$S(t) = \frac{1}{1 + (\alpha t)^\gamma}$	$h(t) = \frac{\alpha \gamma t^{\gamma-1}}{1 + (\alpha t)^\gamma}$
Regression model	$S(t, z) = \frac{1}{1 + (\alpha t)^\gamma \exp(z^T \beta)}$	$h(t, z) = \frac{\alpha \gamma t^{\gamma-1} \exp(z^T \beta)}{1 + (\alpha t)^\gamma \exp(z^T \beta)}$

Table 3: Contains the functions of the Log-logistic regression model

- Accelerated failure time model (AFT): The AFT model for survival time assumes that the relationship of logarithm of survival time T and the covariates is linear and can be written as: $\log T = \beta_0 + \sum_{j=1}^p \beta_j z_j + \sigma \varepsilon$ where $z_j, j = 1, \dots, p$ are the covariates, $\beta_j, j = 0, 1, \dots, p$ are the coefficients, $\sigma (\sigma > 0)$ is an unknown scale parameter, and ε is the error term. It is a random variable with known forms of density function $g(\varepsilon, d)$ and survivorship function $G(\varepsilon, d)$ but unknown parameters d . This means that the survival is dependent on both the covariate and an underlying distribution g . To become more clear why this model is called accelerated, an example will be given. Consider a simple case where there is only one covariate x with values 0 and 1. Then the above equation becomes $\log T = \beta_0 + \beta_1 z + \sigma \varepsilon$. T_0 and T_1 denote the survival times for two individuals with x_0 and x_1 , respectively. Then, $T_0 = \exp(\beta_0 + \sigma \varepsilon)$ and $T_1 = \exp(\beta_0 + \beta_1 + \sigma \varepsilon)$ which becomes $T_1 = T_0 \exp(\beta_1)$. So $T_1 > T_0$ if $\beta_1 > 0$ and $T_1 < T_0$ if $\beta_1 < 0$. This means that the covariate x either “accelerates” or “decelerates” the survival time or time to failure. This is the reason for the name *accelerated failure time model*. [b-1]

Except for the above parametric models, which are the most commonly used, there are some others models that are not so frequently used. These are the Gompertz-Makeham regression model, the Lognormal regression model and the Gamma regression model.

3. Semiparametric models

- Cox model: The most popular form of semiparametric models is the Cox proportional hazards regression model (Cox, 1972). There are two important reasons for the popularity of the Cox model. First, no particular probability distribution needs to be chosen to represent survival times. Second, it is relatively easy to incorporate time-dependent covariates in the model. The hazard function of the Cox model has the following form: $h(t, z) = h_0(t) \cdot \exp(z^T \beta)$ where z are the covariates and β are the regression coefficients. [ws-24]

- Extended Cox model: The extended Cox model is the know model but with the difference that can analyse covariates that depend on time. The hazard function of the Extended Cox model has the following form:

$$h(t, z) = h_0(t) \cdot \exp(z^T(t)\beta(t)).$$
 [b-1]
- Aalen's additive regression model: The additive regression model is an alternative to the Cox model. It results plots that are informative regarding the effect of covariates on survival. The basic equation is the following:

$$h(t, z) = \beta_0(t) + \sum_{j=1}^k \beta_j(t)z_j.$$
 So the hazard at any time is a sum of the baseline hazard $\beta_0(t)$ and a linear combination of the covariates values z_j . The hazard can also be written as: $h(t) = Y(t)\beta(t)$ where the matrix Y contains covariates. [p-1]
- Cox-Aalen Model: The Cox-Aalen model is based on multiplying the two basic models. For this model some covariate effects are believed to result in multiplicative effects, while others effects are better described as additive. The form of the hazard function of the Cox-Aalen is:

$$h(t, z) = \{[Y(t)\beta(t)] \cdot [\exp(z^T \beta)]\}.$$
 [p-19]
- Proportional excess hazards model: The proportional excess hazards model is based on adding the two basic models. The model has the following form :

$$h(t, z) = \{[Y(t)\beta(t)] + [h_0(t) \exp(z^T(t)\beta)]\}$$
 [p-19]
- Stratified Cox model: The stratified Cox model contains different baselines for different strata. The data can be stratified by a covariate, for example sex. If we consider two strata sex: female and sex: male, then we actually have two models: $h_i(t, z) = h_{0i}(t) \exp(\beta z^T_i)$ where $i=1, 2$ for the two sex strata. The baseline survivorship function for each stratum is estimated separately based on the estimated regression coefficients β . [b-1]
- Piecewise Exponential model: The Piecewise Exponential model is a model that has the flexibility of the Cox model and the ability to statistically test the hazard function. In the Piecewise Exponential model, time is divided into intervals. The hazard in each interval is assumed to be constant but can vary across intervals. Let I denote the total number of intervals, and a_0, a_1, \dots, a_{i-1} ,

a_1, \dots, a_l as cut points of intervals, with $a_0 = 0$, and $a_l = \infty$. The hazard can be written as: $h(t) = h_i \exp(z^T \beta)$ where h_i can be considered the baseline hazard, which is the hazard when all the covariates are zero, z and β are vectors of the covariates and corresponding parameters.[ws-24]

- Competing risks model: The most methods deal with a single type of failure time for each study subject. However, in some situations, failure on a person may be due to several distinct causes. The different causes of failure are considered as competing events, so problems with multiple causes are commonly referred to as competing risk problems. The hazard function can be written $h_j(t, z) = h_{0j}(t) \exp(\beta_j z^T)$. The coefficient vector β indicates the effects of the covariates for each event type j . If the covariates are not related to a particular type of event, they are set to 0. If β are the same for all j , the competing risks model reduces to the proportional hazards model (Cox model) [b-1]
- Models for related observations: In Cox's proportional hazards model and other regression methods we make the assumption that the observed survival or the event times are independent. However, in many situations, failure times are observed from related individuals or from successive recurrent events or failures of the same person. For example, in an epidemiological study of heart disease, some of the participants may be from the same family and therefore are not independent. These families with multiple participants may be called clusters[b-1]
- Recurrent event models: Most times events or failures are allowed to occur only once, such as death. When data include recurrences of the same event, such as tumor recurrences after surgeries or successive events of entirely different types such as strokes and heart attacks, then we use the recurrent event model. There are three recurrent event models:
 1. Prentice, Williams, and Peterson (PWP): They proposed two models for recurrent events. Both PWP models can be considered as extensions of the stratified proportional hazards model with strata defined by the number and time of the recurrent events. The hazard function is extended beyond the person's first event to cover subsequent events.

2. Anderson-Gill Model (AG): The model proposed by Andersen and Gill, the AG model, assumes that all events are of the same type and are independent. The risk set of a person at the time of an event would contain all the people who are still under observation, regardless of how many events they have experienced before that time
3. Wei, Lin, and Weissfeld (WLW): Wei, Lin, and Weissfeld proposed also a model for the analysis of recurrent failures. The failures may be recurrences of the same kind of event or events of different natures, depending on how the stratification is defined. If the strata are defined by the times of repeated failures of the same type, similar to the strata defined in the PWP models, it can be used to analyze repeated failures of the same kind. The difference between the PWP models and the WLW model is that the latter considers each event as a separate process and treats each stratum separately. Therefore, each stratum contains everyone in the study. [b-1]
- Artificial Neural Networks: They are non-linear, semiparametric models that are considered as alternative methods for survival analysis in the presence of censorship.[p-18] Most applications of neural networks are classification problems and include diagnosis and prognostic problems. [p-17] Except for the classification problems, neural networks offer the potential of providing more accurate predictions of survival time than do traditional statistical models.[p-16] In order for the neural networks to be used for survival analysis, in the absence or presence of censoring, some new architectures of neural networks have been proposed.
1. Direct Classification: This method considers survival for a fixed time period, and consequently gives a binary classification problem. Censored observations are removed. The neural network output then provides an estimate of the probability that a subject will survive the time period. Above the 50% threshold, the subject is assumed to survive the period. It is clear that this approach is rather basic and does not allow producing individual survival or hazard curves. Furthermore, it does not deal with the problem of censoring and time-varying inputs.[p-7]

2. Ohno-Machado: Ohno-Machado uses multiple neural networks to solve the survival analysis problem. Each neural network has a single output predicting survival at a certain time point. The networks are then trained using their own data subsets consisting of cases that made it to the corresponding time period. Censored observations are included until their time of censoring. So, the number of training instances gradually decreases for the later time intervals making the predictions less reliable. When using these neural networks in isolation, non-monotonic survival curves may result. As a result, the probability of a person surviving two periods could be greater than the probability to survive one period because the interdependencies of the survival probabilities over time are not properly taken into account when isolated neural networks are used.[p-8]
3. Ravdin and Clark: Ravdin and Clark use a multi-layer feed-forward neural network with a single output unit representing the survival status. A time indicator and a survival status indicator are added to each record. The time indicator then records the successive time periods $[1, T_{max}]$ for which a prediction is to be made, with T_{max} the maximum time of follow-up. An uncensored input is then considered T_{max} times whereas a censored input is considered t times with t being the time of censoring. The survival status is the target of the network and is set to zero as long as the subject is alive and to 1 otherwise.[p-9]
4. Biganzoli et al.: A variation on the approach of Ravdin and Clark was suggested by Biganzoli et al. They also trained a neural network with one output and an additional time indicator input. However, unlike Ravdin and Clark, uncensored subjects are only considered for the time intervals in which they were actually observed. So, subjects that have died are not included after the time interval of death. Time dependent inputs might be easily included since each subject has multiple input vectors which may change across the intervals of observation. The neural network predicts discrete hazard rates. [p-10]

5. Lapuerta et al: Lapuerta et al suggested a multi-network strategy to deal with the survival times for the censored cases. For each time period considered, a separate neural network is constructed. These networks are trained using only the observations for which the survival status for the corresponding time period is known. Subsequently, the trained networks are used to predict the outcome for the censored cases. The non-censored observations are then provided for training the principal neural network, which predicts the probability of survival for each time period considered. It is clear that this approach is not suitable for large-scale applications since one need to train as many neural networks as there are time periods considered.[p-11]
6. Faraggi: Faraggi proposes a neural network extension of the Cox proportional hazards model by replacing the linear function $z^T \beta$ in $h(t, z) = h_0(t) \cdot \exp(z^T \beta)$ by the output $g(z, \theta)$ of a neural network with a single, logistic hidden layer and a linear output layer and so the hazard function takes the following form: $h(t, z) = h_0(t) \exp[g(z, \theta)]$. Analogous to the Cox model, no bias input is considered for the output layer since this is incorporated into the baseline hazard $h_0(t)$. This method allows preserving all the advantages of the classical proportional hazards model. [p-12]
7. Street: Street uses a multilayer Perceptron with Tmax output units to deal the survival analysis problem, where Tmax represents the maximum time horizon of the study. A hyperbolic tangent activation function is used in the output layer such that all output neurons take on values between -1 and $+1$. The first output neuron having a value < 0 is considered to be the output neuron that predicts the event time. If all output neurons have values > 1 , then the patient is considered to survive the entire time period of the study. The output units represent the survival probability for the corresponding time period. For the non-censored cases, the output values are set to $+1$ as long as the patient is alive and to -1 after. For the censored cases, the output units are also set to $+1$ until their censoring time. After this period, Street uses the Kaplan-Meier estimator. The Kaplan-Meier hazards are then used to compute the survival probability of the censored observations after the censoring time. These probabilities are then scaled to the range of the hyperbolic tangent function.

The individual survival curve of an observation can then be derived based on the activation values of the output units.[p-13]

8. Mani: A variation on the method of Street was developed by Mani. Again, for every observation in the training set, Tmax output units are computed. But now, these output units now represent the hazard rate instead of the survival probabilities that were used in the approach of Street. For uncensored observations, the hazard is set to zero until the time of death and 1 after. For censored observations, the hazard is set to zero until censoring time and to the Kaplan-Meier estimator after.[p-14]
9. Brown et al: Analogous to Mani, Brown suggests a single neural network with multiple outputs to predict hazard rates. For the non-censored observations, the network output is set to 0 as long as the subject is alive and to 1 when the subject undergoes the event. For the time intervals following the event, the hazard can take any value. The output values for the censored observations are set to 0 until the time of censoring and can take any value for all the time intervals that follow. Then, the neural network is trained to minimize the sum of squared error criterion and to perform weight updates when the hazard can take any value by setting the corresponding errors to 0. [p-15]

1.3 Goal and structure of the thesis

The goal of this study is to implement some of the models that were presented above so as to come up with conclusions that concern the survivability of the patients in the absence and in the presence of covariates. In order to achieve the former, nonparametric models were used that do not make use of the covariates. From the above nonparametric models that were presented, we implemented the Kaplan-Meier method, the Life-table method and the Nelson-Aalen. In order to come up with conclusions that concern the survivability of the patients in the presence of covariates, the Aalen's additive model was used. Finally, we also used artificial neural networks, as an alternative to survival analysis. All the methods that are implemented in this study use the same dataset of the ocular melanoma disease.

The thesis is organised as follows: in chapter 1 we give a description of the ocular melanoma disease and we present some of the most known models that are used in survival analysis, categorised in three main categories: nonparametric, parametric and semiparametric. Also the goal and a structure of this thesis are given. In chapter 2 the theoretical background of the functions of survival time and of the models that are used in this thesis is given. In addition the theoretical background for the covariate selection and censoring is given. In chapter 3, it is presented the application of the models that are used in this study. This means that it is given a more analytical description of the models, it is explained the covariate selection that was done for the dataset of the ocular melanoma disease and finally the censored data of the same dataset are described and explained. In chapter 4, the theoretical background of the artificial neural networks is presented and also the description of neural network that we used is given. In chapter 5 the results of all the methods and the neural networks are presented and a description of these results is given. In chapter 6, comparisons of this study and of others studies are given. The other studies either make use of the same dataset and analyse the data with different methods, such as Cox, either use different datasets of the same disease and analyse the data with the same models that are used in this study, such as the Kaplan-Meier method. Finally, in chapter 7, the conclusions of this thesis are given and also some suggestions for future work are also proposed.

Chapter 2

Theoretical background

This chapter contains all the basic theoretical background that will be needed in the chapters that follow. This means that the theoretical background about the functions of survival time, about the models that are used in this thesis, about censoring and finally about the covariates is given.

2.1 Functions of survival time

As it was mentioned in chapter 1, survival data arise when the aim is to study the time elapsed from some particular starting point of the occurrence of an event. The starting point of observation is usually a medical intervention such as first diagnosis of a given disease, a surgical intervention or the beginning of a treatment in a clinical study. [b-2] The survival time is actually the time up to a certain event. Such event may be a failure, a death, a relapse, or the development of a given disease. These times are randomly distributed and their distribution is usually described by three functions:

1. the survivorship function
2. the probability density function
3. the hazard function.

These three functions are mathematically equivalent, because if one of them is given, the other two can also be derived. The basic problem in survival analysis is to estimate from the sampled data one or more of these three functions, and to draw inferences about the survival pattern in the population.

2.1.1 Definitions of the functions of survival time

Let T denote the survival time. The distribution of T , as mentioned above, can be characterized by three equivalent functions.

- ***Survivorship Function (or Survival Function)***

This function, denoted by $S(t)$, is defined as the probability that an individual survives longer than t :

$$\begin{aligned} S(t) &= P(\text{an individual survives longer than } t) \\ &= P(T > t) \end{aligned} \tag{1}$$

or from the definition of the cumulative distribution function $F(t)$ of T ,

$$\begin{aligned} S(t) &= 1 - P(\text{an individual fails before } t) \\ &= 1 - F(t) \end{aligned} \quad (2)$$

where $F(t)$ = proportion of individuals failing before time t

$S(t)$ is a nonincreasing function of time t with the following properties:

$$S(t) = \begin{cases} 1 & \text{for } t = 0 \\ 0 & \text{for } t = \infty \end{cases}$$

meaning that the probability of surviving at least at time zero is 1 and that the probability of surviving at an infinite time is zero.

The function $S(t)$ is also known as the cumulative survival rate. The graphic representation of $S(t)$ is called the survival curve and it was recommended by Berkson (1942). If there are no censored observations, the survival function is estimated as the proportion of patients surviving longer than t :

$$S(t) = \frac{\text{number of patients surviving longer than time } t}{\text{total number of patients}} \quad (3)$$

When there are censored observations the numerator of cannot always be determined, and that is why nonparametric methods are used in order to calculate the survival function.

- **Probability Density Function (or Density Function)**

The survival time T has a probability density function defined as the limit of the probability of failure in a small interval per unit time. It can be expressed as:

$$f(t) = \frac{\lim_{\Delta t \rightarrow 0} P[\text{an individual dying in the time } (t, t + \Delta t)]}{\Delta t} \quad (4)$$

The density function has the following two properties:

- $f(t)$ is a nonnegative function:

$$f(t) = \begin{cases} \geq 0 & \text{for } t \geq 0 \\ 0 & \text{for } t < \infty \end{cases}$$

- The area between the density curve and the t axis is equal to 1.

If there are no censored observations, the probability density function is estimated as the proportion of patients dying in an interval per unit width:

$$f(t) = \frac{\text{number of patients dying in the time period beginning at time } t}{(\text{total number of patients}) * (\text{width of time period})} \quad (5)$$

Like the estimation of $S(t)$, when censored observations are present, the numerator cannot be always determined. The graph of $f(t)$ is called the density curve.

- **Hazard Function**

The hazard function $h(t)$ of survival time T gives the conditional failure rate. This is defined as the limit of the probability that an individual fails in a very short interval, $t + \Delta t$, given that the individual has survived to time t :

$$h(t) = \frac{\lim_{\Delta t \rightarrow 0} P[\text{an individual fails in the time interval } (t, t + \Delta t) \text{ given the individual has survived to } t]}{\Delta t} \quad (6)$$

The hazard function can also be defined in combination with the cumulative function $F(t)$ and the probability density function $f(t)$:

$$h(t) = \frac{f(t)}{1 - F(t)} \quad (7)$$

If there are no censored observations the hazard function is estimated as the proportion of patients dying in an interval per unit time, given that they have survived to the beginning of the interval:

$$h(t) = \frac{\text{number of patients dying per unit time in the interval}}{\text{number of patients surviving at } t} \quad (8)$$

The cumulative hazard function is defined as:

$$H(t) = \int_0^t h(x) dx \quad (9)$$

2.1.2 Relationships between the survival functions

The three functions defined above are mathematically equivalent. If any one of them is given, the other two can be derived.

From (2) and (7):

$$\left. \begin{aligned} S(t) &= 1 - F(t) \\ h(t) &= \frac{f(t)}{1 - F(t)} \end{aligned} \right\} 1 - F(t) = \frac{f(t)}{h(t)} \Leftrightarrow S(t) = \frac{f(t)}{h(t)} \Leftrightarrow h(t) = \frac{f(t)}{S(t)} \quad (10)$$

The probability density function is the derivative of the cumulative distribution function and as a result the following equation is derived:

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

$$f(t) = \frac{d}{dt}[1 - S(t)] = -S'(t) \quad (11)$$

From the two above equations it is derived that:

$$\left. \begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ f(t) &= -S'(t) \end{aligned} \right\} h(t) = -\frac{S'(t)}{S(t)} = -\log S(t) \quad (12)$$

Integrating the above equation from zero to t and using $S(0) = 1$, it is obtained that:

$$-\int_0^t h(x)dx = \log S(t) \quad \text{or} \quad (13)$$

$$H(t) = -\log S(t) \quad \text{or} \quad (14)$$

$$S(t) = \exp[-H(t)] = \exp\left[-\int_0^t h(x)dx\right] \quad (15)$$

Lastly from equations (10) and (15) it is obtained that:

$$\left. \begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ S(t) &= \exp[-H(t)] \end{aligned} \right\} f(t) = h(t) * \exp[-H(t)] \quad (16)$$

As it was proven from all the above equations if one of the three survival functions is given, then easily the remaining two can be derived. [b-1]

2.2 General description of the models

As it was mentioned above, this study uses some of the models that were presented. The statistical models that we use are the Kaplan-Meier model, the Life-table model, the Nelson-Aalen model and the Aalen's additive model. Below it will be given a basic description of these models. A more analytical description as well as the mathematical background of these models, it will be given in the following chapter. We also mentioned that this study uses neural networks. The description of the neural network that we use in this study is given in chapter 4.

2.2.1 Description of the Kaplan-Meier method

The Kaplan-Meier estimator is a staircase function and is so widely used, that in research papers survival curves are more often called Kaplan-Meier curves. [17] The Kaplan-Meier estimator is used to calculate an estimate of cumulative survival that can then be used to calculate the cumulative hazard rate. At the moment of each death, the proportion of survival decreases and it does not change at any other time. So the curve steps down at each death and is flat between deaths, which lead to the staircase appearance. [ws-23]

It is a method of estimating time-to-event models in the presence of censored cases and it assumes a single cause of failure. The concept of censoring will be explained later in this chapter. The Kaplan-Meier estimator is based on estimating conditional probabilities at each time point when an event occurs and taking the product limit of those probabilities to estimate the survival rate at each point in time. That is why Kaplan-Meier is also called the "product limit estimate." [ws-20]. It involves computing the number of people who died at a certain time point, divided by the number of people who were still in the study at that time. [ws-22] The survival probability is calculated as number surviving divided by number at risk. Patients who have died or dropped out are not counted as "at risk". Patients who are lost are considered censored and are not counted in the denominator. [ws-17] .The equation that gives the Kaplan-Meier estimator is:

$$S(t_i) = \prod_{j=1}^i \left(1 - \frac{d_j}{n_j}\right)$$

A more analytical mathematical analysis will be given in chapter 3. The Kaplan-Meier method is a nonparametric technique and in order to be implemented, exact times of the event that is studied are required. As it was mentioned above, the survival time is the time up to a certain event. Kaplan-Meier method is used mostly to analyze death as event, but it may also be used effectively to analyze time to an endpoint, such as remission. For example instead of analyzing death, it may be used to analyze the remission of a disease.

The Kaplan-Meier survival curve is often illustrated graphically. It looks like a poorly designed staircase, with vertical steps going down at the time of death of each individual subject. [ws-22]

With staircase curves, as the group of patients is larger, the step down caused by each death is smaller and so the staircase will become closer and closer to the ideal of a smooth curve. [ws-23]

Summarizing, Kaplan-Meier estimates the instantaneous risk of death as the number dying in time interval divided by the number currently at risk of experiencing outcome. Moreover,

1. It considers all observations, failures and censored.
2. It considers the ordered sets of all time intervals and estimates the probability of survival in each interval, given that the patient survived through all prior intervals.
3. Time intervals are as small as possible. At each failure, the survivorship function drops in a vertical step, forming a staircase down from 1.0 to the lowest value at the end of the study.
4. It is nonparametric – Kaplan-Meier uses the data to draw the survival function directly.
5. There is absence of competing risk.

All the above characteristics will be confirmed with the mathematical analysis that will be given in chapter 3.

The properties of the survival curve are:

1. Step function: if there are small samples then the steps are big
2. As sample size increases : it gets smoother
3. $S(0) = 1$
4. It is a decreasing function [ws-19]

2.2.2 Description of the Life-table method

The life-table method is one of the oldest techniques for measuring mortality and describing the survival experience of a population. The life tables, summarizing the mortality experience of a specific population for a specific period of time, are called population life tables. As clinical and epidemiologic research become more common, the life-table method has been applied to patients with a given disease who have been followed for a period of time. Life tables constructed for patients are called clinical life tables.

The basic idea of the life table is to subdivide the period of observation into smaller time intervals. For each interval, all people who have been observed at least that long are used to calculate the probability of a terminal event occurring in that interval. The probabilities estimated from each of the intervals are then used to estimate the overall probability of the event occurring at different time points. The mathematical analysis will be given in chapter 3.

The Life-Table method requires a fairly large number of observations, so that survival times can be grouped into intervals. Such as the Kaplan-Meier estimator, the Life-Table estimator makes use of all the survival information until the end of the study. [b-1]

2.2.3 Description of the Nelson-Aalen method

It is a method for the nonparametric estimation of the cumulative hazard function. It gives exactly the same cumulative hazard function with the Kaplan-Meier method but it is a more direct way.

With the Kaplan-Meier method, the survival function is primary computed and then with the relationship $H(t) = -\log S(t)$, the cumulative hazard function was computed. With the Nelson-Aalen method the procedure is exactly the opposite, since first the cumulative hazard function is computed and with the relationship $S(t) = \exp[-H(t)]$, the survival function is computed.

This estimate of $S(t)$ is sometimes called the Fleming-Harrington estimator of the survivor function. Usually, there will not be much difference between the Fleming-Harrington and Kaplan-Meier estimators. Since the same data set is used, it is expected that the cumulative hazard and survival function would be the same. [ws-26]

The risk set is calculated at every point in time where at least one event occurred which means that all events must be sorted according to their ending times. The mathematical analysis of this method will be given in chapter 3.

It is a staircase function with:

- a. Location of drop is random (time at event)
- b. Size of drop is $1/\text{risk set}$ (number at risk: count of persons alive before the death)

2.2.4 Description of the Aalen's additive model

The additive hazard model was suggested for the influence of the covariates on the hazard function. This approach is less vulnerable than the Cox model to problems of inconsistency when some of the covariates are deleted or the precision of some covariates measurements is changed. This method results in plots that may give information on the change over time in the influence of covariates. This information is taken from the change of the slopes of these plots.

The Cox proportional hazards model is the more used model for regression analysis of survival data. The most important aspect of the Cox model is the fact that the hazard function is the basis of the regression model. More importantly, the hazard is a natural function for describing the distribution of life times. Basically the hazard function measures the risk that an event may happen at any given time. This risk is a natural concept and that is why is reasonable to explain the level of risk in terms of covariates.

However there are some weaknesses in the Cox model:

1. The assumptions of the model may not always hold.
2. Changes over time in the influence of covariates are not very easily discovered.
3. The proportional hazard assumption is vulnerable to changes in the number of the covariates and to the accuracy of their measurement. This means that if covariates are deleted from a model or are measured with a different level of accuracy the proportionality is generally destroyed.

The linear model that will be presented below is not supposed to solve all the problems mentioned before, but it may give a partial solution to the second and third problems. It is an alternative to the Cox model but this model does not require

constant proportional hazard. In this model the covariates are modeled as additive risks to a baseline hazard.

An obvious weakness of this linear model and one that basically distinguishes it from the Cox model is that the expression of $h_i(t)$ is not restricted naturally to non-negative values. In this respect the model is not a “natural” one, but this must be weighed against the advantages of linearity. One practical consequence of the lack of restriction to non-negative numbers is that estimated survival functions may not be monotonically decreasing throughout, but can have occasional lapses where they increase slightly. This is a weakness but it is not a great problem unless the data set is quite small or the chosen covariate values are rather extreme. In the data set of ocular melanoma, this is not a problem because the covariate values do not have extreme values and the data set is not small. [p-1]

As in the above models that were described, a more analytical description as well as the mathematical analysis of this model, will be given in chapter 3.

2.3 Theory about censoring

Censoring in survival analysis is a very important issue because in most of the studies, the data that are used in the analysis are not complete. This happens due to various reasons that will be explained analytically below.

There is an important assumption in Survival Analysis that individuals who are censored are at the same risk of failure as those who are still alive and uncensored. The risk set at any time point, the individuals 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 is noninformative, that is those censored at a given time are a representative sample of those at risk, then the estimates of the statistical methods are not biased. However, if the censoring is informative, that is those censored tend to be at a higher or lower risk of failure, then the estimates are biased and the results of the analysis are not representative. If at all possible, censoring from these later situations should be kept to a minimum. [p-6]

In general censored are called the patients that have not experienced the event of interest at the end of the study or time of analysis. For example, some patients may still be alive or disease-free at the end of the study period. Censored are also called the patients that are lost to follow-up after a period of study. This fact did not affect the outcome because all the non-parametric methods that were developed make use of the censored patients. There are three different types of censoring which are summarised below:

- In **type I** censoring, if there are no accidental losses, all censored observations equal the length of the study period.

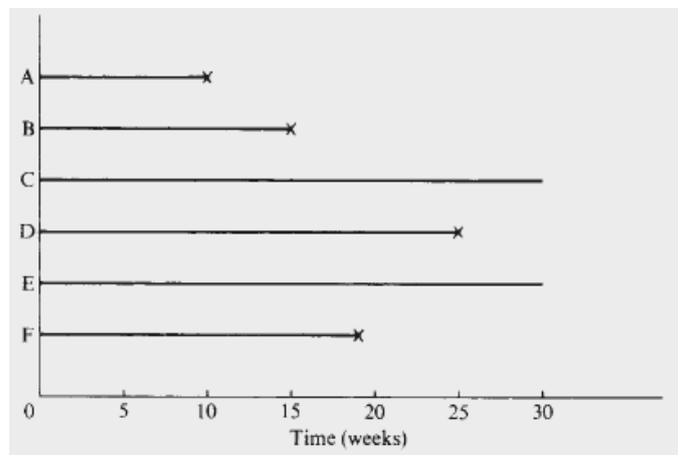


Figure 1: It shows the type I of censoring [b-1]

For example, suppose that six rats have been exposed to carcinogens by injecting tumor cells into their foot pads. The times to develop a tumor of a given size are observed. The experiment is terminated after 30 weeks. The above figure is a plot of the development times of the tumours. Rats A, B, and D developed tumours after 10, 15, and 25 weeks, respectively. Rats C and E did not develop tumours by the end of the study; their tumor-free times are thus 30-plus weeks. Rat F died accidentally without tumours after 19 weeks of observation. The survival data are 10, 15, 30+, 25, 30+, and 19+ weeks. (The plus indicates a censored observation.)

- In **type II** censoring, if there are no accidental losses, the censored observations equal the largest uncensored observation.

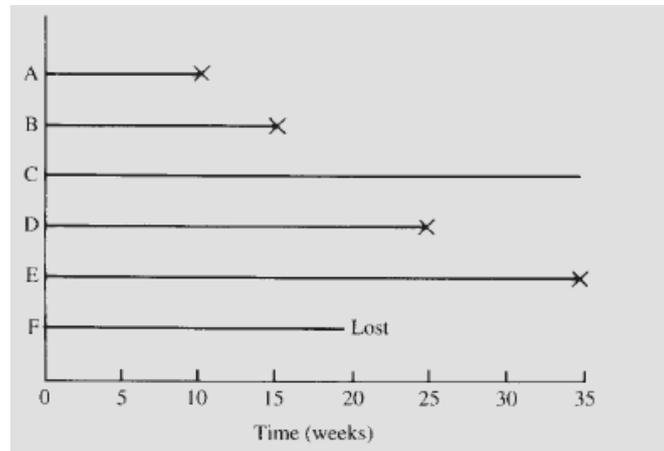


Figure 2: It shows the type II of censoring [b-1]

For example, in an experiment of six rats, the study was terminated after four of the six rats have developed tumours. The survival times are then 10, 15, 35+, 25, 35, and 19+ weeks.

- **Type III**: In most clinical studies the period of study is fixed and patients enter the study at different times during that period. Some may die before the end of the study and so their exact survival times are known but others may withdraw before the end of the study and are lost. Still others may be alive at the end of the study. For those lost patients, survival times are at least from their entrance to the last contact. For patients still alive, survival times are at least from entry to the end of the study. The latter two kinds of observations are censored observations. Since the entry times are not similar, the censored times are also different. [b-1]

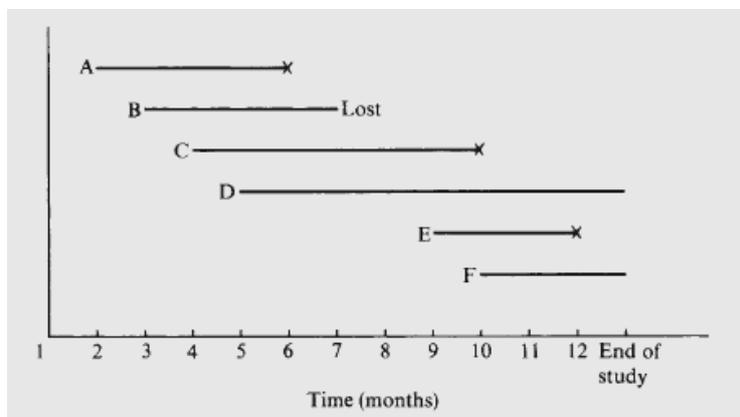


Figure 3: It shows the type III of censoring [b-1]

For example, suppose that six patients enter a clinical study during a total study period of one year. Suppose also that all six respond to treatment and achieve remission. The remission times are plotted in the above figure. Patients A, C, and E achieve remission at the beginning of the second, fourth, and ninth months, and relapse after four, six and three months, respectively. Patient B achieves remission at the beginning of the third month but is lost to follow-up four months later, the remission duration is thus at least four months. Patients D and F achieve remission at the beginning of the fifth and tenth months, respectively, and are still in remission at the end of the study their remission times are thus at least eight and three months. The respective remission times of the six patients are 4, 4+, 6, 8+, 3, and 3+ months. Another commonly used name for type III censoring is random censoring. All of these types of censoring are right censoring.

There are also left censoring and interval censoring cases. Left censoring occurs when it is known that the event of interest occurred prior to a certain time t , but the exact time of occurrence is unknown.

Interval censoring occurs when the event of interest is known to have occurred between times a and b . [b-1]

2.4 Theory about covariates

As with many statistical analyses, interest in survival analysis may be focussed on modelling dependence of the survival time with possible explanatory variables, which are also called covariates.

In survival analysis, the value of a covariate may take a fixed value, that is its value is fixed at the start of the study and do not change throughout the duration of the study, for example sex or treatment. However, in survival analysis, covariates may also be time-varying, taking different values at different time points, for example blood pressure or weight. [ws-25]

The three main categories of the covariates are the following:

- **Numerical:** taking values in an interval (blood pressure).
- **Categorical:** taking a finite number of values (Civil Status).
- **Binary:** a categorical covariate that takes two values (Sex). [ws-30]

Covariates may also be classified as group or individual.

- group covariate: variable that applies to a group of individuals (temperature, environmental factors)
- individual covariate: variable that applies to an individual animal (weight, hormone level)

Both types of covariates can be time variant (body weight) or time invariant (genetic characteristics) [ws-29]

Chapter 3

Model application on specific dataset

This chapter describes in full details the four statistical models that we use in this study. These models, as we have already mentioned, are the Kaplan-Meier model, the Life-table, the Nelson-Aalen and the Aalen's additive model.

3.1 The Kaplan-Meier method

As it was mentioned above, the Kaplan-Meier method is a method of estimating time-to-event data in the presence of censored cases. Censoring is a very important issue in this method and that is why below it will be explained analytically how censored observations appear in our data and how censoring affects the shape of the survival curve practically.

3.1.1 Censoring in the ocular melanoma dataset

The Kaplan-Meier method in this study uses a dataset that contains patients who suffer from the ocular melanoma disease. The data set was made of 1734 patients with ocular melanoma in different parts of the eye. However, many of these patients turned out to be censored because they were still alive until the end of the study.

In censoring type III, patients enter the study at different times during the period of study. For patients still alive, survival times are at least from entry to the end of the study. Since the entry times are not similar, the censored times are also different. It is obvious that the censored patients from our data belong to the third type of censoring. Each patient has entered the study at different time, the date of management, and that is why each one has different survival time. The survival time, for these patients is from entry until a fixed date, the 10th February 2005.

When a patient is censored the curve doesn't take a step down as it does when a patient dies. But because censoring the patient reduces the number of patients who are contributing to the curve, each death after that point represents a higher proportion of the remaining population, and so every step down afterwards will be a little bit larger than it would have been without censoring.

The part of the curve after the first patient is censored is only an estimate of survival for the group rather than the actual survival, which is not yet known since the censored patients are still alive at the time of analysis. [ws-17] As it was mentioned before, censored patients may drop out the study due to various reasons but they are still alive. This is the reason that the survival curve after the first censored patients is only an estimate, since the censored patients are still alive and we do not know when exactly they will die. As a result, we can only estimate the survivability of the remaining group of patients in the study. Thus a censored subject contributes to the number at risk at the time he becomes censored but is not among those at risk immediately after that time.

If the last observed time is censored, then the estimate of the survivorship function does not go to zero and so its smallest value is at the last observed survival time. In this case the estimate is considered to be undefined beyond the last observed time. [ws-20]

Censoring must be accounted for in the analysis because otherwise, results might be misleading. The Kaplan-Meier and life-table methods assume that censoring is noninformative. That is, those censored at a given time are a representative sample of those at risk. If those censored tend to be at higher or lower risk of failure, the censoring is informative and the estimates may be biased. [ws-19] In our data the censored data are noninformative.

Many clinical trials are designed with a minimum follow-up time. This means that the results aren't reported until that amount of time after the last patient signed up for the trial. In this case, no patients will be censored until after that minimum time on the curve.

It is useful to see when patients are censored so as to get a feeling for the reliability of the curve over time. Survival curves often have a tick mark at each point where a patient was censored. They may even show the number remaining "at risk" at several points instead. The number at risk at any point in the survival curve is the number of patients who are still alive and whose follow-up extends at least that far into the curve. [ws-17]

3.1.2 Analytical Description of the method

To compute a survival curve, the time of occurrence of events must be noted: t_1, t_2, t_3, \dots represent the times when a death happens. The starting point of the study must also be defined. Time zero is not some specified calendar date, rather it is the time that each patient entered the study. This means that $t_0=0$.

The basic computations for the Kaplan-Meier survival curve rely on the computation of conditional survival probabilities. In particular, the probability:

$$P[T \geq t_i | T \geq t_{i-1}]$$

which can be interpreted as the probability of the surviving to a specific time, given that someone survived to the previous time.

This probability is easy to calculate if the number of deaths at a specific time is known and if the number of patients at risk at that same time is also known. A more difficult probability is the unconditional probability of survival:

$$P[T \geq t_i]$$

which can be calculated as it is shown in the following lines:

$$\begin{aligned} P[T \geq t_i] &= P[T \geq t_i | T \geq t_{i-1}] \\ &\times P[T \geq t_{i-1} | T \geq t_{i-2}] \\ &\times P[T \geq t_{i-2} | T \geq t_{i-3}] \\ &\times \dots \times P[T \geq t_0] \end{aligned}$$

Therefore, the unconditional probability is equal to the cumulative product of conditional probabilities.

At each time point, it should be counted:

- *di: the number of deaths*
- *ci: the number of censored data.*
- *ni: the number of patients at risk*

which is defined as:

$$n_i = n_{i-1} - d_{i-1} - c_{i-1}.$$

The number at risk at any specific time point is just the number at risk at the previous time point, minus the number of deaths/failures and the number of censored observations. Next the conditional probability of survival is computed:

$$P[T \geq t_i | T \geq t_{i-1}] = 1 - \frac{d_i}{n_i}$$

Finally, the unconditional probability of survival is simply the cumulative product of the conditional probabilities. The following equation is the one that gives the Kaplan-Meier estimator: [ws-22]

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

3.1.3 Confidence Intervals

In order to generalize the knowledge from a sample to the overall population, a survival curve is more informative when it includes a 95% CI. The survival has been measured exactly in a sample but it is needed a more descriptive clarification for the survival of the entire population. There is certainty 95% that the true population survival curve lies within the 95% CI. [ws-21]

If the exact number of patients who are still alive is known at any specific time, then the following equation can be used to calculate an approximate 95% CI for the fraction surviving up to at any time t .

The 95% CI using the Greenwood formula: [ws-19]

$$V[S(t)] = S(t)^2 \sum_{i \leq t} \frac{d_i}{n_i(n_i - d_i)} \quad (1)$$

$$\begin{aligned} lower &= S(t) - 1,96 * S(t) * \sqrt{\sum_{i \leq t} \frac{d_i}{n_i(n_i - d_i)}} \\ upper &= S(t) + 1,96 * S(t) * \sqrt{\sum_{i \leq t} \frac{d_i}{n_i(n_i - d_i)}} \end{aligned} \quad (2)$$

The standard error of the Kaplan-Meier estimator is given by the variance Greenwood's formula. In general, confidence intervals are calculated as: +/- 1.96(standard error), assuming that the errors are normally distributed. In order to conclude to equation 1 the following steps should be followed.

Grouped data are considered and so at first we should group the data using the uncensored times $t_{(1)} < t_{(2)} < \dots < t_{(k)}$.

For each risk set $R_j = \{t_i : t_i \geq t_{(j)}\}$ counting the number of failures is a binomial experiment. So, $d_{(j)} \sim \text{Binomial}(n_{(j)}, h_{(j)})$ where $h_{(j)}$ is the hazard at $t_{(j)}$. Let $q_{(j)} = 1 - h_{(j)}$, for $t_{(i-1)} \leq t < t_{(i)}$. Then,

$$\begin{aligned} \text{var}(\log \hat{S}(t)) &= \text{var}(\log \{\hat{q}_{(1)} \hat{q}_{(2)}, \dots, \hat{q}_{(i-1)}\}) \\ &= \text{var}(\log \hat{q}_{(1)} + \log \hat{q}_{(2)} + \dots + \log \hat{q}_{(i-1)}) \\ &= \sum_{j=1}^{i-1} \text{var}(\log \hat{q}_{(j)}) \end{aligned}$$

The variances are additive because the risk sets at $t_{(1)}, t_{(2)}, \dots, t_{(k)}$ are nested. So, by statistical theory, we can treat $\log \hat{q}_{(1)}, \log \hat{q}_{(2)} \dots$ as uncorrelated terms. By the delta method, which is a method for computing the variances, for a transformation ϕ of an estimate $\hat{\theta}$ we have:

$$\text{var}(\phi(\hat{\theta})) \approx [\phi'(\hat{\theta})]^2 \text{var}(\hat{\theta})$$

Using the delta method:

$$\begin{aligned} \text{var}(\log \hat{q}_{(j)}) &\approx \left[\frac{1}{\hat{q}_{(j)}} \right]^2 \text{var}(\hat{q}_{(j)}) = \frac{1}{\hat{q}_{(j)}^2} \cdot \frac{h_{(j)} q_{(j)}}{n_{(j)}} = \frac{h_{(j)}}{q_{(j)} n_{(j)}} \\ \text{var}(\log S(t)) &= \sum_{j=1}^{i-1} \text{var}(\log \hat{q}_{(j)}) \approx \sum_{y_{(j)} \leq t} \left(\frac{h_{(j)}}{q_{(j)} n_{(j)}} \right) \end{aligned}$$

Using the delta method again,

$$\begin{aligned} \sigma(t)^2 &= \text{var}(S(t)) = \text{var}(\exp(\log \hat{S}(t))) \\ &\approx [S(t)]^2 \cdot \text{var}(\log \hat{S}(t)) \end{aligned}$$

We put in $\hat{h}_{(j)} = \frac{d_{(j)}}{n_{(j)}}$ and $\hat{q}_{(j)} = \frac{n_{(j)} - d_{(j)}}{n_{(j)}}$.

The Greenwood's formula, for estimating the variance of the Kaplan-Meier estimate, is:

$$\text{var}(S(t)) \approx [S(t)]^2 \sum_{y_{(j)} \leq t} \frac{d_{(j)}}{n_{(j)}(n_{(j)} - d_{(j)})},$$

which is the same equation as equation 1 [oth-5]. By using the standard error, we can obtain approximate confidence intervals for the survival function $S(t)$. [ws-26] These confidence intervals are given by equations 2. Using these two equations, the upper and the lower 95% CI can be found of the survival curve and so a more complete image for the entire population can arise.

3.1.4 Implementation of the method in the ocular melanoma dataset

As it was mentioned, in this study the censored data belong to the third type of censoring, which is also called random censoring. The patients entered the study at different times (date of management) and all the data are selected up to a specific time (21/3/2005). Some died before that day and so their exact time of death is known and some are still alive until that day. Those who are alive are recognised because their date of death is missing (DOD). Those patients that their DOD is missing are considered as censored patients.

In order to plot the Kaplan-Meier survival curve, the exact time events have to be known. The variable *survy* (survival years) is very helpful because for those patients that are dead, the value of this variable is the time from date of management until the data of death. For those patients that are still alive the survival years are from the date of managements until the 10th February 2005. This is very convenient because the exact starting point for each person has not to be clarified as it common for all the patients. The counting begins from time zero.

Because the Kaplan-Meier Method is a method that takes advantage of the censored patients, these patients are identified as these who do not have a specific date of death and so this value is missing.

By counting all these patients, those who are still alive, as censored, there are left 490 patients who have died. This means that 1244 patients were still alive until the last date of this study (21/3/2005). From the 490 patients who died some of them died from cancer and some of them died from another cause.

Total N	N of Events	Censored	
		N	Percent
1734	490	1244	71,7%

Table 4: Contains the number of the censored and dead patients

If the survival curve is plotting deaths due to a particular form of cancer, it must be decided what to do with those patients who died from another cause and not from cancer. In our study the patients who died from cancer are recognised because there is a characteristic value **metastatic death** which is equal to 1. On the contrary, those who died from another cause have characteristic value **death** equal to 1 and **metastatic death** equal to 0.

Metastatic death: 0 and death: 1 → person died from another cause
 Metastatic death: 1 and death: 1 → person died from cancer

Some investigators count those who died from another cause as censored subjects and others count them as deaths. In this case, they were counted as deaths. [ws-21]

Below it is given a small sample of four patients of the dataset so as to explain when a patient is thought to be dead for cancer, dead from another cause or censored.

mand	DOD	md	death	survy
19-Jan-2005	missing	0.00	0.00	0.06
19-Mar-1993	missing	0.00	0.00	11.90
21-Oct-1994	8-Apr-1997	0.00	1.00	2.46
3-Jun-1985	29-Nov-1989	1.00	1.00	4.49

Table 5: Contains a sample of four patients of the dataset

1. The first two data represent patients who are still alive until the date 10th February 2005. These patients are thought to be censored.
2. The third patient is someone who has died from another cause and so he is also considered as dead.
3. The fourth patient has died from cancer.

Because in some small time intervals, more than an event took place, the data had to be grouped so as to make easier the computation of the survival probability with the Kaplan-Meier estimator.

survy	deaths di	censored ci	at risk ni	di/ni	1-di/ni
0,01	0	2	1734	0	1
0,02	1	1	1732	0,000577	0,999423

Table 6: Contains a sample of the Kaplan-Meier analysis

- The first column contains the survival years which are sorted in ascending way.
- The second column contains the patients who died.
- The third column contains the patients who were censored.
- The fourth column contains the patients at risk, meaning the patients who were alive in the previous time minus those who died and those who were censored.
- The fifth and the sixth columns contain computations that will eventually help to compute the cumulative probability. Actually the sixth column contains the conditional probability.

In order to implement the Kaplan-Meier estimator, the cumulative probability must be calculated.

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

The cumulative probability, the 95% CI and the survival curve, are all computed with the help of Matlab [Appendix A] and the results are given in chapter 5.

3.2 The Life-table method

3.2.1 Analytical description of the method

As it was mentioned above, The Life-table method is like the Kaplan-Meier method with the difference that the time is separated in intervals. Since censoring is an important issue for the Kaplan-Meier method, is logical that the Life-table method also considers censoring in its analysis. The detail analysis and description of the characteristics of the Life-table method follows. This study makes use of the clinical Life-Tables and not of the population Life-tables. The former have usually the following columns:

1. **Interval (t_i - t_{i+1}):** The first column gives the intervals into which the survival times and censored times are distributed. The last interval has an infinite length and all the intervals are fixed.
2. **Midpoint (t_{mi}):** The midpoint of each interval is included for convenience in plotting the hazard and the probability density functions.
3. **Width (b_i):** The width of each interval is needed in the calculation of the hazard and density functions, and the width of the last interval is infinite.
4. **Number lost to follow-up (l_i).** This is the number of people who are lost to observation and whose survival status is thus unknown in the i th interval. In this dataset there are no people lost to follow-up.

5. **Number withdrawn alive (w_i):** People withdrawn alive in the i th interval are those known to be alive at the closing date of the study. The survival time of these people is the time from the date of management until the 10th February 2005.
6. **Number dying (d_i):** This is the number of people who die in the i th interval. The survival time of these people is the time from date of management to death.
7. **Number entering the i th interval (n_i'):** The number of people entering the first interval is the total sample size. The other entries are calculated from $n_i' = n_{i-1}' - c_{i-1} - d_{i-1}$ which means that the number of persons entering the i th interval is equal to the number studied at the beginning of the previous interval minus those who are censored and died
8. **Number exposed to risk (n_i):** This is the number of people who are exposed to risk in the i th interval. It is assumed that the censored times are approximately uniformly distributed in the interval. Therefore, people lost in the interval are exposed to risk of death for one-half the interval. $n_j' = n_j - c_j / 2$. If there are no censored patients $n_i = n_i'$
9. **Conditional proportion dying (q_i):** It is defined as $q_i = d_i / n_i$. It is an estimate of the conditional probability of death in the i th interval.
10. **Conditional proportion surviving (p_i):** It is given by $p_i = 1 - q_i$ and is an estimate of the conditional probability of surviving in the i th interval
11. **Cumulative proportion surviving ($S(t)$):** It is an estimate of the survivorship function at time t_i and some times it is referred as the cumulative survival rate. For $S(t_i) = \begin{cases} 1, & i = 0 \\ S(t_i) = p_{i-1} * S(t_{i-1}), & i > 0 \end{cases}$
12. **Estimated probability density function ($f(t)$):** It is defined as the probability of dying in the i th interval per unit width. $f(t) = \frac{S(t) * q_i}{b_i}$
 $i = 1, \dots, s-1$

13. **Hazard function (h (t)):** The hazard function for the i th interval is the number of deaths per unit time divided by the average number of survivors at the midpoint of the interval

$$h(t) = \frac{f(t)}{S(t)} = \frac{S(t) * \frac{q_i}{b_i}}{\frac{1}{2} * [S(t_{i+1}) + S(t_i)]} = \frac{S(t) * \frac{q_i}{b_i}}{\frac{1}{2} * S(t) * (p_i + 1)} = \frac{2 * q_i}{b_i * (1 + p_i)}$$

$S(t)$ is defined as the probability of surviving at the beginning and not at the midpoint of the interval. In the following table it is given the format of the clinical Life-Table. [b-1]

In order to compute the Life-table estimator, the data of the ocular melanoma set had to be formed as the following table. This was done with Microsoft Excel and then the data with the new format were used in Matlab. The results of the analysis of this method will be given in chapter 5.

3.2.2 Confidence intervals

Confidence intervals will also be used in the Life-Table method. With the use of the confidence intervals we can conclude to clearer results. At this point, the standard errors will be given since the confidence intervals, as we mentioned before, are calculated as : +/- 1.96(standard error).

The standard error of the survival function is derived from the Greenwood's

formula and is given by: $\sigma(\hat{S}(t_i)) = \hat{S}(t_i) \sqrt{\sum_{j=1}^{i-1} \frac{\hat{q}_j}{n_j \hat{p}_j}}$.

The standard errors that follow and refer to the density function and to the hazard function accordingly, are not so commonly used and they are not derived from the Greenwood (1929) formula but they were suggested by Gehan (1969). The standard

error of the density function is given by: $\sigma(f(\hat{t}_{mi})) = f(\hat{t}_{mi}) \sqrt{\sum_{j=1}^{i-1} \frac{\hat{q}_j}{n_j \hat{p}_j} + \frac{\hat{p}_i}{n_j \hat{q}_j}}$.

Finally, the standard error of the hazard function is given by:

$$\sigma(h(\hat{t}_{mi})) = h(\hat{t}_{mi}) \sqrt{\frac{1 - (b_i h(\hat{t}_{mi}) / 2)^2}{n q_i}}$$

Interval	Midpoint	Width	Number Lost to Follow-up	Number Withdrawn Alive	Number Dying	Number Entering Interval	Number Exposed to Risk	Conditional Proportion Dying	Conditional Proportion Surviving	Cumulative Proportion Surviving	Probability Density	Hazard
$t - t_2$	t_{m1}	b_1	l_1	w_1	d_1	n'_1	n_1	\hat{q}_1	\hat{p}_1	$\hat{S}(t_1) = 1.00$	$\hat{f}(t_{m1})$	$\hat{h}(t_{m1})$
$t_2 - t_3$	t_{m2}	b_2	l_2	w_2	d_2	n'_2	n_2	\hat{q}_2	\hat{p}_2	$\hat{S}(t_2)$	$\hat{f}(t_{m2})$	$\hat{h}(t_{m2})$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$t_i = t_{i+1}$	t_{mi}	b_i	l_i	w_i	d_i	n'_i	n_i	\hat{q}_i	\hat{p}_i	$\hat{S}(t_i)$	$\hat{f}(t_{mi})$	$\hat{h}(t_{mi})$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$t_{s-1} - t_s$	$t_{m,s-1}$	b_{s-1}	l_{s-1}	w_{s-1}	d_{s-1}	n'_{s-1}	n_{s-1}	\hat{q}_{s-1}	\hat{p}_{s-1}	$\hat{S}(t_{s-1})$	$\hat{f}(t_{m,s-1})$	$\hat{h}(t_{m,s-1})$
$t_s - \infty$	—	—	l_s	w_s	d_s	n'_s	n_s	1	0	$\hat{S}(t_s)$	—	—

Table 7: Contains the format of the Life-table analysis [b-1]

3.3 The Nelson-Aalen method

3.3.1 Analytical description of the method

The Nelson-Aalen method, as it was mentioned in the previous chapter, computes directly the cumulative hazard function and with the transformation $S(t) = \exp[-H(t)]$, the survival function can be derived. The Kaplan-Meier method does exactly the opposite analysis. Once again, censoring is considered in the analysis. The Nelson-Aalen is given by:

$$H(t) = \sum_{t_j < t} \frac{d_j}{n_j}$$

where d_j is the number of dead and n_j is the number of the at risk patients. [ws-18] As it was mentioned again, the number at risk at any specific time point is just the number at risk at the previous time point, minus the number of deaths/failures and the number of censored observations.

In order to compute the Nelson-Aalen estimator, Matlab was used [Appendix A]. The results are given in chapter 5.

3.3.2 Confidence Intervals

The Confidence intervals will also be used in the Nelson-Aalen method as an important piece of the analysis. The standard error is given by:

$$\sigma(\hat{H}(t)) = \hat{H}(t) \sqrt{\sum_{t_j < t} \frac{d_j}{n_j^2}}$$

As mentioned again before the confidence intervals, are calculated as: +/- 1.96(standard error).

3.4 The Aalen's additive model

As it was mentioned in previous chapters, the Aalen's additive model is a semiparametric model. It makes use of covariates in order to analyse the survivability of the patients. The main purpose of this method is to result in plots that give information on the change over time in the influence of covariates. This information is taken from the change of the slope. In this section of this chapter, it will be given the analytical description of the method and also an analysis of the covariates of this dataset will be presented.

3.4.1 Analytical description of the method

A number of individuals are observed over time to see if a specified event occurs. The individuals are assumed to be independent and any events happening to the individuals are also assumed to be independent between individuals. The life times that are observed may be complete or right censored. [p-1] The basic equation may be formulated as follows: [p-3]

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

The hazard at any time is thus a sum of a baseline hazard $\beta_0(t)$ and a linear combination of the covariate values, Z_j .

In Aalen's terminology, the hazard $h_i(t)$ denotes the probability (intensity) of the event occurring at time t for individual i , given that it has not occurred before. Let's clarify that:

- n is the number of subjects
- r is the number of covariates.

So, the vector of intensities $h_i(t), i = 1, 2, \dots, n$ 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 i th individual is a member of the risk set at time t then the i th row of $Y(t)$ is the vector $Y^i(t) = (1, Z_1^i(t), Z_2^i(t), \dots, Z_r^i(t))$, where $Z_j^i(t), j = 1, \dots, r$ 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.

It should be noted that the value of the Y vector must be used just before a relevant event time and that the covariates are not time-dependent, which means that the covariates are not a function of time.

The first element of the vector $\beta(t)$ is a baseline parameter and the remaining elements $\beta_i(t), i = 1, 2, \dots, r$ are called regression functions and estimate the influence of the covariates. [p-1] 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. [p-4]

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 impractical and difficult to estimate the individual regression functions and instead the cumulative regression functions are estimated. This happens for the same reason that it is easier to estimate a cumulative distribution than a probability density.

The elements $B_j(t), j = 0, 1, \dots, k$ of the column vector $\mathbf{B}(t)$ are the cumulative regression functions and are defined as:

$$B_j(t) = \int_0^t \beta_j(s) ds$$

If $T_1 < T_2 < \dots$ are the ordered event times, meaning the times when an actual event, not censoring, occurs. Then it may be proved [p-20] that a reasonable estimator of $\mathbf{B}(t)$ is given by:

$$\mathbf{B}^*(t) = \sum_{T_k \leq t} X(T_k) I_k$$

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

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

It should be noted that the estimator $\mathbf{B}^*(t)$ is only defined as long as $\mathbf{Y}(t)$ has full rank and therefore $\mathbf{Y}'\mathbf{Y}$ is invertible. Therefore estimates are restricted to the time interval where $\mathbf{Y}'\mathbf{Y}$ is not singular or alternative to the time when the rank of the matrix $\mathbf{Y}(t)$ falls below $r+1$. In fact this time, R will be considered as a final censoring time. This means that all subjects remaining in the risk set at this time are considered as censored. The introduction of such a censoring does not affect the statistical analysis, since it is a stopping time.

Also the estimates of the baseline hazard are not constrained to non-negative numbers. This should not be a problem because the scope of this model is not to conclude to a baseline hazard function but to plots that will give information on the change over time in the influence of covariates.

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.

The survival function as well as the cumulative hazard can also be estimated. Let $Z_i = (1, Z_1^i, \dots, Z_r^i)$ be a set of covariate values for the i th individual that are fixed at given times. In analogy to the Kaplan-Meier estimator, the survival function may be estimated as follows for each patient:

$$S_i^*(t) = \prod_{T_k \leq t} [1 - (X(T_k)I_k)' Z_i]$$

where Z_i is the vector of the covariate values. What makes the above equation suitable for computing the survival function for each patient separately is that vector Z_i of the covariates, which is unique for each of the patients. The above equation computes the unconditional probability of survival and that is why, as in the Kaplan-Meier method where the product of the unconditional probabilities was used in order to compute the unconditional probability, the product of the integrals that compute the cumulative regression functions is used in the above equation. Accordingly the cumulative hazard function is given by the relationship:

$$H^*(t) = -\log S(t)$$

Another way of computing the survival function is through the cumulative hazard function. This means that the inverse procedure must be followed from the one that we described previously. After computing the cumulative hazard function, the survival function will be computed by the following equation: $S(t) = \exp[-H(t)]$

The procedure that must be followed is the following: The hazard function of the additive model is the following: $h(t, z) = \beta_0(t) + \sum_{j=1}^k \beta_j(t)z_j(t)$. By integrating

this equation, we get the cumulative hazard function: $H(t, z) = B_0(t) + \sum_{j=1}^k B_j(t)z_j(t)$

In the case of our study that we have three covariates, the above cumulative hazard function becomes: $H(t, z) = B_0(t) + B_1(t)z_1(t) + B_2(t)z_2(t) + B_3(t)z_3(t)$. After the computation of the cumulative hazard function, we use the equation $S(t) = \exp[-H(t)]$ and so we get the survival function. The results that come up from the one way and the other are exactly the same.

It should be mentioned that the estimated survival functions are not necessarily monotonic over the whole observation period. In fact they may increase at some values of t and accordingly the cumulative hazard functions may decrease for some time t . [p-1]. In chapter 5 there will be given results about the survivability of the patients. Also results about how their survivability is influenced by the presence of the covariates will be given. These results will be given for the entire population, but also for separate patients.

3.4.2 Description of the covariates of the ocular melanoma dataset

For the Aalen's additive model not all of the patients were used and this is due to the fact that this model makes use of covariates. Many covariates are not complete and as a result there were many missing values that had to be detracted, in order to implement the additive model. After the removal of the missing values, they were left 743 patients.

Below it is given a detailed description of the data set and a table that summarizes the data and their values.

At first the data that have to do with time are described and after that, the covariates and their meaning are presented. The information description of the covariates was given by the online support of the Bioppatern group.

- **DOB**: date of birth, which is obtained from hospital records
- **DOD**: date of death, which is obtained from death certificate, family, medical practitioner, or any other valid source.
- **Mand**: date of management, which is the date of primary treatment of intraocular tumour
- **Surv**: survival years, which are measured from the date of primary ocular treatment (mand) to the date of death or if there is no notification of death then to the 10th February 2005.
- **Sex**: 0 → female, 1 → male
- **Md**: metastatic death : 0 → no, 1 → yes
- **Death**: patient diseased : 0 → alive, 1 → dead
- **Uh**: ultrasound height, which is the tumour height measured by ultrasonography

- **Lubd**: Longest ultrasound basal dimension, which defines the longest basal tumour dimension as measured by ultrasonography. [oth-1] There is a staging system that is widely used and is devised by the Collaborative Ocular Melanoma Study group. They use a system that refers to small, medium, and large melanomas, which is revised below. [ws-1]

Type	Thickness	Diameter
Small	1.0 mm to 2.5 mm	5 mm
Medium	2.5 mm to 10 mm	5 mm to 16 mm
Large	10 mm or more	16 mm or more

Table 8: Contains a system that categorize the dimension of the basal tumor

- **Epi**: Epithelioid cellularity, which describes the presence or not of epithelioid melanoma cells. 0→no, 1→yes
- **Antmar**: anterior tumor margin, which is estimated by ophthalmoscopy, slit-lamp examination and ultrasonography. This variable takes 14 different values, each of them corresponds to a different part of the eye that the tumor is located. [oth-1]

Number	Location
1	Disc & Fovea
2	Fovea
3	<=1DD Fovea
4	1-2DD Fovea
5	Disc
6	<=1DD Disk
7	1-2DD Disk
8	Posterior choroid
9	Anterior Choroid
10	Pars Plana
11	Pars Plicata
12	Iris
13	Angle
14	Cornea

Table 9: Contains the description of the anterior tumor margin covariate

This covariate can also be categorized in two subgroups:

Subgroups	Value of antmar
Choroidal (value 0)	Antmar <=9
Ciliary (value 1)	Antmar >9

Table 10: Contains the two subgroups that can be derived from categorization of the anterior tumor margin covariate

If the data get categorized according to the above values of antmar then we get two subgroups of patients.

Subgroups	Type of antmar	Number of patients
Group A	Choroidal	446 patients
Group B	Ciliary	297 patients

Table 11: Contains the number of the patients of the two subgroups

An example of the data described above is given, for two patients.

DOB	antmar	mand	DOD	lubd	uh	epi	md	death	survy	sex
25-Feb-1933	9	19-Jan-2005	missing	13.60	6	1.00	0.00	0.00	0.06	0.00
24-Oct-1925	9	21-Oct-1994	8-Apr-1997	14.80	7	1.00	0.00	1.00	2.46	1.00

Table 12: Contains an example of two patients and their covariates

Below it is given a brief description of the above parts of the eye and a picture of the anatomy of the eye. [ws-15]

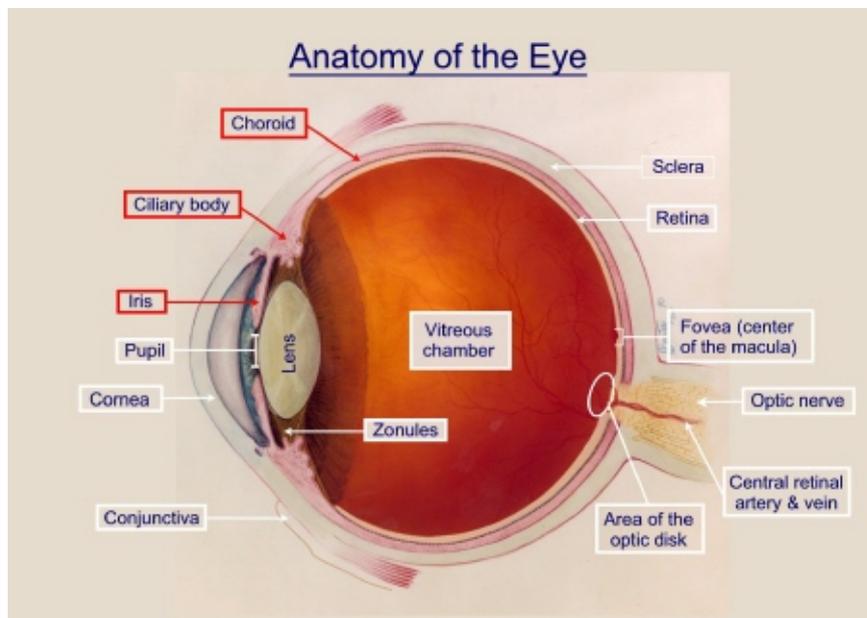


Figure 4: The anatomy of the eye

The **choroid** is part of the lining of the eyeball and is dark-colored to prevent light being reflected around the inside of the eye. The **ciliary body** extends from the choroid and focuses the eye by changing the shape of the lens. The ciliary body is divisible into two parts: the pars plana and the pars plicata [ws-5] The **pars plana** is the part of the eye lying between the retina and the pars plicata. The **retina** converts light energy into a signal that is sent to the brain. The **pars plicata** makes intraocular fluid, the aqueous humor. [ws-14] The **iris** is the clearly visible colored disc at the front of the eye, which controls the amount of light entering the eye. [ws-5] The **fovea** is the centre most part of the macula. This tiny area is responsible for our central, sharpest vision. [ws-16] The **macula** is located roughly in the centre of the retina, temporal to the optic nerve. It is a small and highly sensitive part of the retina responsible for detailed central vision. The macula allows us to appreciate detail and perform tasks that require central vision such reading. [ws-13] The **cornea** is the transparent, dome-shaped window covering the front of the eye. It is a powerful refracting surface, providing 2/3 of the eye's focusing power. Like the crystal on a watch, it gives us a clear window to look through. [ws-12] The area in the anterior chamber where the cornea and iris join is known as the **angle**. [ws-10] The optic nerve transmits electrical impulses from the retina to the brain. It connects to the back of the eye near the macula. When examining the back of the eye, a portion of the optic nerve called the optic **disc** can be seen. [ws-11]

3.4.3 Implementation of the method in the ocular melanoma dataset

As it was mentioned previously, the main focus of the analysis in the additive Aalen's model is on cumulative regression plots, where the slope at any given time gives information on the influence of the covariates at that specific moment.

The meanings of the covariates, as well as the different covariates of the ocular melanoma dataset were explained in previous sections. In this section it will be presented the implementation of the Aalen's additive model in the ocular melanoma dataset as well as the solution of some problems that came up during the analysis.

Originally there were 1743 patients in the data set but for the purposes of the additive model only 743 patients were used. Since information on the selected covariates is not complete, the number of the patients had to get deduced. For the implementation of the additive model, three covariates were selected in order to get

examined for their effect over the lifetime distribution and for their influence in the disease of ocular melanoma.

These covariates had to be time independent or else constant over time. Moreover these covariates had to have medical interest. This means that after the analysis, we had to conclude to results that would have some meaning. In this way doctors could use them so as to come up with useful conclusions for the disease of ocular melanoma. These covariates describe certain characteristics of the tumour and they were obtained after medical observation. From all the covariates that are contained in the data set the only three covariates that fulfil the above requirements are the:

- *Longest ultrasound basal dimension*, which defines the longest basal tumour dimension as measured by ultrasonography
- *ultrasound height*, which is the tumour height measured by ultrasonography
- *Epithelioid cellularity*, which describes the presence or not of epithelioid melanoma cells.

All the other covariates contain information about time (DOB, DOD, mand, survy) or are indicators about death (md, death). In addition, since the ocular melanoma has almost no difference between males and females, as it was mentioned in the description of the disease in chapter 1, the covariate sex was not examined. Finally, the covariate antmar which describes in which part of the eye the tumour was located will be used in order to categorize the data in two subgroups for further analysis. The two subgroups are the “ciliary body” subgroup and the “choroidal” which are the main points of the eye that the disease appears.

The analysis of the additive model is going to be implemented with the three covariates that are summarised in the following table. According to the results that will arise, the influence of these covariates is going to get examined.

Covariate	Label	Value
Ultrasound height	uh	Mean=15mm, range(0-30mm)
Longest ultrasound basal dimension	lubd	Mean=15mm, range(0-30mm)
Epithelioid cellularity	epi	Yes: n= 464 ,No: n=279

Table 13: Contains the three covariates that will be examined in the Aalen’s additive model

In the beginning of the analysis the data set had to be determined. As mentioned from the 1734 patients the analysis of the additive model was implemented with 734 patients. This happened because the three covariates that were chosen had missing values and so the individuals with these covariates had to be removed.

In addition the matrix **Y** had to be computed for discrete time moments. This means that for each separate time point, only a single matrix **Y** should be computed. This was a problem because in the data set of ocular melanoma in a specific time, more than one event took place. For example in time 0,73 years one person dies and three others are still alive, meaning that they are censored. So, in the separate time point of the 0,73 years four events took place. What it had to be done was to count all the death events and the censoring events that took place in a specific time separately, so as to have summarized in all the different times the dead and censored patients. In this way, the calculation of the **Y** matrix will be held without problems. The code that computes the summarized events as well as the **Y** matrix is in Appendix B. After that computation, 560 different **Y** matrixes were created, for each different time. This is logical since before, we had in the same time four different events, while now we have in the same time all the events summarized and in this way there is no need for the time point to be repeated. This difference will be shown in the following two tables.

Survival	Dead patients	Censored patients
0,73	1	0
0,73	0	1
0,73	0	1
0,73	0	1
0,75	0	1
0,75	0	1

Table 14: Contains the events that take place in two time point before to be summarized

The table that follows contains the same time points, but the dead and censored patients are now summarized from each time point. By counting the events, we see that in time 0,73 we have 4 events, 1 dead patient and 3 censored, and in time 0,75 we have two events, no dead patients and two censored.

Survvy	dead	censored
0,73	1	3
0,75	0	2

Table 15: Contains the events that take place in two time point after to be summarized

After the computation of the \mathbf{Y} matrix, the \mathbf{X} matrix had to be computed. As it was mentioned, this matrix can only be computed until the time when the rank of the matrix \mathbf{Y} falls below $r+1$. Because the covariates in this analysis are three, when the rank of \mathbf{Y} falls below 4, then the computation of the \mathbf{X} matrix stops. In this case the computations stop at 20, 68 years and as a result 546 different \mathbf{X} matrixes are created. The individuals that take place in the analysis remain 743. The computation of the \mathbf{X} matrix is presented in Appendix B.

In order for the analysis to be continued, the cumulative regression functions will be calculated. As it was mentioned in the description of this method previously in this chapter, in the calculation of the estimator $B^*(t) = \sum_{T_k \leq t} X(T_k)I_k$, which has as elements the cumulative regression functions, only the ordered event times must be computed in the analysis. This means that will be used only the times that an event, or else death, takes place and not the times that censoring occurs.

This leads to a deduction in the number of patients because between the 743 individuals, there are patients that are censored. These patients have to be left out from the computation of the cumulative regression functions, as the model indicates. After this deduction that was required for the analysis of the additive model, 271 patients were left and 238 different matrixes B were computed. This difference in the number of patients is due to the fact that at each different time more than one event can take place, as it was explained before. The analysis after this deduction stops at 20, 58 years. The computation of the B matrix is presented in the Appendix B.

The cumulative regression functions that were computed have to be plotted against time so as to give a description of how the covariates influence the survival over time. The change in the slope gives that information. These plots will be given in chapter 5.

Chapter 4

Artificial Neural Networks

In the beginning of this chapter is given an explanatory theoretical background about the artificial neural networks, and later in this chapter, it is given the architecture of the neural network that use in this study. In chapter 2, there were presented the neural networks that are used in survival analysis. In this study, we did not use any of these neural networks because our goal was not to implement a separate survival analysis with the neural networks but to train a neural network with an already used survival method. This means that the inputs of the neural network that we used are known. If we wanted to implement survival analysis not with statistical methods, but with a neural network, then one of the model that were described in chapter 2 would be used. The neural network that we used in this study was trained with the Kaplan-Meier method. In this way the results that came up from the statistical analysis and with the neural network analysis, can be compared.

4.1 Theoretical background of the neural networks

Artificial neural networks are inspired from the biological neural system with the perspective to reproduce some of the flexibility and power of the human brain in order to solve problems which are difficult for traditional sequential computers. They can be defined as synthetic networks that emulate the biological neural networks found in biological living organisms. [oth-4]

Another definition which may be given is that “a neural network is an interconnected assembly of simple processing elements, *nodes*, which functionality is loosely based on the animal neuron. The processing ability of the network is stored in the interunit connection strengths, *weights*, obtained by a process of adaptation to, or *learning* from, a set of training patterns.” [b-3]

As it can be derived from the above definition, the building blocks of artificial neural nets are the artificial neurons, which work in parallel in order to solve a specific problem. These neurons are tied together with weighted connections and have many inputs and one output. In order for the weights to be changed and adapt to the training set, a learning rule must be followed. The learning rule is also the way in which the neurons are interconnected.

The following definition describes in a very effective way what exactly the learning procedure is: “Learning is process by which the free parameters of a neural network are adapted through a process of simulation by the environment in which the network is embedded. The type of learning is determined by the manner in which the parameter changes take place” [oth-4]

The central idea of neural networks is that parameters, like the weight w and the bias b , can be adjusted so that the network exhibits some desired or interesting behavior. Thus, the network can be trained to do a particular job by adjusting the weight or bias parameters, or perhaps the network itself will adjust these parameters to achieve some desired end. Every network needs a transfer function which takes the argument n , the sum of the weighted input wp and the bias b , and produces the output a . The scalar input is p and is usually transmitted through a connection that multiplies its strength by a scalar input w and the result is the product wp . This procedure is summarized in the following diagram. [oth-3]

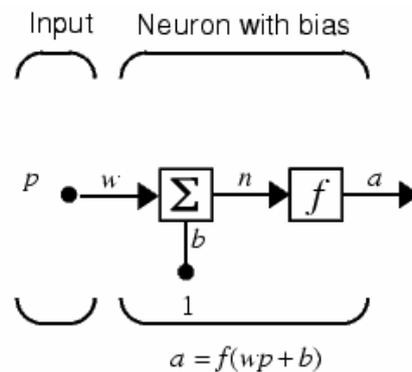


Figure 5: Diagram that shows the central idea of the neural networks

4.1.1 Transfer functions

The transfer functions are very essential in the formation of the neural networks. There are many transfer functions but the most commonly used are three.

- **The hard-limit Transfer function:** it limits the output of the neuron to either 0, if the net input argument n is less than 0, or to either 1, if n is greater than or equal to 0. It is usually used to make classification decisions.

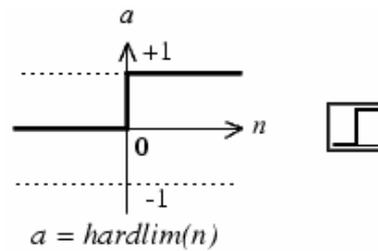


Figure 6: diagram that shows the hard-limit Transfer function

- **The linear Transfer function:** it is usually used as linear approximators.

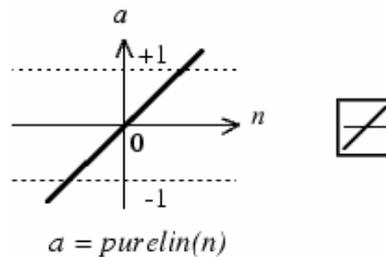


Figure 7 : diagram that shows the linear transfer function

- **The sigmoid Transfer Function:** it takes the input, which may have any value between plus and minus infinity, and produces the output into the range 0 to 1. It is commonly used in backpropagation networks.

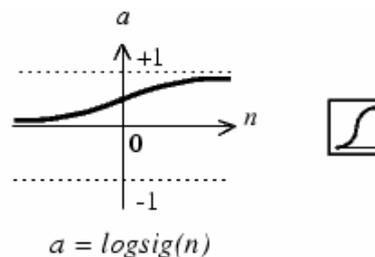


Figure 8: diagram that shows the sigmoid transfer function

4.1.2 Network Architectures

There are different network architectures, which are usually identified in three different classes:

- **Single Layer feedforward networks:** it consists of an input layer of source nodes that feed an output layer of neurons. Feedforward means that the neurons are connected to output neurons, but not vice versa.
- **Multilayer feedforward networks:** it consists of one or more hidden layers. The function of hidden layers is to intervene between the input and network output. When the size of the input is large, it is best to add more hidden layers because the results tend to be better and more precise. In multilayer

feedforward networks, the input neurons supply input features to the first hidden layer whose outputs constitute to the second hidden layer and so on. The weights connected to inputs, are called *input weights* and the weights coming from layer outputs, are called *layer weights*. The final output of the network represent the overall response of the network to the inputs fed by the inputs neurons of the input layer. The multiple layer networks are very powerful. Usually the first layers are sigmoid and the last is linear. With this kind of network, which is used in backpropagation, can be trained to approximate almost any function. The following network is a three-layer network with two hidden layers (layer 1 and 2) and one output layer (layer 3).

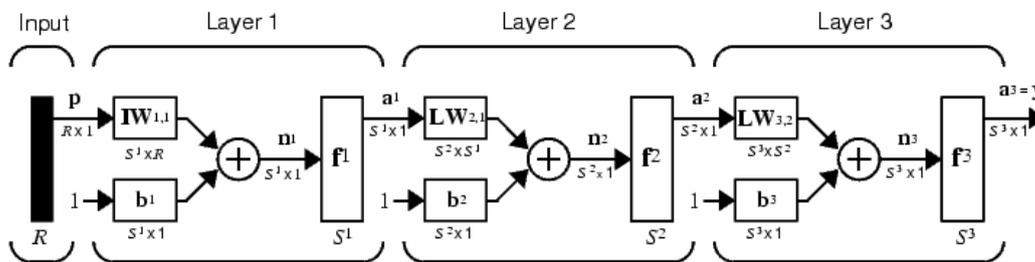


Figure 9: Diagram that shows a multilayer feedforward network

- **Recurrent networks:** it contains at least one feedback loop.

4.1.3 General information about networks

There are two different styles of training a network: the *incremental* training and the *batch* training. In the incremental training the weights and biases are updated each time an input is presented to the network. In batch training the weights and biases are only updated after all of the inputs are presented.

There are also two basic types of input vectors: those that occur at the same time, *concurrently*, and those that occur *sequentially* in time. For the former the order is not important but for the latter the order in which the vectors appear is important.

Lastly, there are *static* networks, they have no feedbacks or delays, and *dynamic* networks that contain delays.

As mentioned before, the learning rule is applied to train the network to perform some particular task. There are two main categories of learning rules: *supervised* training and *unsupervised* training. In supervised learning, both the inputs and the outputs are provided. The network then learns a set of input-output pairs which is called the training set. So the weights are adjusted such as the network responds to any input by the desired output or with an accepted accuracy or error. In unsupervised learning, the weights and biases are modified in response to network inputs only, since the network is only provided with inputs and no with the desired response. [oth-3]

There are many different learning rules used in neural networks such as the Hebbian rule, the Perceptron rule, the Delta rule, the Widrow-Hoff rule and others. The most commonly used rule in multilayer feedforward networks is the Backpropagation rule. Later in this chapter, a fully description of the neural network that we use, based on the above characteristics will be given.

4.1.4 Backpropagation learning rule

Backpropagation was created by generalizing the Widrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer functions. Input vectors and the corresponding target vectors are used to train a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way as defined by you. Networks with biases, a sigmoid layer, and a linear output layer are capable of approximating any function with a finite number of discontinuities.

Standard backpropagation is a *gradient descent algorithm*, in which the network weights are initialized with random values and then are changed in the negative direction of the error gradient in order to minimize the error function.

The term backpropagation refers to the manner in which the gradient is computed for nonlinear multilayer networks. Properly trained backpropagation networks tend to give reasonable answers when presented with inputs that they have never seen. Typically, a new input leads to an output similar to the correct output for input vectors used in training that are similar to the new input being presented. The target vector is known (supervised learning). This generalization property makes it possible to train a network on a representative set of input/target pairs and get good results without training the network on all possible input/output pairs.

The multilayer feedforward networks that use the backpropagation learning algorithm often have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons. This happens because if the last layer of a multilayer network has sigmoid neurons, then the outputs of the network are limited to a small range but if linear output neurons are used, then the network outputs can take on any value.

During training, the weights and biases of the network are iteratively adjusted to minimize the network performance function. The default performance function for the feedforward network is the mean *square error*. As each input is applied to the network, the network output is compared to the target. The error is calculated as the difference between the target output and the network output.

There are many variations of the backpropagation algorithm. The simplest implementation of backpropagation learning updates the network weights and biases in the direction in which the performance function decreases most rapidly, the negative of the gradient.

The backpropagation algorithm in rules is given by:

- weight change = some small constant \times error \times input activation
- For an output node, the error is:

$$\text{error} = (\text{target activation} - \text{output activation}) \times \text{output activation} \\ \times (1 - \text{output activation})$$

- For a hidden node, the error is:

$$\text{error} = \text{weighted sum of to-node errors} \times \text{hidden activation} \\ \times (1 - \text{hidden activation})$$

Because the backpropagation algorithm often takes a long time to learn, the learning rule is often augmented with a so called momentum term. This consists in adding a fraction of the old weight change. The learning rule then looks like:

$$\text{weight change} = \text{some small constant} \times \text{error} \times \text{input activation} + \text{momentum} \\ \text{constant} \times \text{old weight change}$$

In this point, the backpropagation algorithm in equation will be given:

- If j is a node in an output layer, the error δ_j is:

$$\delta_j = (t_j - a_j) a_j(a_j - 1)$$

where a_j is the activation of node j , t_j is its target activation value, and δ_j its error value

- If j is a node in a hidden layer, and if there are k nodes $1, 2, \dots, k$, that receive a connection from j , the error δ_j is:

$$\delta_j = (w_{1j} \delta_1 + w_{2j} \delta_2 + \dots + w_{kj} \delta_k) a_j (a_j - 1)$$

where the weights w_{1j} , w_{2j} , ..., w_{kj} belong to the connections from hidden node j to nodes $1, 2, \dots, k$. The backpropagation learning rule (applied at time t) is:

$$\Delta w_{ji}(t) = \mu \delta_j a_i + \Delta \beta w_{ji}(t-1)$$

where $\Delta w_{ji}(t)$ is the change in the weight from node i to node j at time t , the learning constant μ is typically chosen rather small, and the momentum term β is typically chosen around 0.5. [oth-6]

It is very important to measure the performance of a trained network. This can be measured to some extent by the errors on the training, validation and test sets, but it is often useful to investigate the network response in more detail. One option is to perform a regression analysis between the network response and the corresponding targets. The routine *postreg* is designed to perform this analysis.

It returns three parameters. The first two, m and b , correspond to the slope and the y-intercept of the best linear regression relating targets to network outputs. If the fit is perfect (outputs exactly equal to targets), the slope would be 1, and the y-intercept would be 0. The third variable returned by *postreg* is the correlation coefficient (R-value) between the outputs and targets. It is a measure of how well the variation in the output is explained by the targets. If this number is equal to 1, then there is perfect correlation between targets and outputs. [oth-3]

4.2 Implementation of the neural networks in the ocular melanoma dataset

The neural network was used in this study in order to extract results that could be used later in future studies, since the neural network will already be trained and so only the inputs and the outputs will be required in order for the results to be derived.

The data of the Kaplan-Meier method were used for the training of the neural network. This means that it has one input, a matrix that contains the dead and the at risk patients, and one output the Kaplan-Meier estimator. Because the number of patients was big enough, 1743, more than one hidden layer had to be used. In particular five hidden layers were used. The first hidden layer had 2 nodes and the rest four had five nodes each.

The learning rule that was used was the *backpropagation* because in this study the inputs and the target vectors are known and so they can be used so train the network until a function is approximated. As a training function the *traingd* could be used, since it uses batch descent training and so the weights and biases are updated in the direction of the negative gradient of the performance function. In order to have a faster training, the *trainlm (Levenberg-Marquardt)* function was finally used because the performance function will always be reduced at each iteration of the algorithm.

In addition, as learning function the *learnqdm* was used. It calculates the weight change for a given neuron from the neuron's input p and error E , the weight w , learning rate and momentum constant according to the gradient descent momentum. A momentum constant was used so as to make it less likely for a backpropagation networks to get caught in a local minimum. With momentum a network can slide through such a minimum.

Lastly, as a transfer function between the hidden layers the *tansig* was used rather than the *logsig* because the former is good for neural networks where speed is important and the exact shape of the transfer function is not.

By summarizing, in order to calculate the weight changes in the hidden layer, the error in the output layer is propagated 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. In the following table the characteristics of the neural network are summarized.

Architecture	Multilayer feedforward network
Number of inputs	1
Number of outputs	1
Number of hidden layers	5
Number of nodes of each hidden layer	2 & 5
Training style	Batch training
Type of input vector	Sequential
Type of network	Static
Category of learning rule	Supervised
Learning rule	Backpropagation
Training rule	Trainlm
Performance function	Mean square error(mse)
Leaning function	learngdm
Transfer functions	Logsigmoid(hidden layers) Purelin(output layer)

Table 16: Contains the characteristics of the neural network that we used in this study

Chapter 5

Results of the statistical methods and of the neural networks

In this chapter of this study, the results from the statistical methods that we used will be presented. As it was mentioned again previously, the methods that this study uses are the Kaplan-Meier method, the Life-Table, the Nelson-Aalen and the Aalen's additive model. Also in this study we use neural networks, the results of which will be presented later in this chapter.

5.1 Results from the Kaplan-Meier method

As it was mentioned in chapter 2, the equation that gives the Kaplan-Meier method is:

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

The survival curve that is derived for the dataset of the ocular melanoma disease is the following:

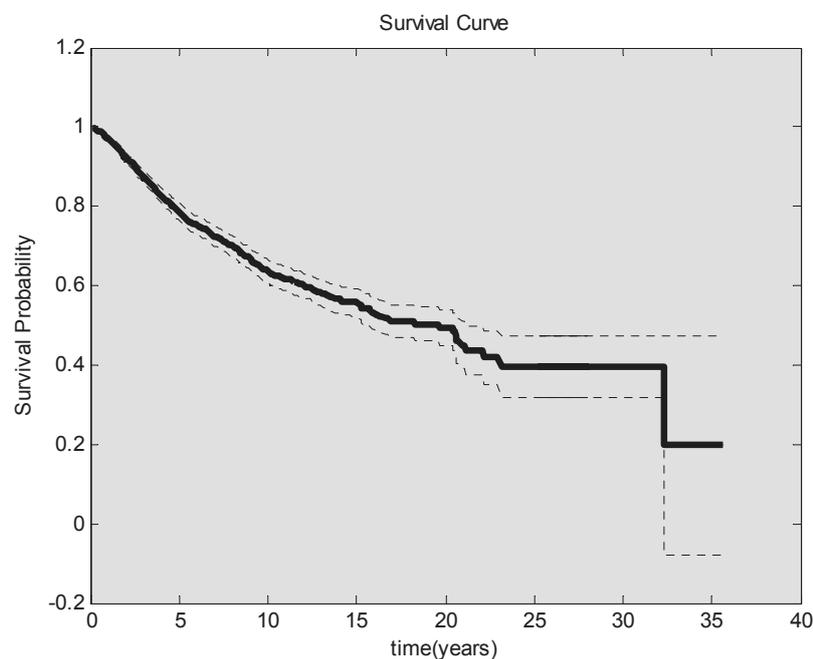


Figure 10a: The survival curve of the 1743 patients from the Kaplan-Meier method

The above survival curve is smoother in the beginning because there are more patients in the first 20 years and after that it takes a staircase appearance, which is logical because after the twenty years the patients are not so many. As it can be seen from this survival curve after the 23 years there are no deaths, and that is why the survival curve is stable, until the 32 years that a death occurs and so the survival curve takes a step down. Because the above survival is gradual, it can be derived that this curve represents high survival rate or longer survival. Moreover, this survival can be interpreted as follows: The chances, of those who managed to live 32 years, are about 20% to be cured.

Because the last observed time is censored, the estimate of the survivorship function does not go to zero and so the smallest value is at the last observed survival time. In this case the estimate is considered to be undefined beyond the last observed time, which is the case in this data set.

In this survival curve the 95 % CI is also plotted. This means that there is 95% certainty that the overall survival curve for the entire population lies within the dotted lines. The CIs are not wide in the beginning because there are many patients in the first 20 years, but they get wider after the 20 years where there are fewer patients.

As it can be seen from the survival curve, after the 23 year there are not so many deaths. Actually at 23,21 years a death happens and after that time point, the next death occurs at 32,33 years. By taking into account this last death our results are affected since from the survival curve is shown that the chances of those who managed to live 32 years, are about 20% to be cured. But if we do not take into account this last death, then the probabilities of the patients to be cured are augmented at about 40%. This means that the patients after the 23,31 years might be thought as a statistical error since, as we already said, affect our results. In order to have a clearer image, the survival function was plotted again but this time the last death was not taken into account. As it can be seen from the following plot, the probabilities of the patients to be cured are now 40%, as we already said.

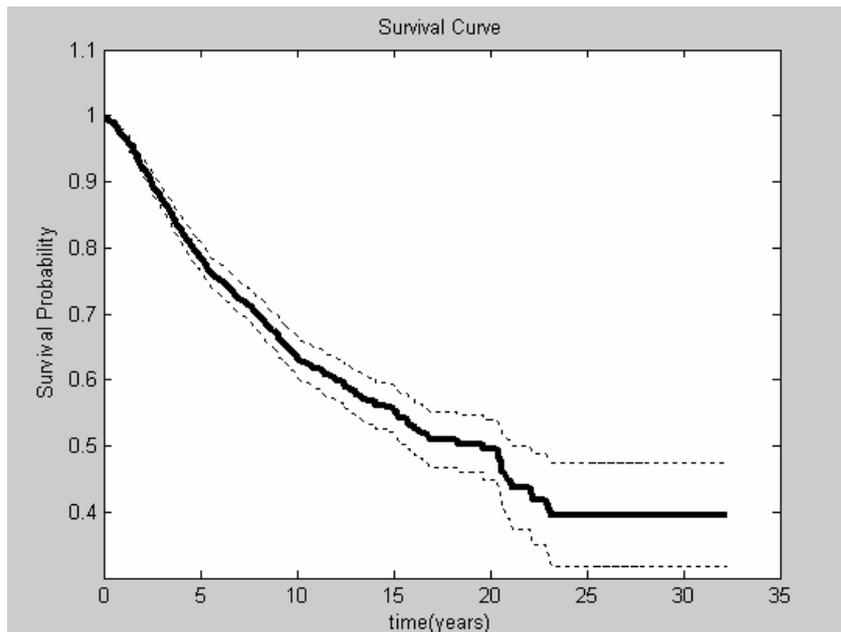


Figure 10b: The survival curve of the 1743 patients from the Kaplan-Meier method without the last death taken into account.

From the above survival functions, the median survival can also be derived, which is the 50th percentile. The results will be the same either we derive the median survival from the figure 10a, either from the figure 10b. As it can be seen from the following plot, the median is derived from the first figure.

The median survival time is the time at which half the patients have died and half are still live. It is easy to derive the median survival from the survival curve, because a horizontal line must be drawn at 50% survival and then it must be observed where this line crosses the curve. From the following plot it is shown that the median survival is at 18 years almost.

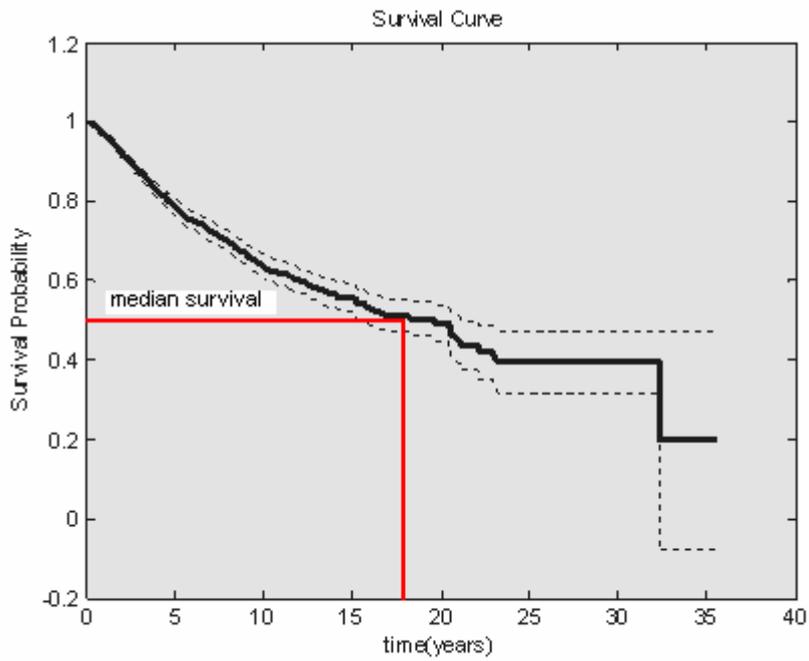


Figure 11: The survival curve with the median survival

The cumulative hazard function can easily be derived by using the relationship between the survival and the hazard function: $H(t) = -\log S(t)$

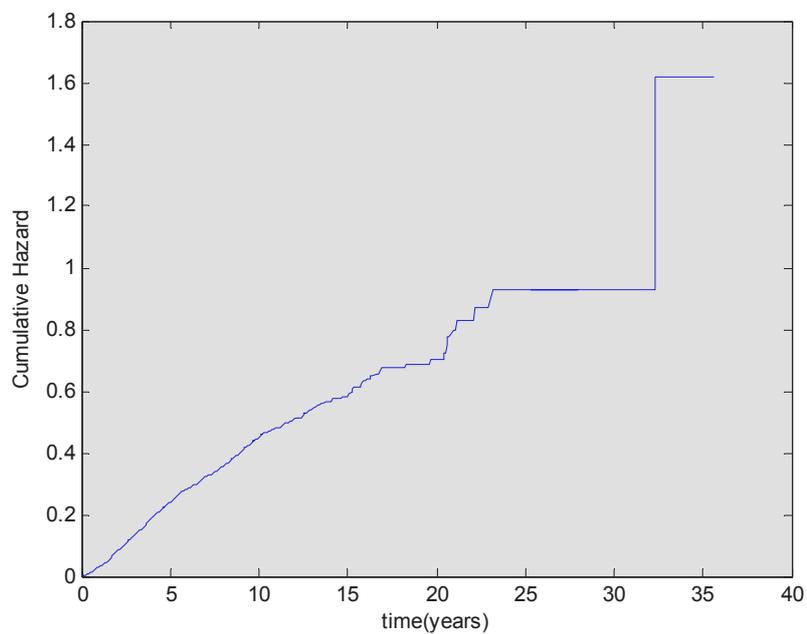


Figure 12a: The cumulative hazard function

The possible statistical error, between the years 23,21 and 32,33, which exists in the survival function, also exists in the Cumulative hazard function. If we ignore the last dead patient at the 32,23 years, then the cumulative rate will be at about 1. But if we don't, as it is seen from the above plot, the cumulative hazard goes up to 1,6. As it was said above, the last patient that dies affects our results, and that is why it can be ignored. The following figure shows the cumulative hazard function without the last death taken into account.

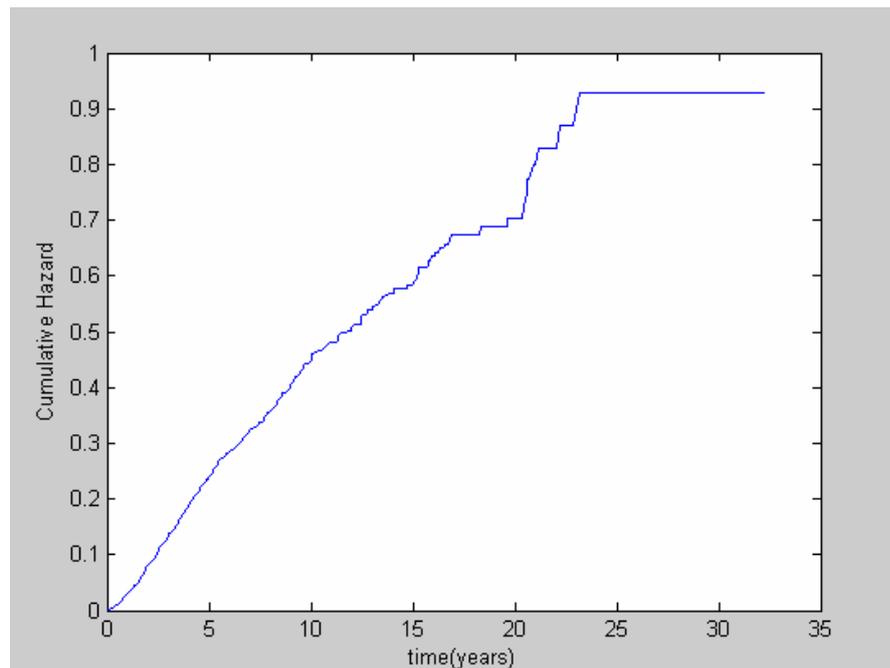


Figure 12b: The cumulative hazard function without the last death taken into account

The survival curve and the cumulative hazard function that are derived from the Kaplan-Meier estimator were also computed with the SPSS program so that they can be comparable.

In order to obtain results from the SPSS program no code has to be written. The only thing that has to be done is to import the data into the data editor which will contain two columns, one with the survival years and another one with the number of deaths. Then from the menu the following order has to be followed: Analyse → Survival → Kaplan-Meier. Then a few adjustments have to be done and the plots of the cumulative survival and hazard functions are ready. [Appendix A]

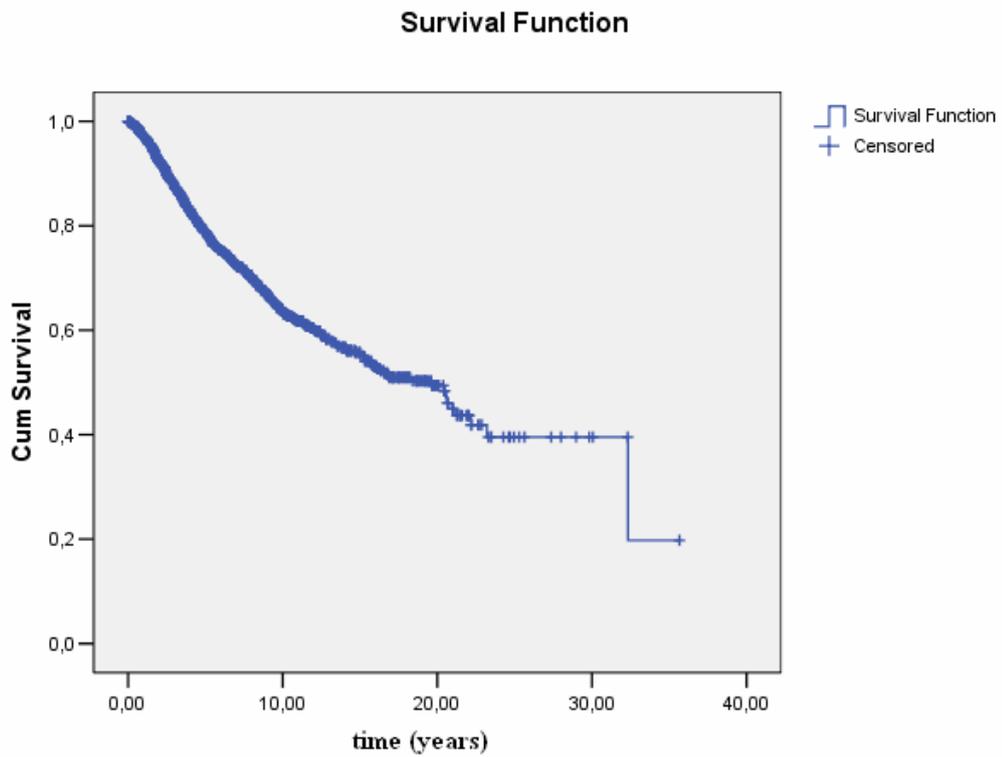


Figure 13: The survival function from the SPSS program

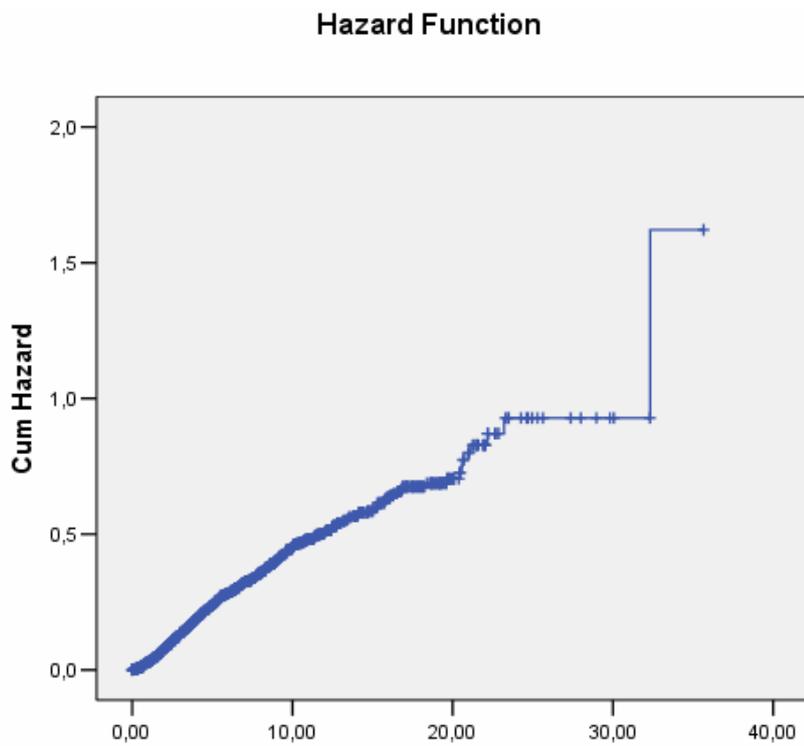


Figure 14: The cumulative hazard function from the SPSS program

In the survival function as well as in the cumulative hazard function the censored patients can be identified by the mark “+”. As it can be seen the survival functions and the cumulative hazard functions are very much alike since they describe the same events in the exact times.

From the SPSS program the following table came up that summarizes the confidence intervals and the median survival. As it can be seen the median that was taken from the survival function of the Kaplan-Meier method and the median that came up from The SPSS program are very close. The former is at 18 years and the later is at 19, 6 years. Also the upper and lower bound of the Confidence Intervals are very close.

Mean(a)				Median			
Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound			Lower Bound	Upper Bound
19,289	,788	17,746	20,833	19,660	1,297	17,118	22,202

Table 17: Contains information from the SPSS about the confidence intervals and the median of the survival function

We can also use the Kaplan-Meier method so as to come up with the survival functions of two subgroups. In this point, we have categorized our data in two subgroups according their sex. So in the following plot, the survival functions of the male and female patients of our data are shown. With the blue line, the survival function of the females is shown and with the green the line the survival function of the males is shown. As it can be seen, the survivability of the females is greater than the survivability of the males, since for the former the survivability goes up to 50%, while for the later it goes up to almost 20%. In this point, we can say that the death that affects our results is a death of a male patient. If this death, as we already said before, can be thought of as a statistical error and be ignored, then the survivability of the male patients go up to 30%. In either case, ignoring or not the last patient, the female patients have greater survivability than the male patients.

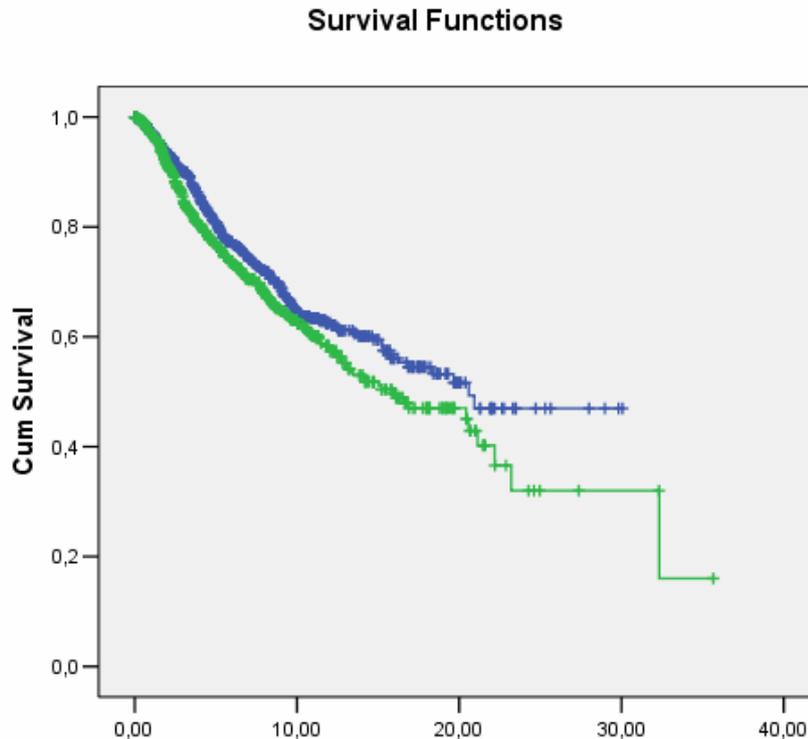


Figure 15: Survival functions of the male and female patients

5.1.1 Fitting procedure of the survival function

The above survival plots were derived from the Kaplan-Meier estimator which means that the plot was constructed by the points of the set that came up during the computation of the estimator. These points have to be fitted to known survival distributions so as the survival data to be approximated. Up to this point, the survival functions that were plotted were just points that came up from the statistical analysis, and not an exact plot of a known survival distribution. With the fitting procedure, instead of having a plot that will represent the survivability in points, we will have a plot of the survival function, since after the following procedure, the function of the survivability will be known.

Two survival distributions have been chosen and tested but only the one of them fitted suitably the data.

- Exponential distribution
- Log-Logistic distribution

The fitting procedure was done with the “cftool” of Matlab. [Appendix A]

The other distribution was rejected because its goodness of fit was not that good. The Curve Fitting Toolbox supports the following goodness of fit statistics for parametric models and each of them must have a specific value which ensures the good fit.

- The sum of squares due to error (SSE)
- R-square
- Adjusted R-square
- Root mean squared error (RMSE)

Sum of Squares Due to Error: This statistic measures the total deviation of the response values from the fit to the response values. It is also called the summed square of residuals and is usually labelled as SSE. A value closer to 0 indicates a better fit.

$$SSE = \sum_{i=1}^n w_i (y_i - \hat{y}_i)^2$$

R-Square: This statistic measures how successful the fit is in explaining the variation of the data. In another way, R-square is the square of the correlation between the response values and the predicted response values. R-square is defined as the ratio of the sum of squares of the regression (SSR) and the total sum of squares (SST). SSR is defined as:

$$SSR = \sum_{i=1}^n w_i (\hat{y}_i - \bar{y})^2$$

SST is also called the sum of squares about the mean, and is defined as:

$$SST = \sum_{i=1}^n w_i (y_i - \bar{y})^2$$

where $SST = SSR + SSE$. Given these definitions, R-square is expressed as:

$$R - square = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

R-square can take on any value between 0 and 1, with a value closer to 1 indicating a better fit.

For example, an R2 value of 0.8234 means that the fit explains 82.34% of the total variation in the data about the average.

Degrees of Freedom Adjusted R-Square: This statistic uses the R-square statistic defined above, and adjusts it based on the residual degrees of freedom. The residual degrees of freedom is defined as the number of response values n minus the number of fitted coefficients m estimated from the response values.

$$v = n - m$$

v indicates the number of independent pieces of information involving the n data points that are required to calculate the sum of squares. The adjusted R-square statistic is generally the best indicator of the fit quality when additional coefficients are added to the model.

$$\text{adjusted } R\text{-square} = 1 - \frac{SSE(n-1)}{SST(v)}$$

The adjusted R-square statistic can take on any value less than or equal to 1, with a value closer to 1 indicating a better fit.

Root Mean Squared Error: This statistic is also known as the fit standard error and the standard error of the regression

$$RMSE = s = \sqrt{MSE}$$

where MSE is the mean square error or the residual mean square

$$MSE = \frac{SSE}{v}$$

A RMSE value closer to 0 indicates a better fit. [oth-1]

The Exponential distribution

The exponential survival distribution is characterized by a constant hazard rate λ , its only parameter. A high λ value indicates high risk and short survival while a low λ value indicates low risk and long survival. [b-1]

$$S(t) = e^{-\lambda t}$$

The goodness of fit statistics for this distribution where:

sse	0,4506
rsquare	0,9819
adjrsquare	0,9819
rmse	0,0215

Table 18: Contains the goodness of fit of the exponential distribution in the Kaplan-Meier method

The fitting procedure was done with the curve fitting tool of the Matlab. The exponential survival distribution had to be created, since it was included in the custom fitting equations. The plot that shows the Kaplan-Meier estimator with the fitting of the exponential distribution is the following:

As it can be seen from table 15, the goodness of fit is not very good, since the sse and the rmse must be close to zero so as to be a good fit, and the rsquare and the adjrsquare must be close to 1. There is a deviation of these values and that is why, we will also test the log-logistic survival distribution.

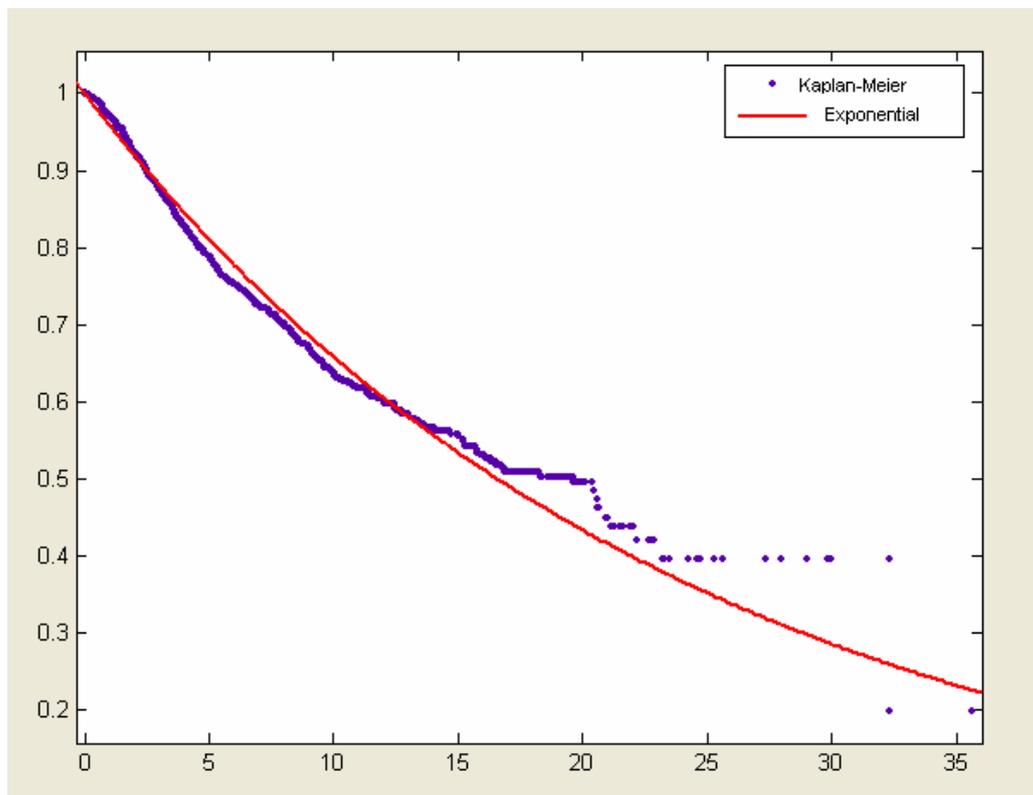


Figure 16: The fit of the exponential survival distribution in the Kaplan-Meier method

The Log-Logistic distribution

The log-logistic distribution is characterized by two parameters α , and γ . [b-1]

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

The goodness of fit statistics for this distribution where:

sse	0,09658
rsquare	0,9961
adjrsquare	0,9961
rmse	0,009958

Table 19: Contains the goodness of fit of the log-logistic distribution in the Kaplan-Meier method

These values reveal that the fit will be good as they are very close to the suggested values, which is confirmed by the following plot.

As it can be seen, the estimated points and the log-logistic survival distribution are very close and especially the first 20 years.

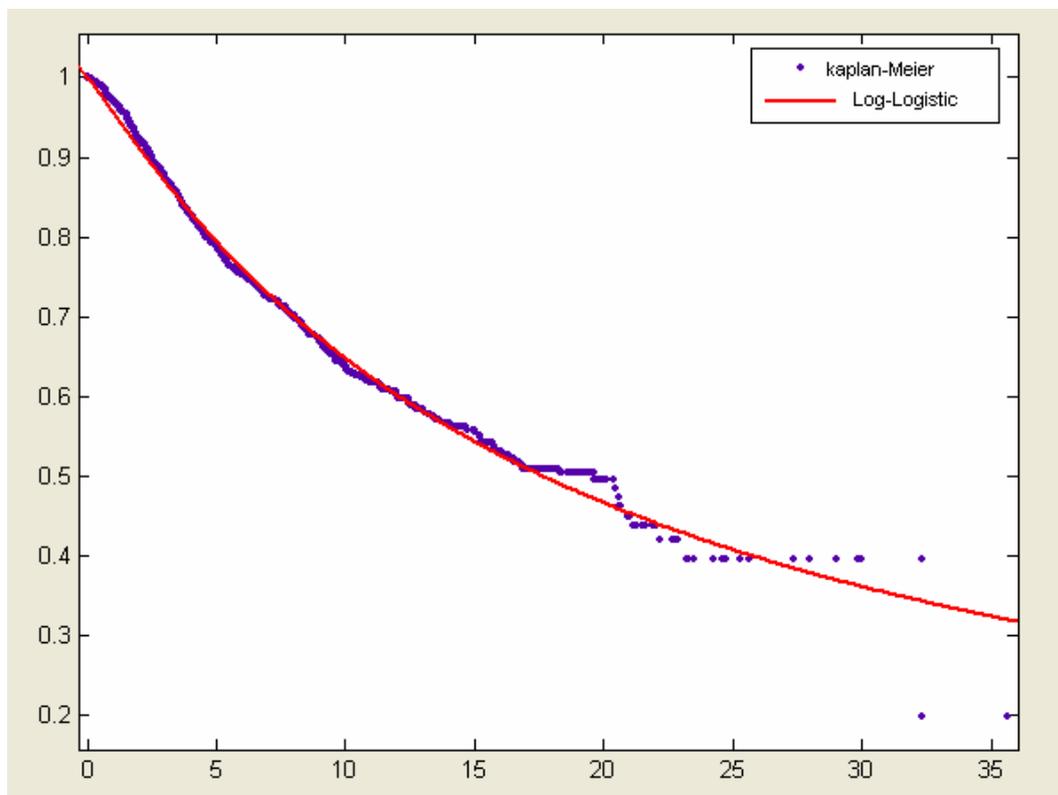


Figure 17: The fit of the log-logistic survival distribution in the Kaplan-Meier method

Between the two survival distributions, it is obvious that the log-logistic survival distribution gives better results, since the goodness of fit is obviously better. The sse and the rmse must be really close to zero, and the rsquare and the adjrsquare are very close to 1. This is why, the log-logistic distribution is chosen so as to fit the Kaplan-Meier estimator and to derive the function of the survivability.

The parameters α and γ have the following values:

α	0,0454
γ	1,079

Table 20: Contains the parameters of the log-logistic distribution in the Kaplan-Meier method

The survival function, based on the above parameters and the log-logistic distribution is: $S(t) = \frac{1}{1 + 0,0454t^{1,079}}$

5.2 Results from the Life-table method

In this point of this chapter, the results of the Life-table method will be given. The survivorship function, the hazard function and the probability density function were computed with the help of Matlab and of SPSS so as to be comparable. [Appendix A]. The SPSS procedure is exactly the same for the Life-table method, except the fact that the following path must be followed: Analyze \rightarrow Survival \rightarrow Life-table. Also the time intervals must be defined.

1. Survival function

- Matlab

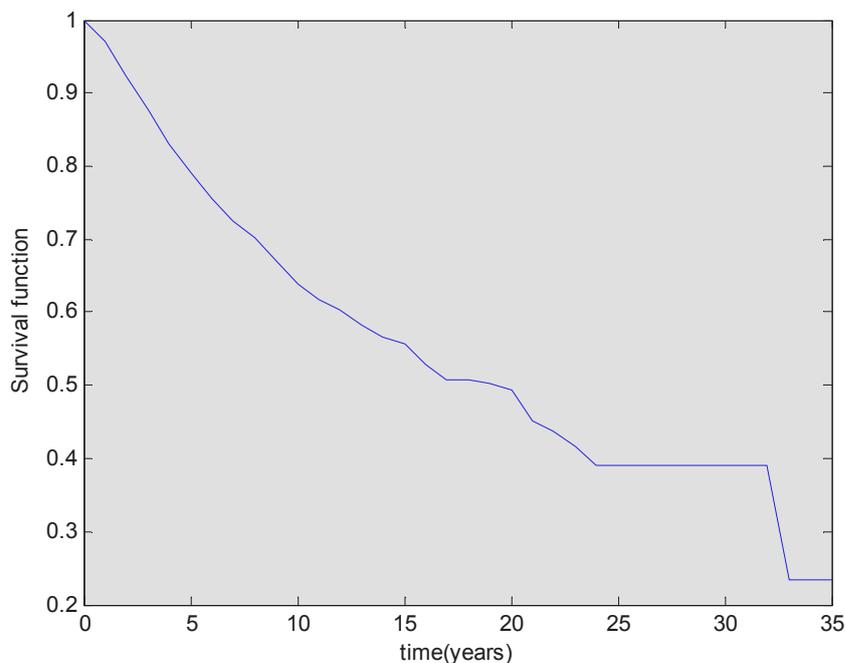


Figure 18a: The survival function of the Life-table method from Matlab

As we mentioned before in the Kaplan-Meier method, the last death provokes problems in the analysis since it can be thought of as statistical error. If the last death is not taken into account, the probabilities of cure are augmented from 20% to 40%. The following figure shows exactly this fact. Also in the following plot the confidence intervals are shown. During the last years, the confidence intervals become wider because not so many events take place as during the first years.

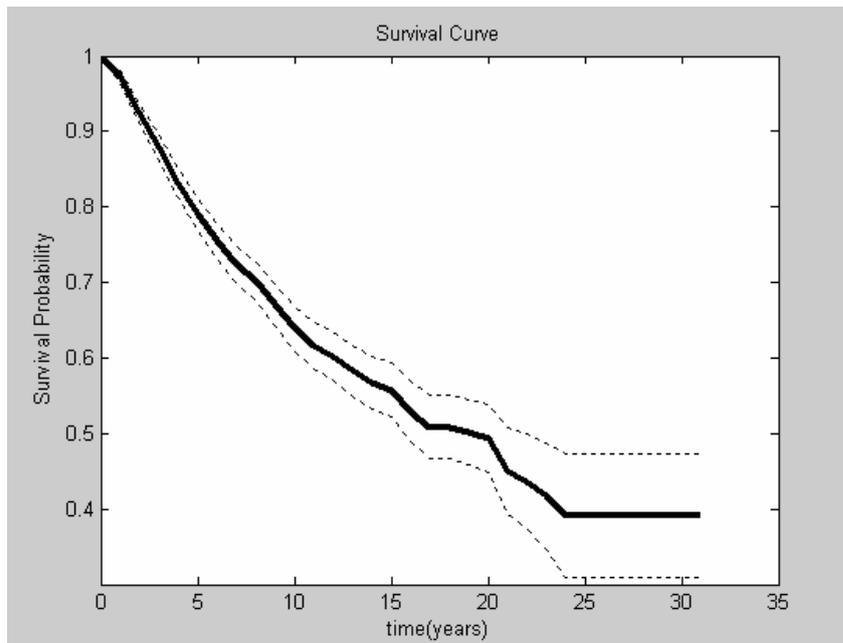


Figure 18b: The survival function of the Life-table method from Matlab that does not contain the last death.

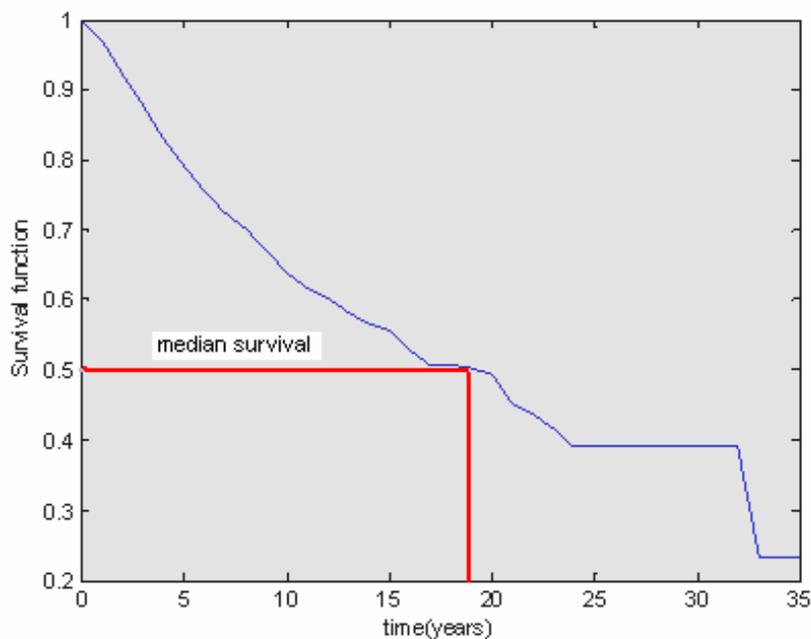


Figure 19: the survival function with the median survival of the Life-table method

As it can be seen from the above plot, the median survival is almost at 19 years, which is really close with the median survival that occurred from the Kaplan-Meier method, where the median was at the 18 years. The median survival is the same either it is derived from the survival function that contains all the data, either is taken from the survival function that does not contain the last death.

- SPSS

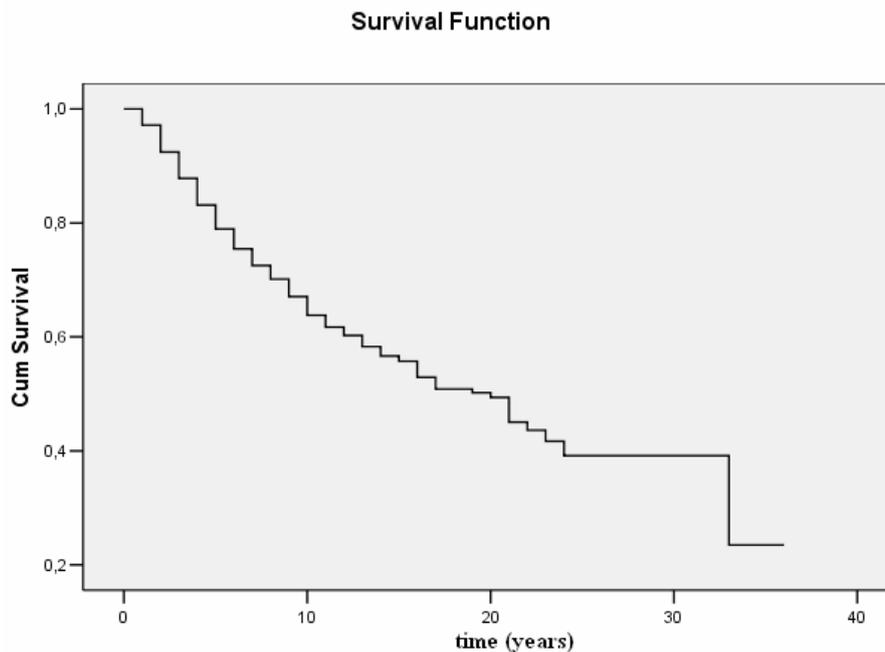


Figure 20: The survival function of the Life-table method with the SPSS program

As it can be observed the Kaplan-Meier method and the Life-table method give almost the same survival probability, which means that these two methods are very much like, except the fact that the former does not use intervals of time, while the latter does. Besides, this is the main difference between these two methods. The above survival is gradual, as the survival from the Kaplan-Meier method was too, which means that high survival rate or longer survival is represented. Just because the Life-Table uses intervals the steps down are clearer and so the survival curve looks like a staircase function.

2. Probability Density Function

- Matlab

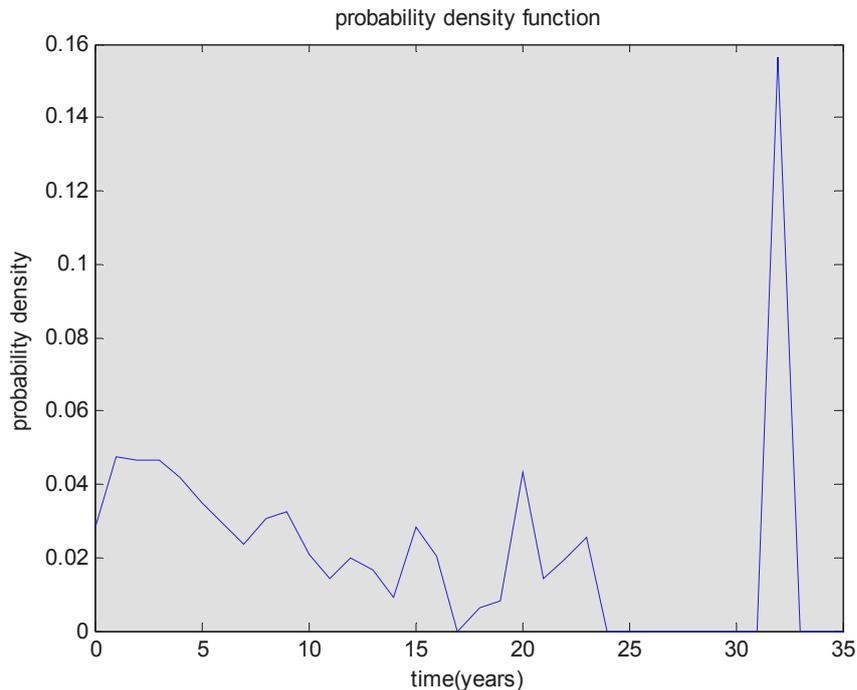


Figure 21a: The probability density function with Matlab

The proportion of individuals that fail in any time interval and the peaks of high frequency of failure can be found from the density function. The probability density function is defined as the probability of dying in the i th interval per unit width.

As it can be observed from the above plot, in the 32,3 years, the biggest probability density is observed because for a long time (23 years-32 years) no patients have died and then suddenly a patient died and that is why a peak is shown at that time. The conditional probability of death is high at that time interval. The peak that is biggest, after that of the 32,3 years is at almost 2 years. This is expected, since during the first two years the most of the deaths take place, and that is why at that time point there is a big frequency of failure. The above plot may be misleading, since it shows that the biggest probability of dying is at the last interval. The real fact is that the biggest probability of dying is at the second year since during that time the most of the patients die. In particular, 73 patients die during that time. The following plot does not take into account the last death and so it gives a clearer image of what is happening.

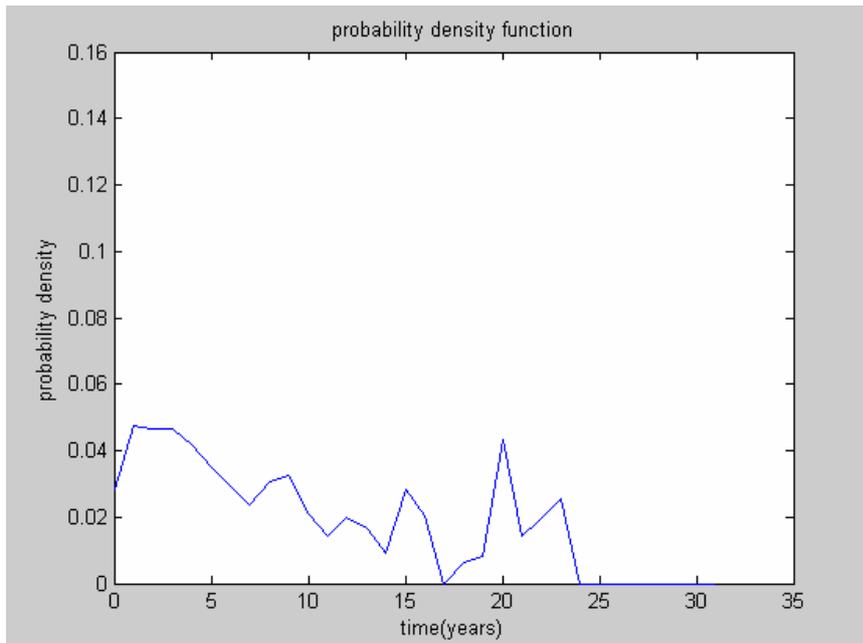


Figure 21b: The probability density function with Matlab without the last death taken into account

- SPSS

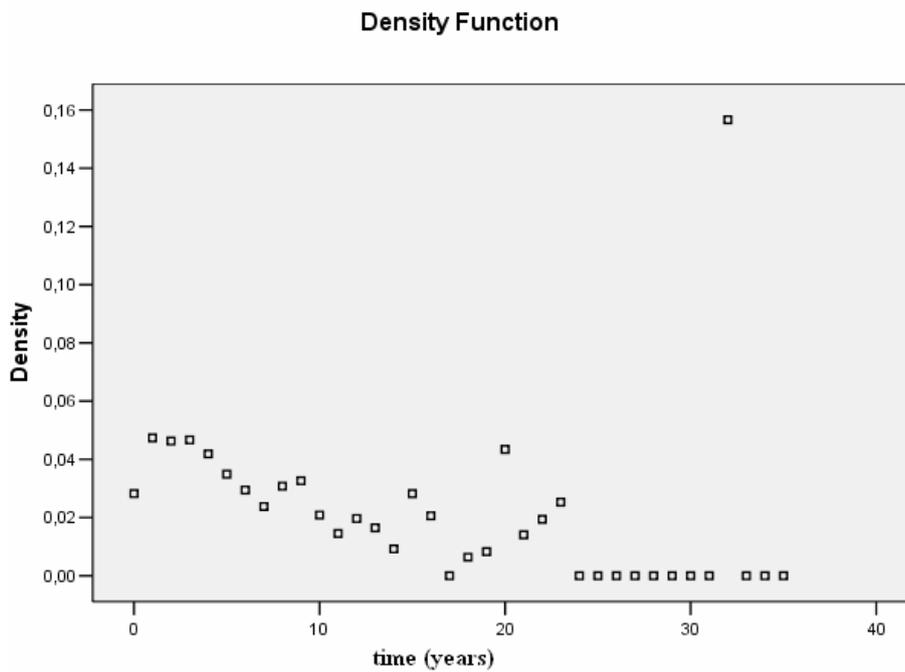


Figure 22: The probability density function with SPSS

As it can be observed the two density functions from the two programs are alike.

3. Hazard function

- Matlab

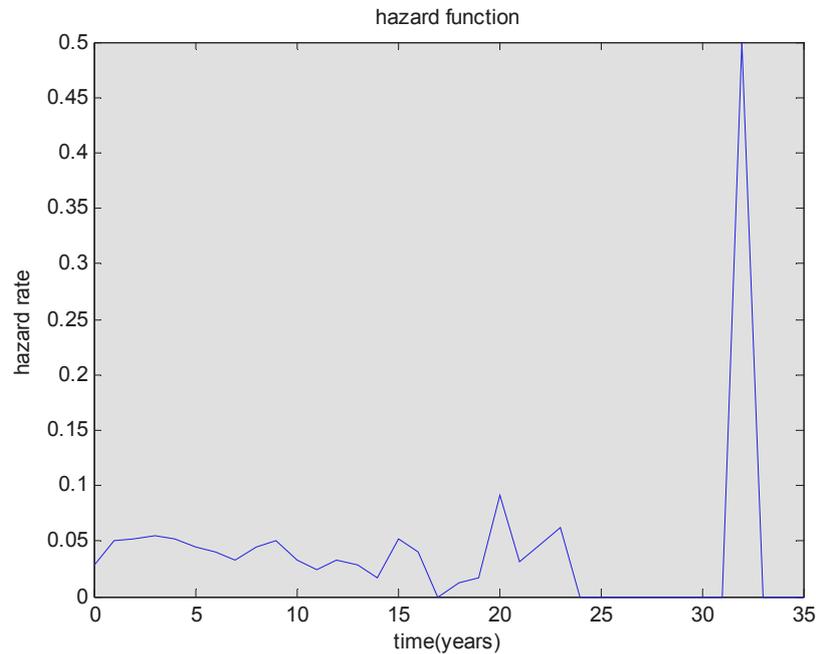


Figure 23a: The hazard function with Matlab

The hazard function for the i th interval is the number of deaths per unit time in the interval divided by the average number of survivors at the midpoint of the interval. From the above plot it can be said that the death rate is highest at the 32 years of the study. For a period of 10 years (23-32) the hazard rate is zero and then suddenly the peak of the hazard rate is observed. This happens due to the fact that for a long period there were not any deaths. During the first 10 years the death rate is relatively constant, taking values from 0.03-0.051. Then it is observed a gradual attenuation in the death rate which is logical since after the first 10 years the number of deaths is obviously less. Only during the fifteenth and twentieth year, the hazard rate is higher because during that time there is a small increase of deaths. Once again, because the above plot may become misleading, the last death was not taken into account and the hazard function was computed again. As it was mentioned before, this last death may be thought of as statistical error and that is why it affects our results.

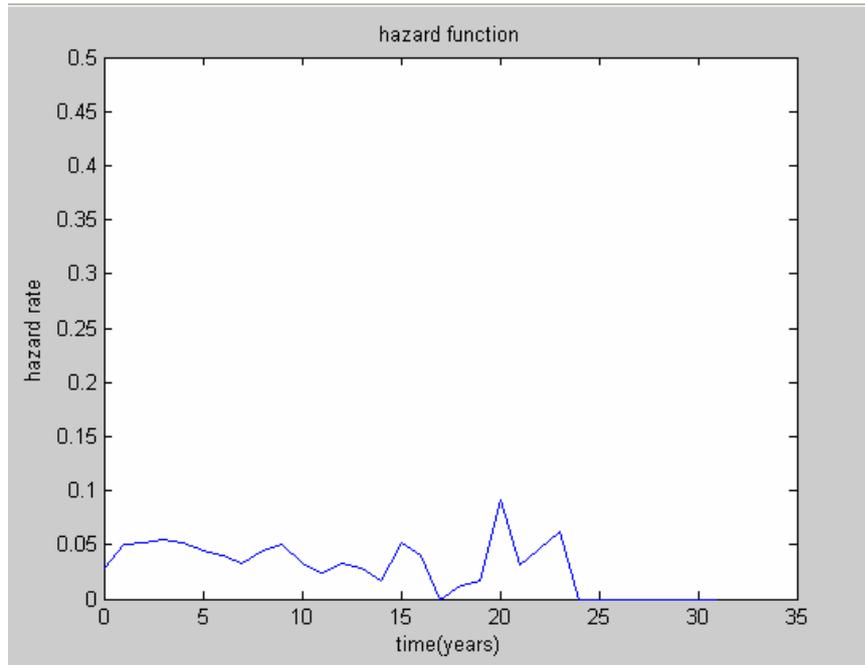


Figure 23b: The hazard function with Matlab without the last death

- SPSS

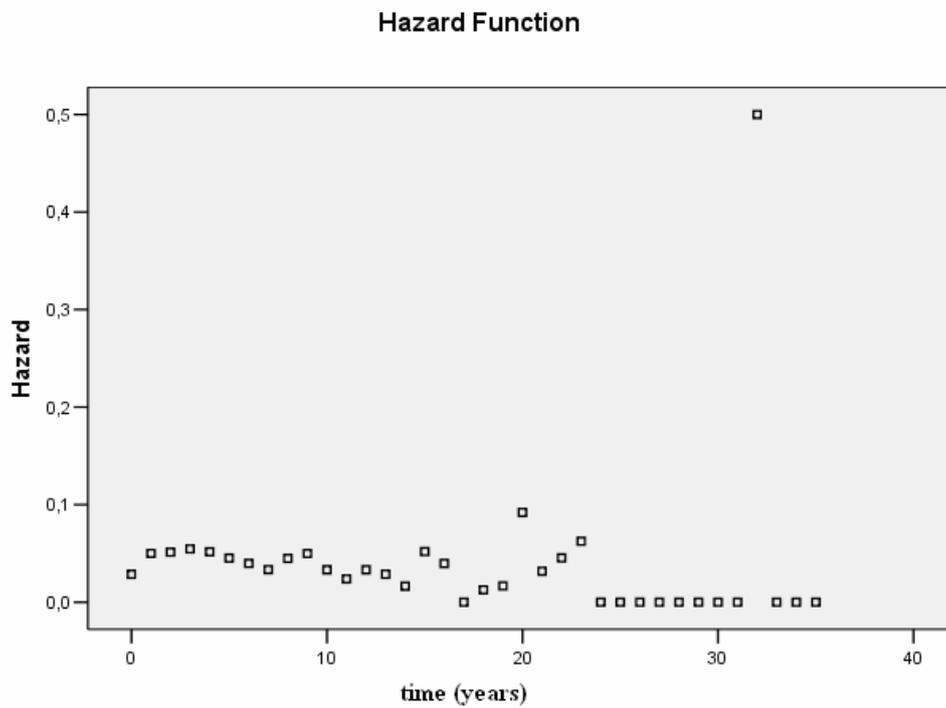


Figure 24: The hazard function with SPSS

As it can be observed, the two hazard functions from Matlab and SPSS are alike.

Even though, the statistical analysis of the Life-table method does not contain the computation of the Cumulative hazard function, we thought that it would be good to derived it from the known relationship $H(t) = -\log S(t)$. In this way, by having the cumulative hazard function, a more complete image of the all the nonparametric models is going to be achieved. In addition, in chapter 6, where all the comparisons are going to be presented, a more complete work will be done.

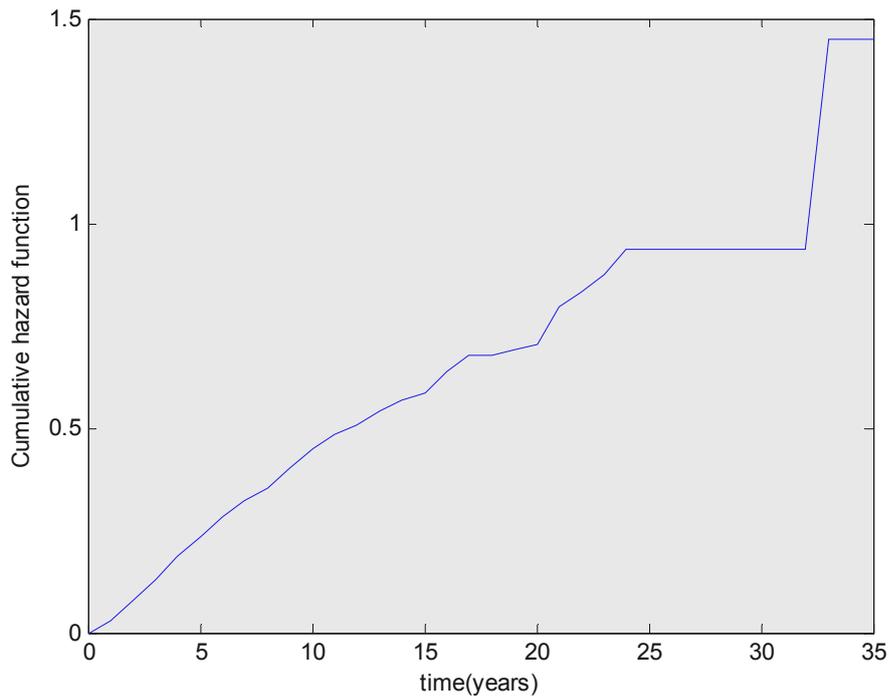


Figure 25: The cumulative hazard function of the Life-Table method

5.2.1 Fitting procedure

In this section of the chapter, the plot of the survival function was derived from the Life-Table method. The survival curve has to be fitted to known survival distributions so as the survival data to be approximated. Since the exponential distribution was checked in the Kaplan-Meier method and was not suitable, it will not be checked again. Only the log-logistic distribution will be tested.

The log-logistic distribution is the following:

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

The goodness of fit statistics for this distribution where:

sse	0,034562
rsquare	0,9764
adjrsquare	0,9757
rmse	0,031883

Table 21: Contains the goodness of fit of the log-logistic distribution in the Life-Table method

These values reveal that the fit will be good as they are very close to the suggested values, which is confirmed by the following plot.

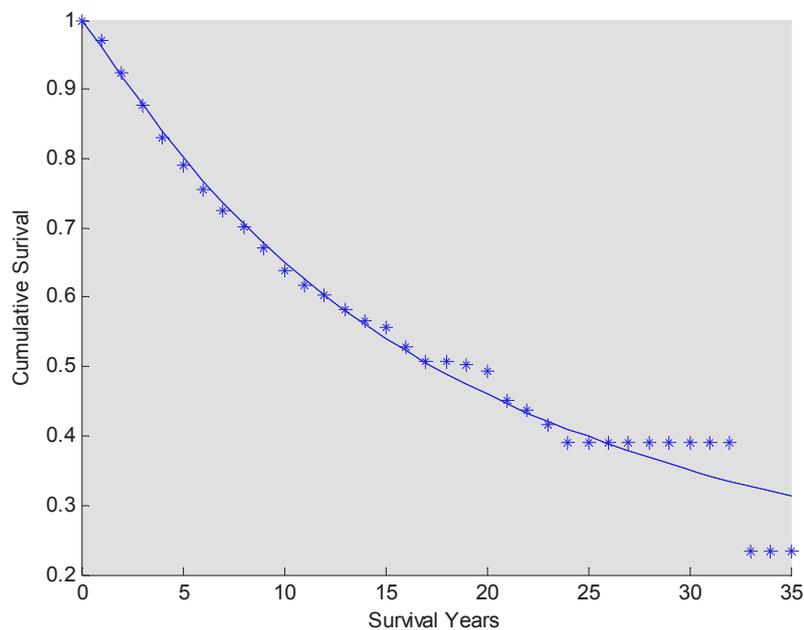


Figure 26: The fit of the log-logistic survival distribution in the Life-table method

The parameters α and γ have the following values:

α	0,04005
γ	1,127

Table 22: Contains the parameters of the log-logistic distribution in the Life-table method

The survival function, based on the above parameters and the log-logistic distribution is is:

$$S(t) = \frac{1}{1 + 0,04005t^{1,127}}$$

5.3 Results from the Nelson-Aalen method

As we already mentioned, the Nelson-Aalen method gives the cumulative hazard function and with the relationship $S(t) = \exp[-H(t)]$, the survival function is computed. At this point, the results of the analysis of the Nelson-Aalen method are given:

The plot of the cumulative hazard function is:

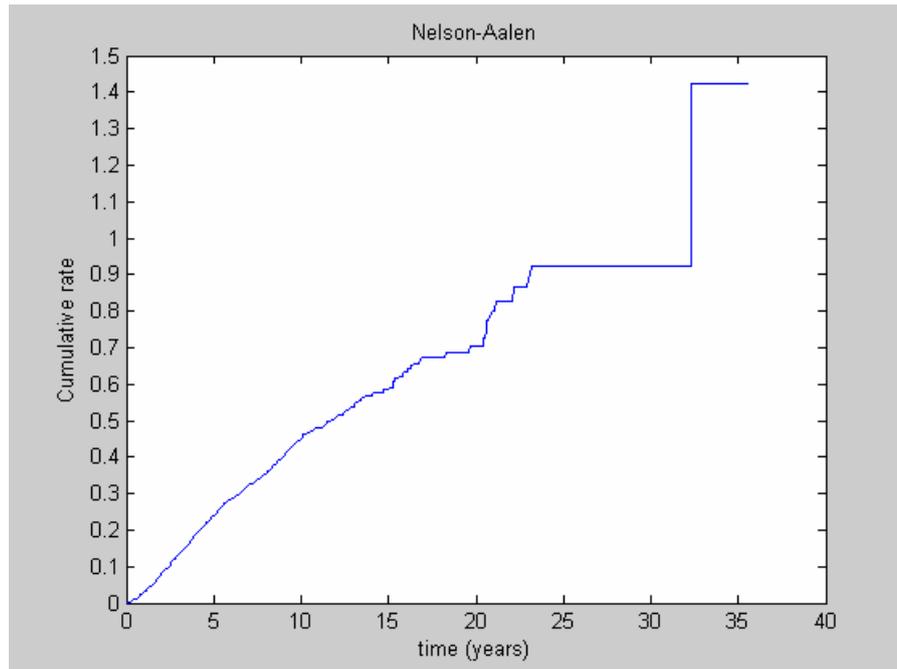


Figure 27a: The cumulative hazard function of the Nelson-Aalen method

Since the data set is the same for all the three methods, the possible statistical error, between the years 23,21 and 32,33, which exists in the previous methods, also exists in the Nelson-Aalen method, and especially in the cumulative hazard function. If we ignore the last dead patient at the 32,23 years, then the cumulative rate will be at about 1. But if we don't, the cumulative hazard goes up to 1,6. As it was said again previously, the last patients that dies affects our results, and that is why it can be ignored. If we ignore the last patient we conclude to the following plot. In addition, this plot contains the confidence intervals which may be very informative. Since during the first years the patients are more, the confidence intervals are narrower. However, during the last years where fewer events take place, the confidence intervals are wider because there is more uncertainty about the results.

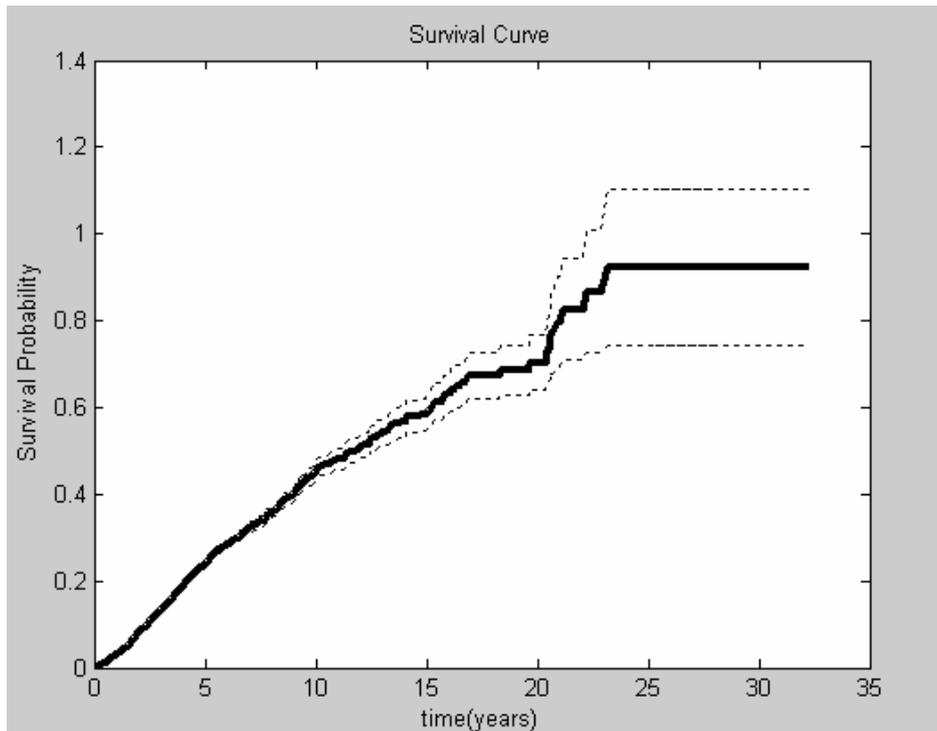


Figure 27b: The cumulative hazard function of the Nelson-Aalen method without the last death taken into account.

The survival function will also be plotted. It can be derived from:

$$S(t) = \exp[-\Lambda(t)]$$

The plot of the survival function is:

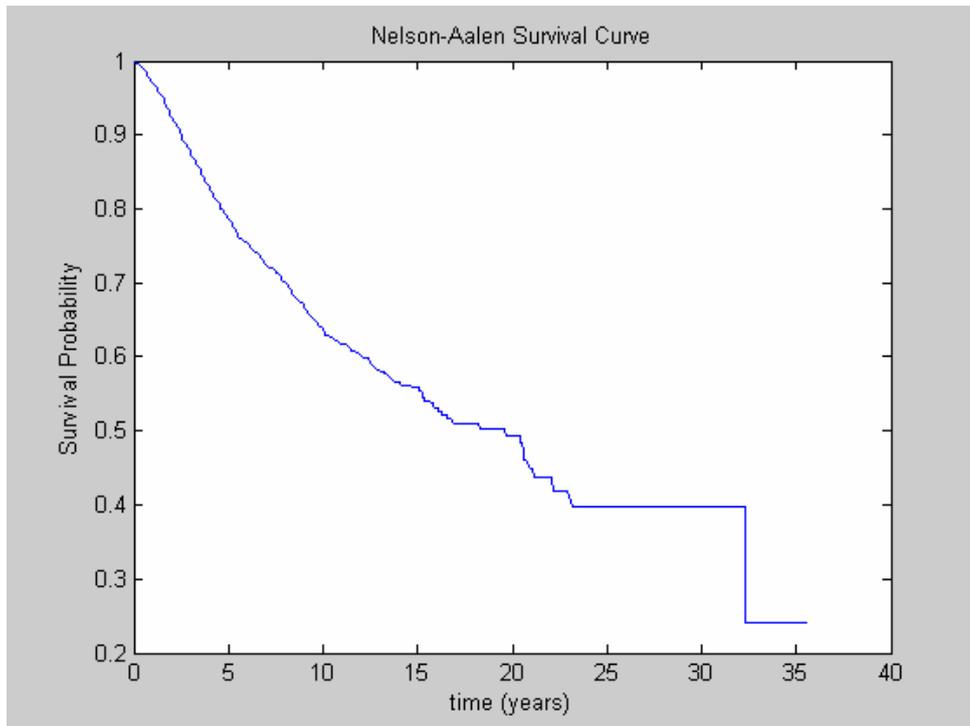


Figure 28: The survival function of the Nelson-Aalen method

At this point we must underline that if the last death is not taken into account, the probabilities of cure will arise to 40%, since the last step down will not take place.

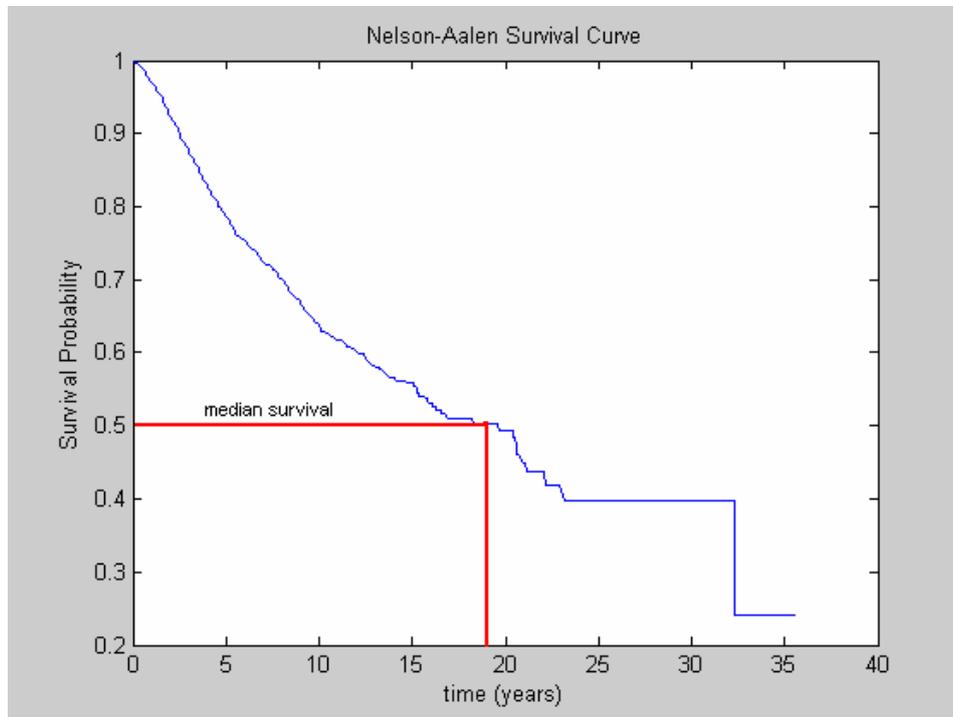


Figure 29: the survival function with the median survival of the Nelson-Aalen method

The median survival is about at 19 years, just like the other two methods. All the three methods, the Kaplan-Meier, the Life-table and the Nelson-Aalen method, have similar results. The survivability of the patients and also the median survival is about the same in the three methods. A more complete comparison of the three methods, their differences and their similarities, will be given in chapter 6.

5.4 Results from the Aalen's additive model

At this point, the cumulative regression functions that were computed by the statistical analysis of this model are going to be plotted against time. This is the scope of this model, to result in plots that will give information on the influence of the covariates. Since the regression functions vary with time, statistical analysis of them reveals changes in the influence of the covariates over time. The information is taken from the slope of the plots. The slope can give three different kinds of information.

- If there is an increasing slope, the covariate influences the evolution of the disease.
- If there is a negative slope, the covariate loses its influence or it has a decreasing effect.
- If the slope is constant, the covariate can no longer be considered as a prognostic factor.

It must be noted at this point that we must not focus on small changes in the slope of the plots but we must restrict to more dramatic changes, like the disappearance of a clearly increasing slope.

It is to be expected that the covariates measured at the beginning of the observation period may often lose their influence on survival after a while. [p-1]. This is something that will be shown from the plots of the regression functions that follow. The plots of the cumulative regression functions for each covariate are presented below.

1. Covariate “lubd”

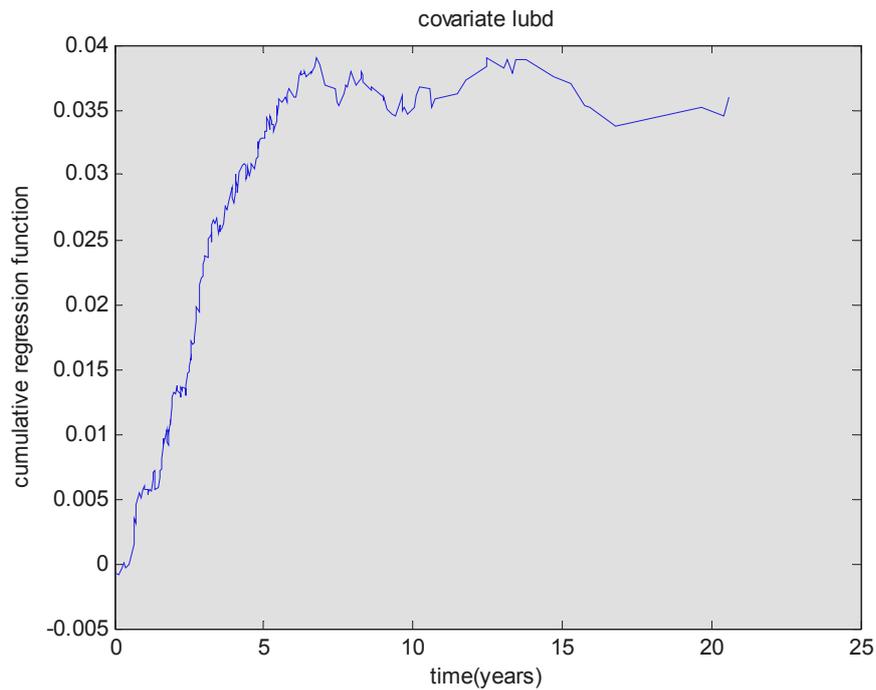


Figure 30: The plot of the regression function that corresponds to the longest ultrasound basal dimension covariate

The covariate “Longest ultrasound basal dimension” has a clear influence during the first seven years, since the slope is clearly increasing. After that time the influence seems to disappear since the slope of the cumulative regression function is almost constant. This means that during the first years of the disease, this covariate plays an important role to the survivability of the patients and it influences the evolution of the disease. Since that crucial time passes, this covariate starts to lose its influence, since the increasing slope disappears.

It is very important to be able to get information about the influence of each one of the covariates because in this way we are able to find out which covariate influences more. In addition, we are able to know the exact time point where a covariate loses its influence. All this piece of information can become very helpful for the doctors since they can adjust the cure of the disease and they can change it while the time passes and the covariates start to lose their effect.

2. Covariate “uh”

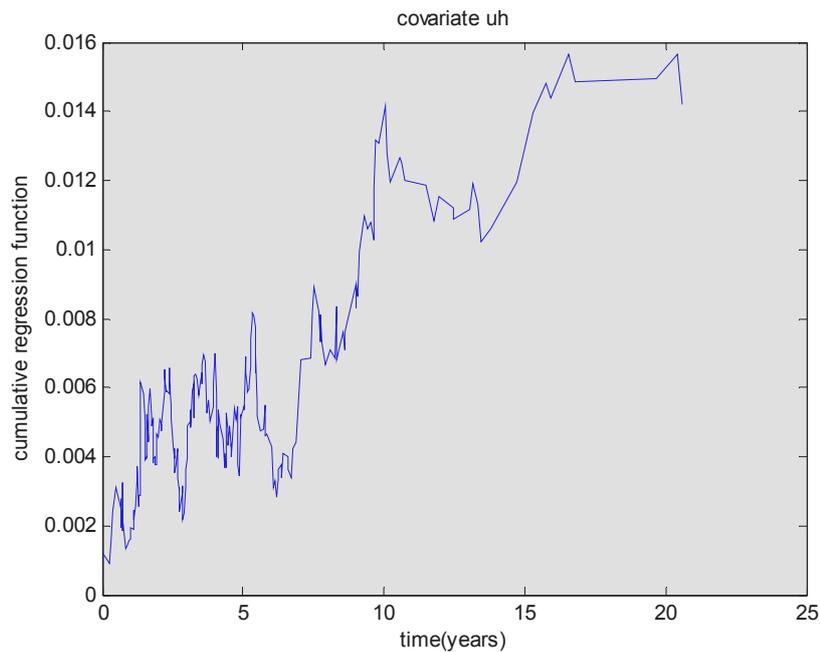


Figure 31: The plot of the regression function that corresponds to the ultrasound height covariate

The covariate “ultrasound height” is not stable throughout the years nor has a clear increasing or decreasing effect. This is due to the fact that the slope increases and decreases constantly while the time passes and so no final result can be taken. In order to conclude some results, the time should be split to small time intervals so as to observe the influence of the covariate in each time interval.

This is not however the scope of the additive method, because the influence of the covariate should be observed through the entire time so as to conclude to results. So what it can be said about this covariate is that it has not a clear influence in the disease of ocular melanoma and so it does not influence the evolution of the disease in a determining and clear way.

This will also be verified below where we put all the covariates together and we try to find which one of them influences more the disease. We will find out that the covariate that influences the disease less is the “ultrasound height” covariate.

3. Covariate “epi”

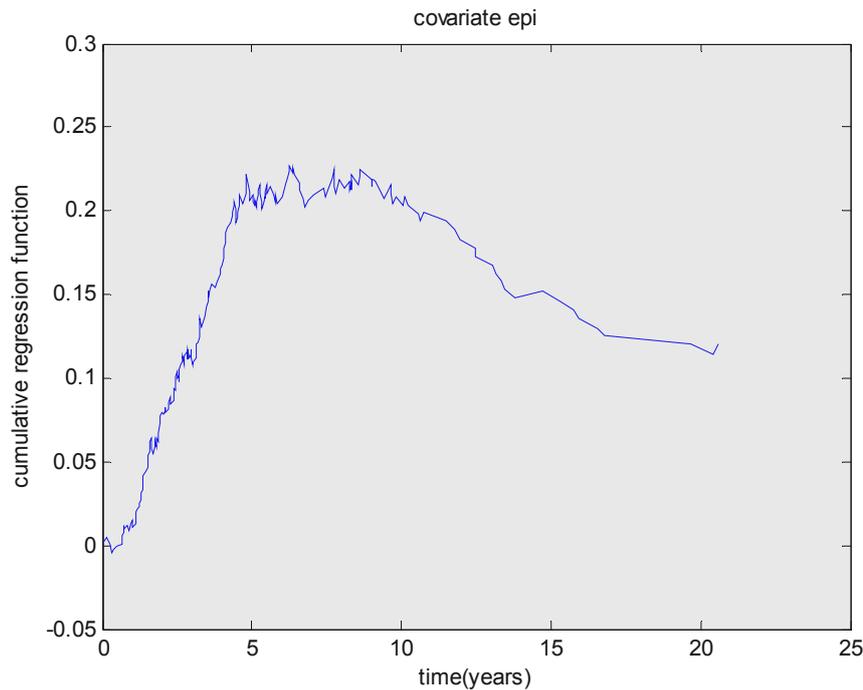


Figure 32: The plot of the regression function that corresponds to the epithelioid cellularity covariate

The covariate “Epithelioid cellularity” has a clear influence during the first five years, since the slope is clearly increasing. After that time point, the slope of the cumulative regression function remains constant and so the covariate does not influence the disease for the next five years. After the ten years the influence seems to disappear, which means that the covariate has a decreasing effect. This is something that could be said after the first five years, where there is a clear influence, because the slope of the cumulative regression function begins to disappear.

At this point, we must remind that a negative slope means that the covariate loses its influence or that it has a decreasing effect, an increasing slope means that the covariate influences the evolution of the disease and that a constant slope means that the covariate is no longer considered as a prognostic factor, since it has no influence on the disease. In other words only when we have an increasing slope the covariate influences the disease. When we have a decreasing and constant slope, the covariate starts to lose its influence until the point where it does not influence the evolution of the disease at all, which is the point of the constant slope.

What it can be said for this method is that when the plots are interpreted, they represent a useful supplement to other kind of analysis, such as the Cox analysis. In particular, it is expected that covariates measured at the beginning of the observation period may often lose their influence on survival after a while. This is something that happened with the above covariates. It is very useful to have a method that can reveal the change of the influence of the covariates during the time.

In general, as we already mentioned before, in order to interpret the influence of the covariates, one should not focus on smaller changes in the slope of the plots, but restrict to more dramatic changes, like the disappearance of a clearly increasing slope.

At this point, the three cumulative regression functions have been put together in a plot so as to observe which one of the covariates influences more the evolution of the disease. So far we have only tested how each covariate separately influences the disease while time passes. Now we will examine, which one of the covariates influences more the disease.

This will be done only for the first six years, because as it was observed from the previous plots, the covariates had a clear influence only during the first years. While time passes the covariates started to loose their influence and that is why there is no meaning in examining the rest of the years.

Once again, the results depend on the slope of each regression function. Since the slope of the entire function must be found in order to conclude to a result, and not the slope at each point, the cumulative regression functions were fitted to linear functions. In this way it is easy to come up with clear conclusions. The fit was done with the cftool of Matlab.

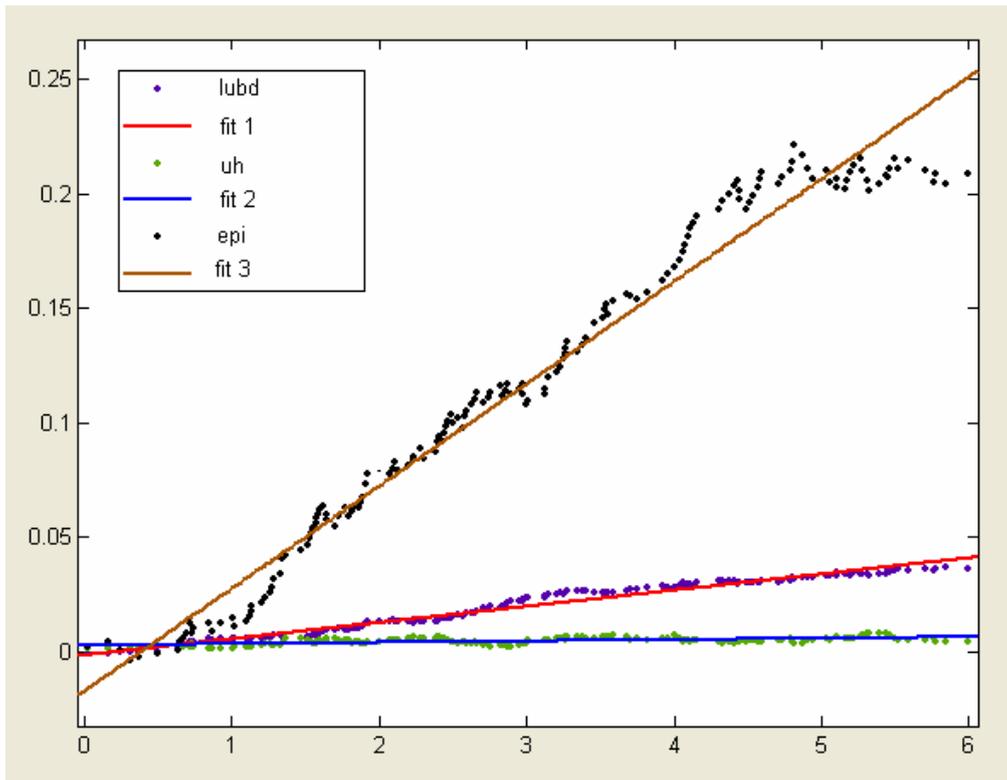


Figure 33: This diagram gives information about which covariate influences more the disease

The black plot represents the cumulative regression function for the “epi” covariate. The green plot corresponds to the “uh” covariate and finally the mauve one corresponds to the “lubd” covariate. The fit was done to linear functions so as to be able to find the slope.

From the above plot it is clear that the covariate “epi” influences more the survival, since it has the bigger and the steepest slope. After the “epi” covariate, the “lubd” covariate influences more the survival and finally the “uh” covariate influences the survival but in a less profound way. This confirms the result that came up from the interpretation of the plot of the cumulative regression function that corresponded to the “uh” covariate. The conclusion of the interpretation of the plot was that the covariate “uh” had no clear and determining influence in the survival.

5.4.1 Division of the dataset into two subgroups

As it was said, the analysis of the additive model was done for 743 patients. At this point, the analysis will be done again from the beginning but this time not for the entire population of the 743 patients but for two subgroups of the population. These subgroups were divided according to the value of the covariate antmar. The antmar, as we mentioned earlier in this thesis, is a categorical covariate and it actually divides the patients into two subgroups according to the point where the ocular melanoma is located. Let's repeat that the ciliary body and the choroidal are the two main points where the ocular melanoma is located. In these two subgroups the patients are divided. So, what is important at this point is to understand that the covariate antmar is not included in the analysis of the additive model, the covariates that are studied are still three, but is only used in order to divide the patients into two main subgroups. The two subgroups are summarized below.

Subgroups	Value of antmar
Choroidal (value 0)	Antmar ≤ 9
Ciliary (value 1)	Antmar > 9

Table 23: Contains the two subgroups that can be derived from categorization of the anterior tumor margin covariate

Subgroups	Type of antmar	Number of patients
Group A	Choroidal	446 patients
Group B	Ciliary	297 patients

Table 24: Contains the number of the patients of the two subgroups

The analysis of the additive model was done for these two subgroups separately from the beginning so as to come up with conclusions for the influence of the covariates and compare these results with those from the population of the 743 patients. The code is exactly the same and for that reason it will not be presented again. The only thing that changes is the size of each subgroup.

In order to become clearer the fact that the patients are divided into two subgroups according to the type of the covariate antmar, the survival functions of these two subgroups are given. These functions were plotted by taking the mean value of the three covariates in the analysis.

It should be noted that the analysis for the subgroup “ciliary body” stops at 15, 77 years and for the subgroup “choroidal” stops at 20, 58 years.

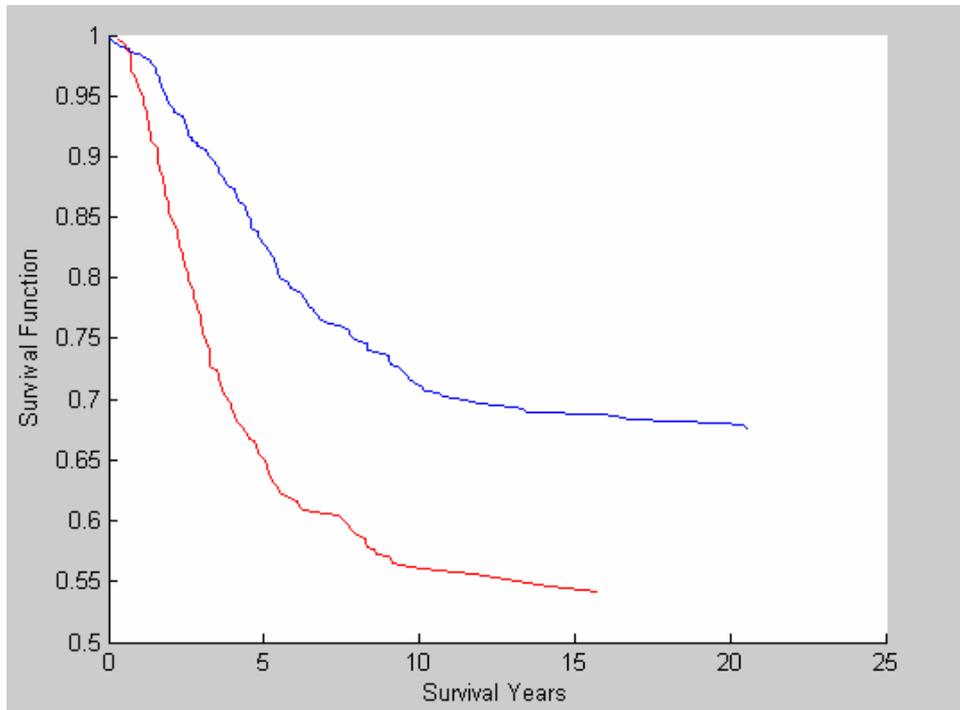


Figure 34: The plot that contains the survival functions of the two subgroups choroidal and ciliary body.

It is clear from the above plot that the blue line corresponds to the subgroup “choroidal”, since the survival function of this subgroup goes up to 20, 58 years. The other survival function, the red one, corresponds to the subgroup “ciliary body”, since the analysis for this subgroup stops at 15, 77 years. These two survival functions are very informative because they show that the patients that suffer from ocular melanoma which is located at the ciliary body have less probabilities of cure than those patients that suffer from ocular melanoma located at the choroidal. In addition, it becomes clear the fact that the covariate antmar divides effectively the population into two subgroups.

At this point it would be suitable to remind that the survival functions are not necessarily monotonic over the whole time period. They may increase at some points and this is a weakness of the additive model, as we mentioned earlier at the description of the model.

From the analysis of the additive model for the entire population, we concluded that the covariate “epithelioid cellularity” is the one that influences more the evolution of the disease. The main reason that we divided the population of the patients into these two subgroups is to check whether the same covariate is the one that influences more the disease either a patients suffers from ocular melanoma located at the “ciliary body” either a patients suffers from ocular melanoma located at the “choroidal”.

In order to conclude to the plots of the cumulative regression functions, the same procedure that was followed in the previous section of this chapter is going to be followed also here. This means that the same plots are going to be given.

5.4.1.1 Ciliary body

a. Covariate “lubd”

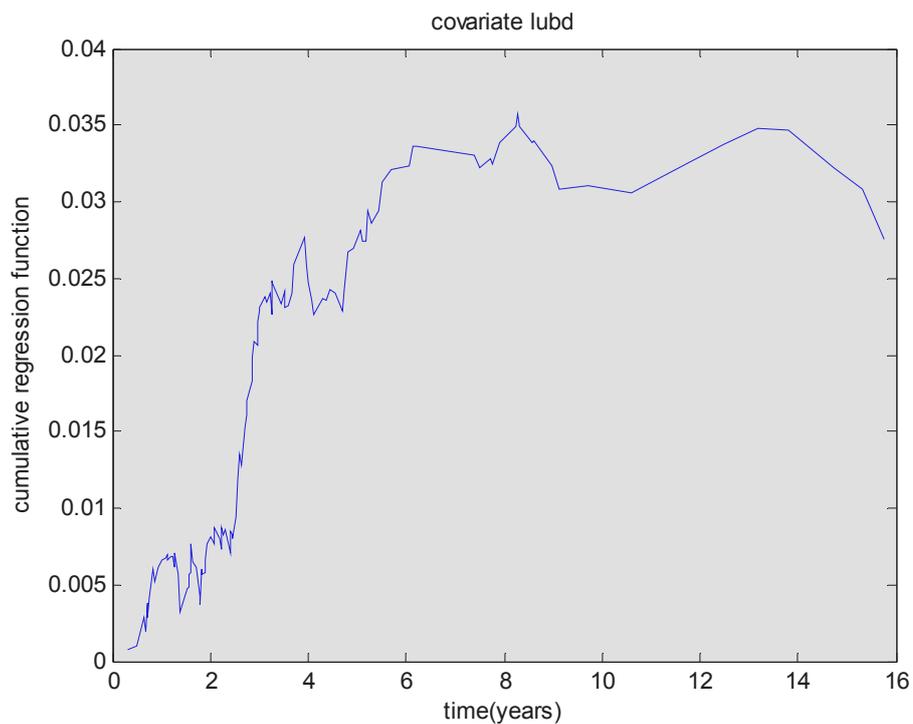


Figure 35: The plot of the regression function that corresponds to the longest ultrasound basal dimension covariate for the ciliary body

The covariate “Longest ultrasound basal dimension” has an influence in the survival during the first six years, but after that time the influence seems to disappear since the slope of the cumulative regression function is almost constant. As it was mentioned before, in order to make conclusions, big changes in the slope of the plots must be observed and not small ones. In this plot the influence of the covariate “lubd” is not as obvious and clear as it was for the population of the 743 patients, since in some time points during the first six years, the slope decreases but in the overall plot, it is obvious that at first there is an increasing effect and then this effect disappears. This result is satisfying enough to interpret the influence of the covariate.

b. Covariate “uh”

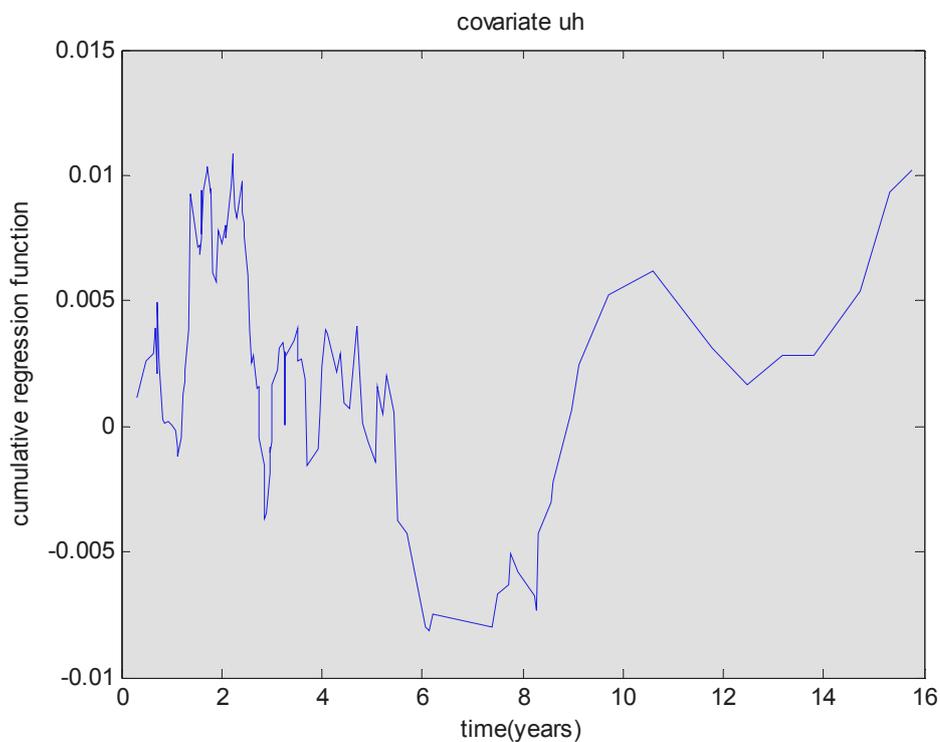


Figure 36: The plot of the regression function that corresponds to the ultrasound height covariate for the ciliary body

Once again, no clear and explicit results can arise from this cumulative regression plot. The slope is not clearly decreasing or increasing throughout the time period but is constantly changing. For this reason, it can be said that no specific results can come up from the interpretation of the above plot for the influence of the covariate “ultrasound height” to the evolution of the disease. For the subgroup of the 297 patients, the covariate “ultrasound height” has the same behaviour as the

population of the 743 patients. This means that it is expected that this covariate will have the smallest influence in the disease, when all the covariates will be put together. This will be checked in a following plot.

c. Covariate “epi”

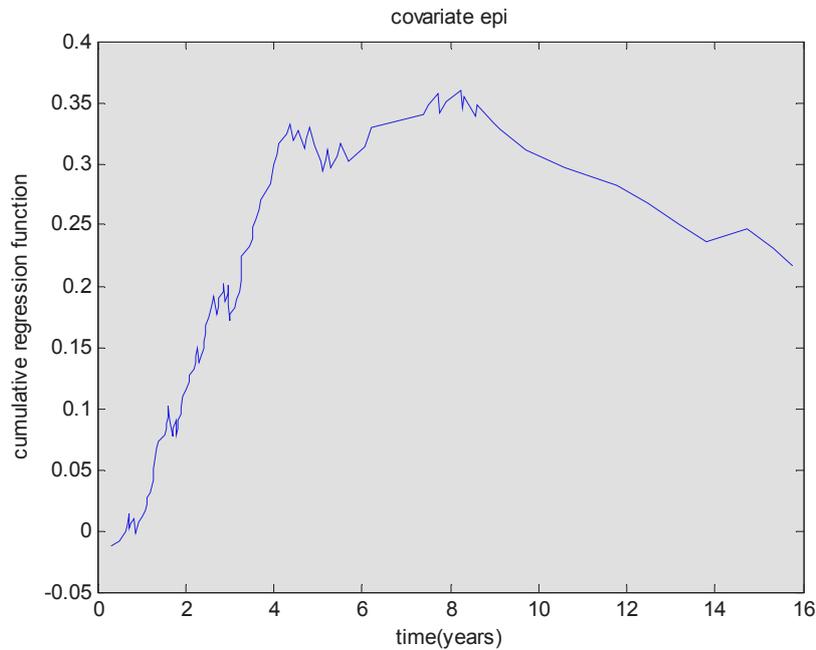


Figure 37: The plot of the regression function that corresponds to the epithelioid cellularity for the ciliary body

The covariate “Epithelioid cellularity” has a clear influence during the first four years. After that, the slope of the cumulative regression function remains almost constant for the next five years and so the covariate does not influence the disease for the following years. After the ninth year, the influence seems to disappear.

The three plots of the regression functions will be put together so as to conclude to summarized results for the influence of the three covariates. This will be done only for the first four years, because as it was observed from the previous plots, the covariates had a clear influence only during the first years. While time passes the covariates started to loose their influence and that is why there is no meaning in examining the rest of the years.

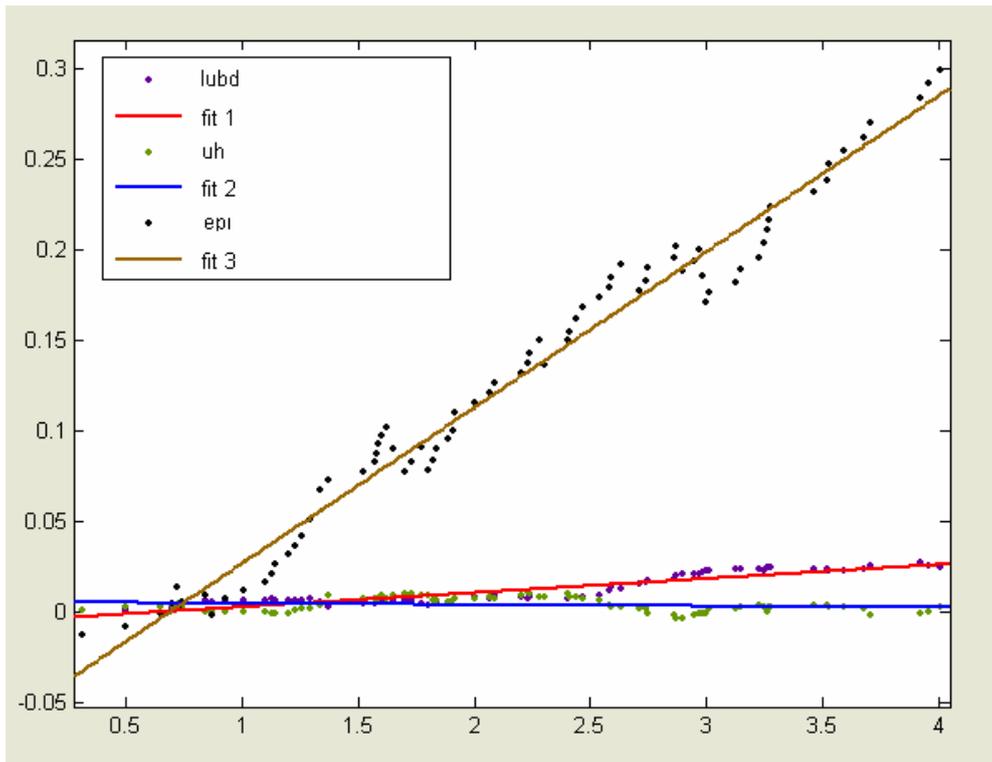


Figure 38: This diagram gives information about which covariate influences more the disease in the ciliary body

The black plot represents the cumulative regression function for the “epi” covariate. The green plot corresponds to the “uh” covariate and finally the mauve one corresponds to the “lubd” covariate. The fit was done again to linear functions so as to be able to find the slope.

From the above plot it is clear that the covariate “epi” influences more the survival, since it has the bigger and the steepest slope. After the “epi” covariate, the “lubd” covariate influences more the survival and finally the “uh” covariate influences the survival but in a less profound way. The results are the same as those for the population of the 743 patients.

The covariate “epithelioid cellularity” is the covariate that influences more the evolution of the disease at the subgroup ciliary body. Since this is by far the most important covariate, we want to see how this covariate influences the survivability of the patients. In particular we want to see which patients have more probabilities of survival, those that have epithelioid cells or those that do not have. Let us remember that the covariate “epithelioid cellularity” is a categorical one and it describes the presence or the absence of epithelioid melanoma cells.

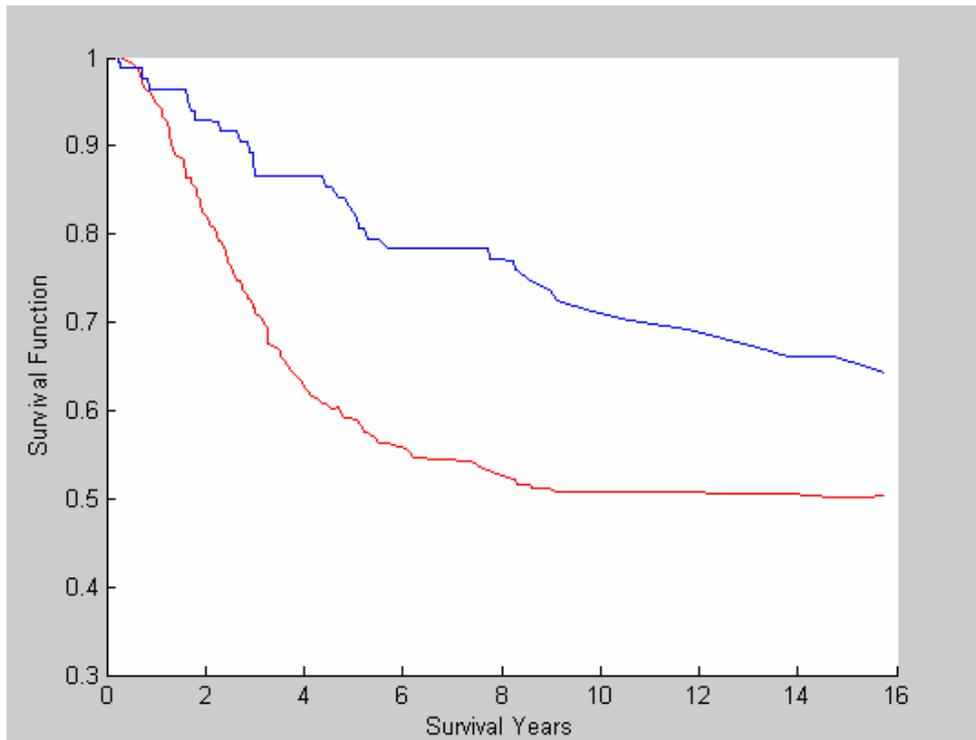


Figure 39: This figure contains the survival functions that show the survivability of the patients under the presence or absence of epithelioid cells for the subgroup ciliary body

The blue line corresponds to the survival function for those patients that do not have epithelioid cells. The value of the covariate “epithelioid cellularity” is 0 when there is absence of epithelioid cells and when there is presence of these cells the value of the covariate is 1. It becomes clear that those patients that do not have epithelioid melanoma cells have more probabilities of survival than those that have these cells. This is logical since these cells have the ability to accelerate the evolution of the disease and as a result those patients that are characterized by the presence of these melanoma cells have less probabilities of survival.

An analogous plot that shows how the survivability is affected from the absence or presence of the epithelioid cells for the entire population of the 743 patients is going to be given in a following section. In that section we also have plots that show how the survivability of the entire population of the patients is affected by the other covariates that are included in the analysis.

After this point the same procedure is going to be followed in order to verify if the covariate “epithelioid cellularity” has the greatest influence in the subgroup choroidal.

5.4.1.2 Choroidal

a. Covariate “lubd”

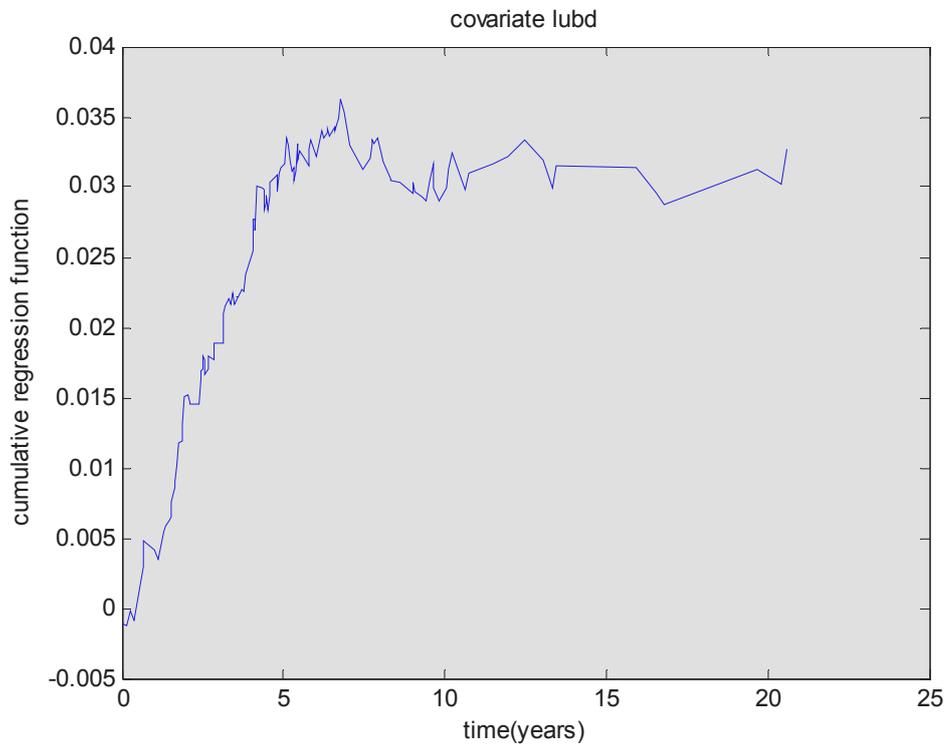


Figure 40: The plot of the regression function that corresponds to the longest ultrasound basal dimension covariate for the choroidal

The covariate “Longest ultrasound basal dimension” has a clear influence in the survival during the first eight years, but after that time the influence seems to disappear since the slope of the cumulative regression function is almost constant.

b. Covariate “uh”

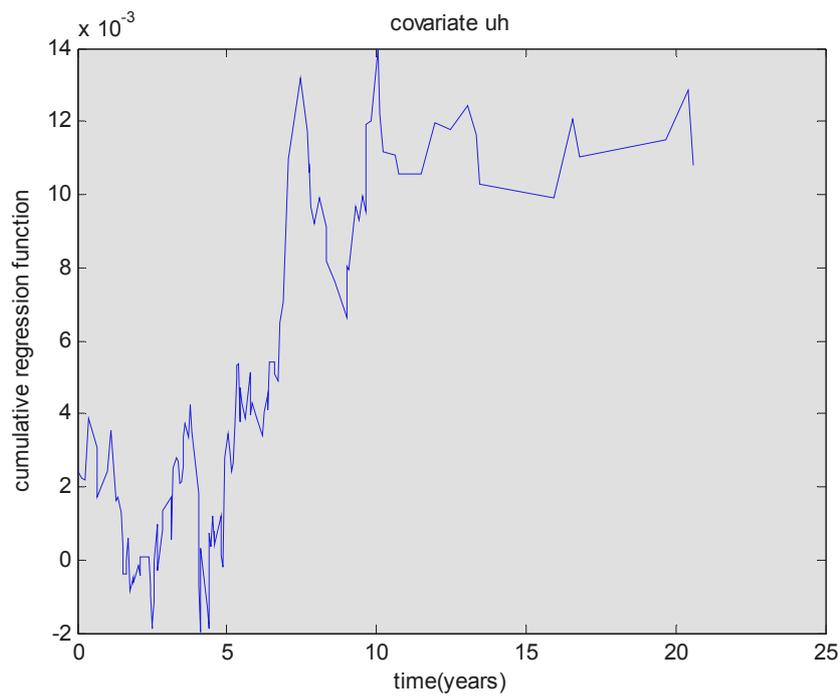


Figure 41: The plot of the regression function that corresponds to the ultrasound height covariate for the choroidal

This plot of the covariate “uh” can not lead to clear and explicit results. The slope is not clearly decreasing or increasing throughout the time period but is constantly changing. Once again, no specific results can come up from the interpretation of the above plot for the influence of the covariate “ultrasound height” to the evolution of the disease.

c. Covariate “epi”

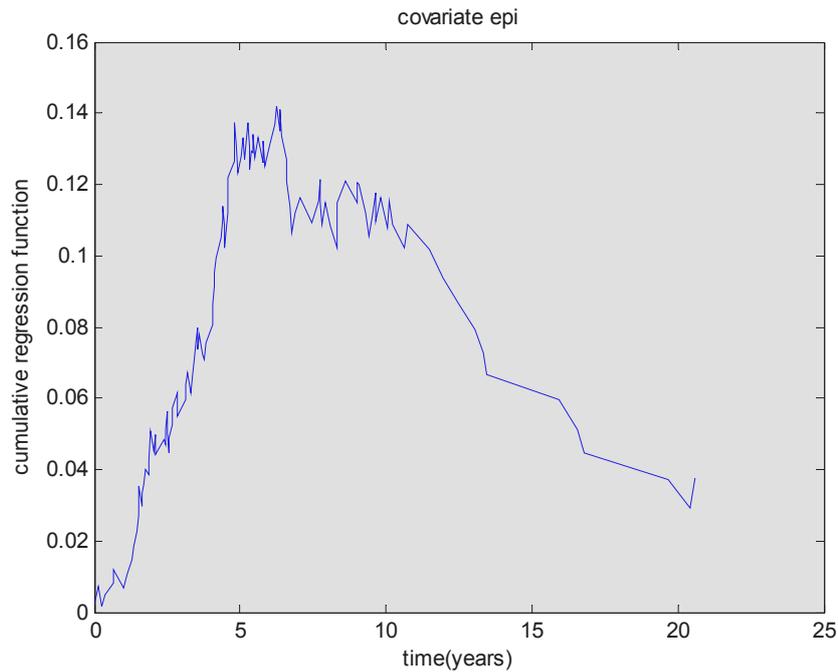


Figure 42: The plot of the regression function that corresponds to the epithelioid cellularity for the ciliary

The covariate “Epithelioid cellularity” has a clear influence during the first seven years. After that, the slope of the cumulative regression function remains almost constant for the next four years and so the covariate does not influence the disease for the following years. After the tenth year, the influence seems to disappear.

The three plots of the regression functions will be put together so as to conclude to summarized results for the influence of the three covariates. This will be done only for the first five years, because as it was observed from the previous plots, the covariates had a clear influence only during the first years. While time passes the covariates started to loose their influence and that is why there is no meaning in examining the rest of the years.

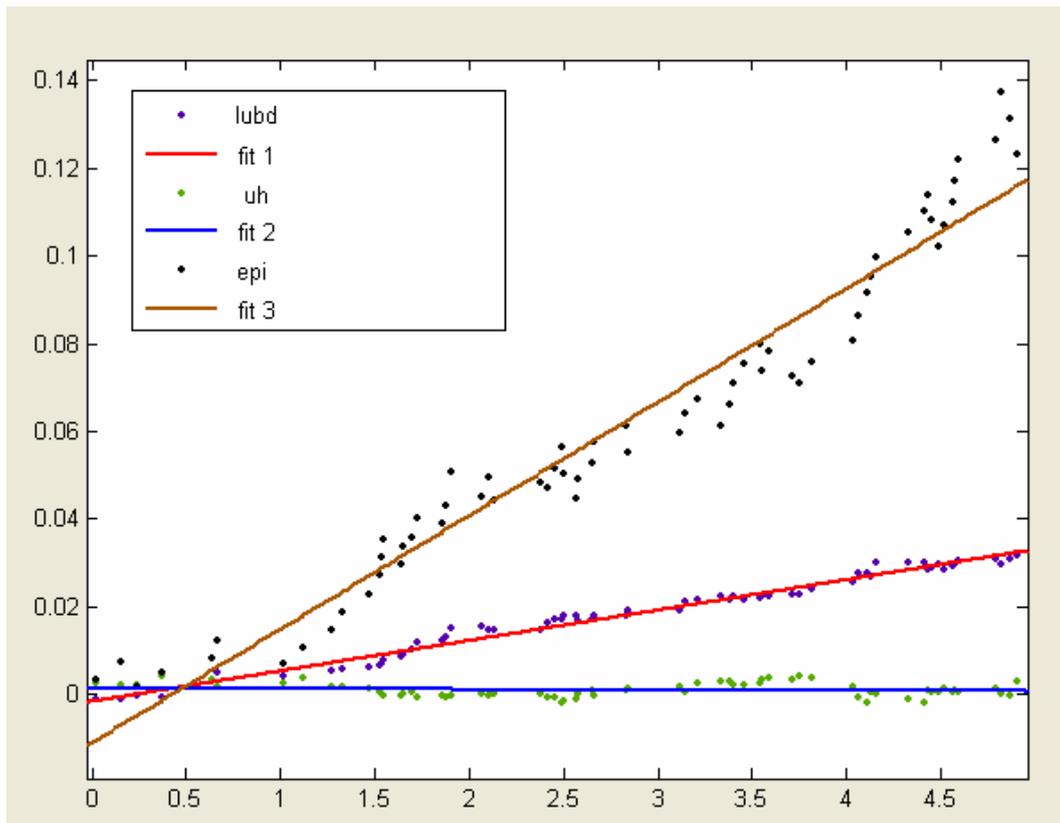


Figure 43: This diagram gives information about which covariate influences more the disease in the choroidal

Once again the black plot represents the cumulative regression function for the “epi” covariate. The green plot corresponds to the “uh” covariate and finally the mauve one corresponds to the “lubd” covariate. The fit was done again to linear functions so as to be able to find the slope.

From the above plot it is clear that the covariate “epi” influences more the survival. After the “epi” covariate, the “lubd” covariate influences more the survival and finally the “uh” covariate influences the survival but in a less profound way.

Since once again the covariate “epithelioid cellularity” is the one that influences more the survivability of the patients, we will plot the survival functions that show how the patients are affected by the presence or absence of the epithelioid melanoma cells.

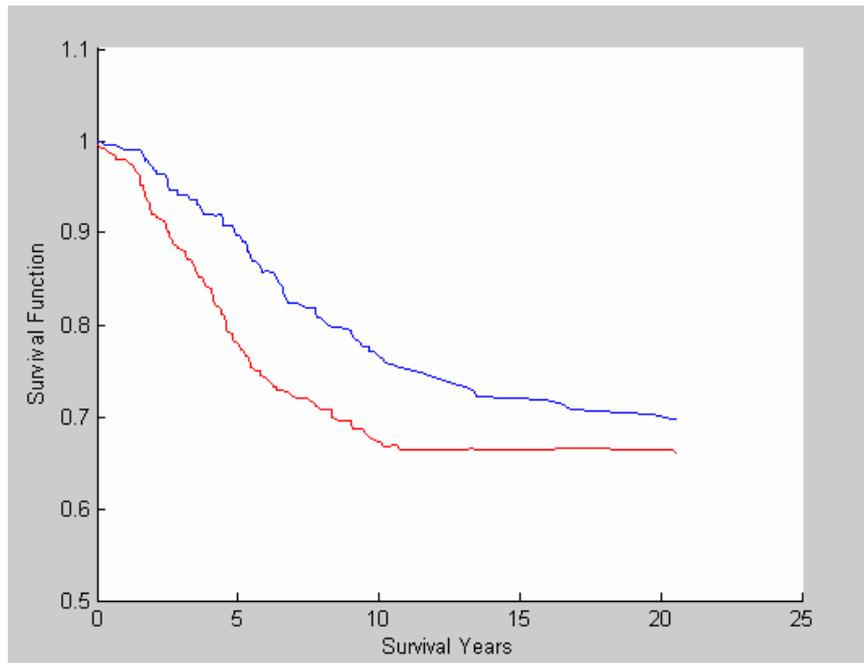


Figure 44: This figure contains the survival functions that show the survivability of the patients under the presence or absence of epithelioid cells for the subgroup choroidal

The above diagram shows that the patients that are characterized by the presence of the epithelioid melanoma cells have less probabilities of survival than those that do not have this type of cells. Obviously, the blue line corresponds to the survival function that shows the survivability under the presence of the epithelioid melanoma cells and the red line corresponds to the survival function that shows the survivability of the patients that do not have this type of melanoma cells.

As a final result it could be said that the covariates influence the survival in the same way either it is for a subgroup of the population, either it is for the whole population. The covariates “epi” and “lubd” have a clear influence during the first years and after some years this influence drops. The covariate “uh” does not have a clear influence meaning that it is a not a very helpful covariate since its interpretation does not lead to results that can be used in an effective way.

The influence of the covariates, for the subgroups of the “ciliary body”, “choroidal” and for the population of the 743 patients, in a descending way is summarized below. As is can be seen the influence is the same for all the groups, meaning that there is no difference for the covariates and their effect.

Rank of covariate in influence	Covariate name
1	Epithelioid cellularity
2	Longest ultrasound basal dimension
3	ultrasound height

Table 25: Contains the rank of the covariate in influence

5.4.2 Survival function

As it was mentioned before, the survival function as well as the cumulative hazard function can be estimated and plotted for each individual separately. What changes during the computation is the set of the covariates which is different for each individual. Below it is given a plot of survival functions that correspond to fifteen different patients. These functions are plotted together so as to clarify in a more clear way the probability of survival.

The survival functions were chosen to be plotted because they give more information about the survivability of the patients than the cumulative hazard functions. From the survival functions, it becomes really clear what chances have someone to survive, which is something that it can not be derived from the cumulative hazard functions.

The individuals were chosen randomly from all the data set. This could be done because there is independency between the individuals. The code for the survival function is given to Appendix B.

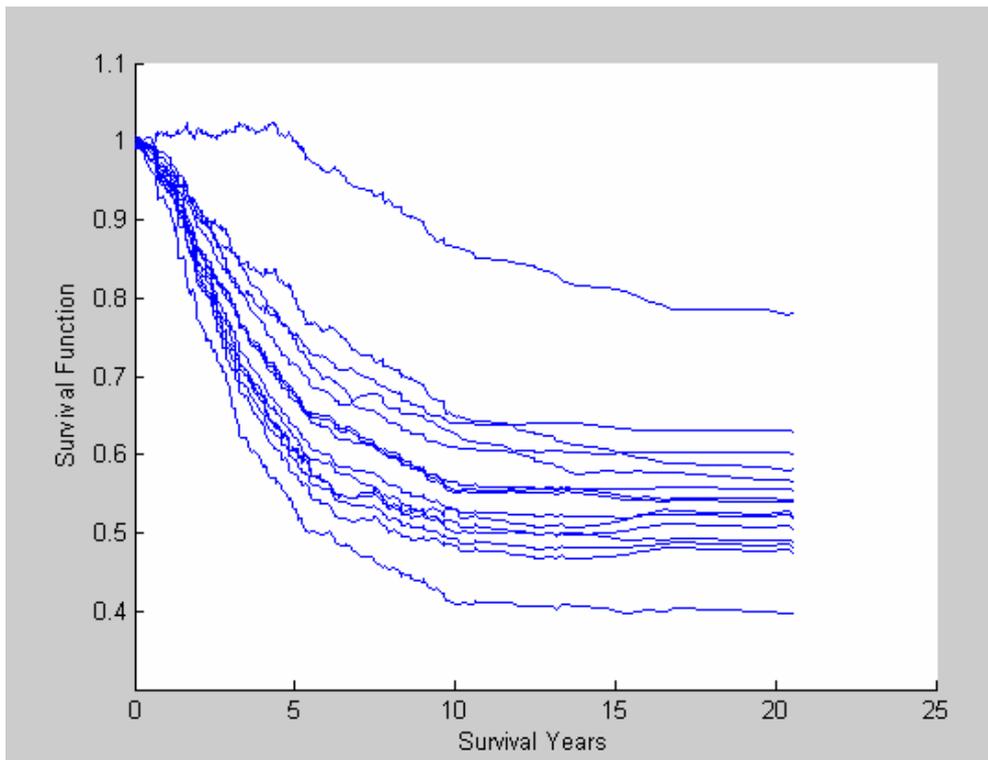


Figure 45: This diagram contains the survival functions of fifteen random patients from the additive model

As it can be seen from the above plot, the majority of the patients have a probability of survival 50% - 60%, after 10 years. Only two patients do not belong to this range. The one of them has a probability of survival 40% and the other one 80%. This is logical because in a large set of patients some may differ from the normal distribution. These differences happen because of the variety that exists in the set of covariates. In general, it can be said that the patients of this study have a satisfying probability of survival since most of them are about to live with probability 50%. This can also be proved from the diagram that follows. In this diagram, it was plotted the survival function for the entire population of the patients. This was done by taking the mean value of the three covariates. In this way the survival function that follows is representative for all the patients and verifies the above results, where we concluded that the patients of this study have a probability of survival at about 50%. In the following plot this is shown clearly, since the survival of the patients after 10 years is almost 50%.

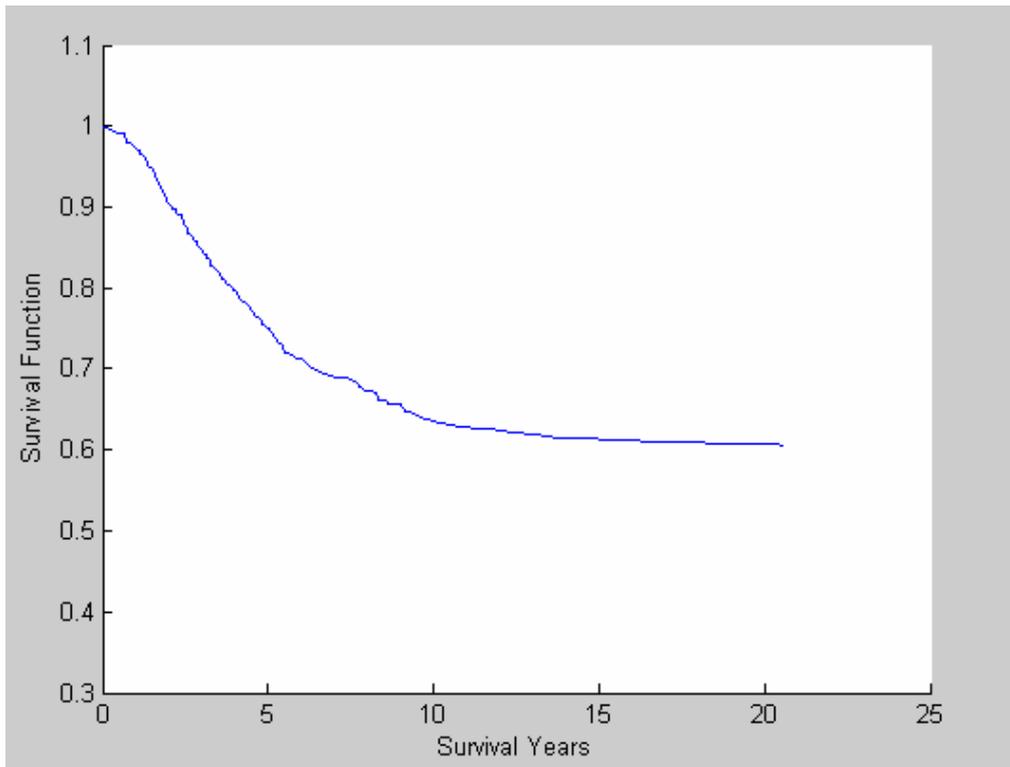


Figure 46: The survival function of the entire population of the patients of the additive model

5.4.2.1 Survivability versus Covariates

In order to have more thorough results about the influence of the covariates in the survival function, the three covariates were spited into two categories:

- “Epithelioid cellularity” covariate: was split into epi 0 and epi 1
- “Longest ultrasound basal dimension” covariate: was split into two categories of values, those that have value more than 14.9, which is the mean value, and those that have value less than 14.9
- “ultrasound height” covariate: was also split into two categories of values, those that have value more than 10, which is the mean value, and those that have value less than 10.

This was done so as to be able to come up with results that will explain which one of the covariates influences the survivability in a more effective way. In order to conclude to these results some changes had to be done in the analysis. The most important was that the dataset had to be split into the above categories. This was done with Matlab and the code is presented in the Appendix B.

The first plot shows how the survival functions are distributed for the covariate epi.

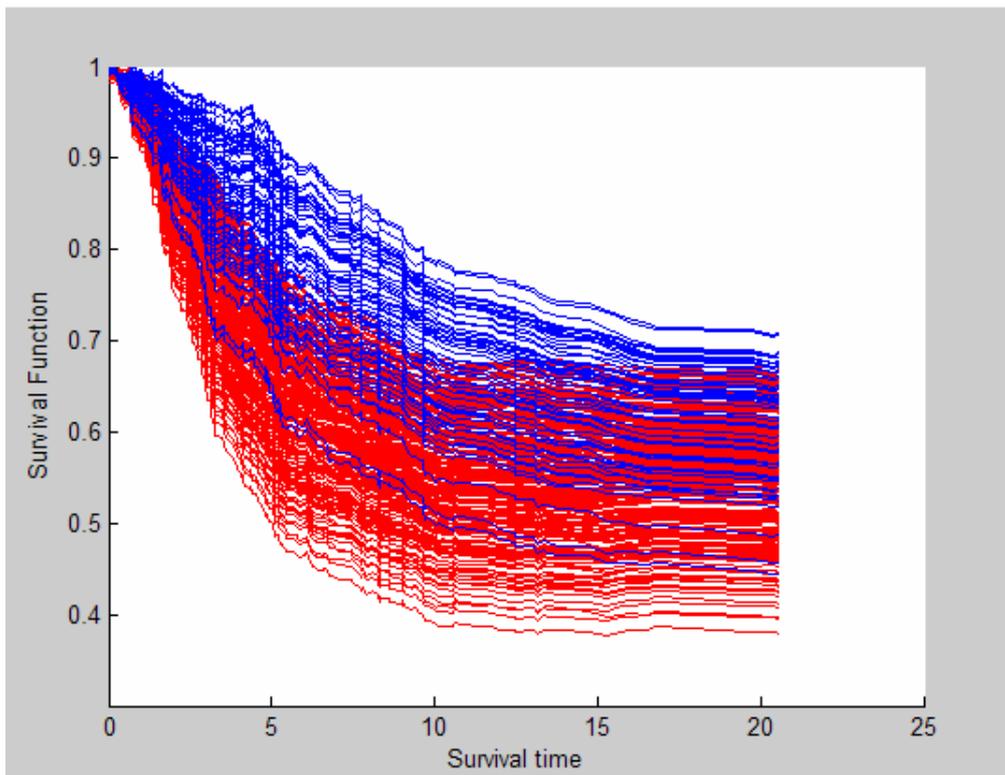


Figure 47: Shows how the survival functions are distributed for the covariate epi

With the blue colour are the survival functions for epi 0 and with the red colour are the survival functions for epi 1. As it was explained before, the covariate epi affects the disease and with this plot it can be proven that those patients that have epithelioid melanoma cells, epi 1, have more probabilities to die than those that do not have epithelioid melanoma cells, epi 0. If we take for example the twentieth year, the group of patients that have epi 1 has a mean probability of survival 55% while the group of patients that have epi 0 has a mean probability of 60%.

These results can also be verified with the plot that follows. This plot shows the distribution of the survival of the entire population. This was done by taking the mean value of the covariate epi 1 and epi 0. As it can be shown, the patients with covariate epi 1, the red line, have survivability about 55%, while the patients with epi 0 have greater survivability which is about 60%. The plot that follows is analogous to the figures 39 and 44 that show how the presence or the absence of the epithelioid melanoma cells influence the survivability of the patients for the subgroups ciliary body and choroidal. The figure below shows the results for the entire population of the patients.

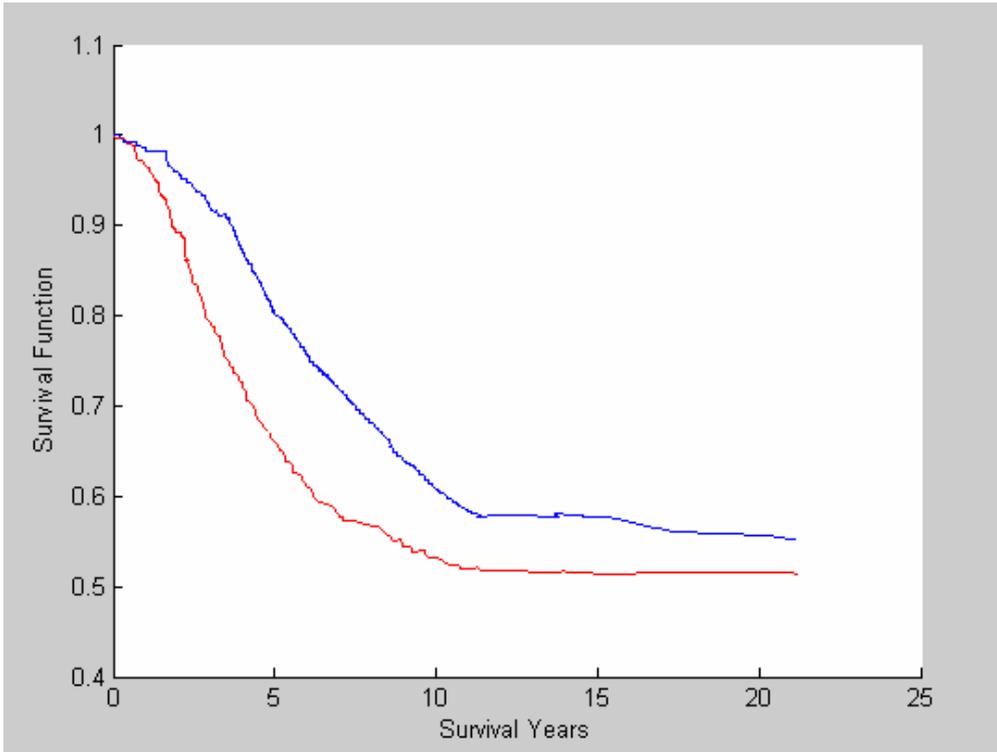


Figure 48: The survival function for the entire population with the mean value of epi 1 and epi 0

The second plot shows the distribution of the survival functions for the covariate “longest ultrasound basal dimension”

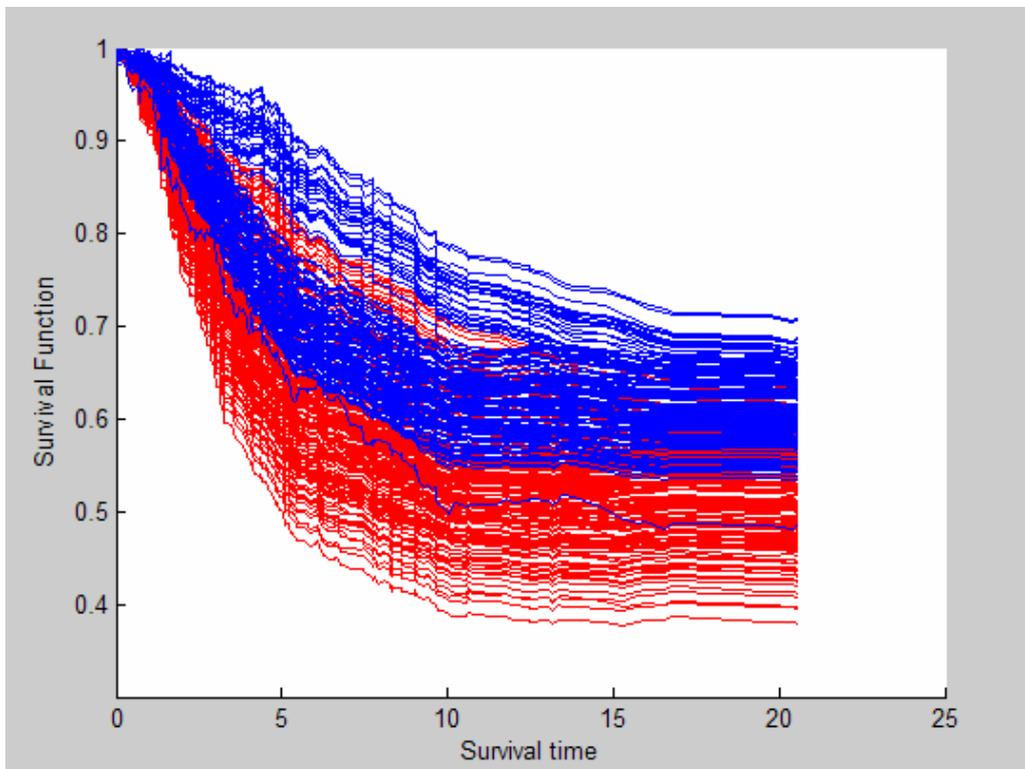


Figure 49: Shows the distribution of the survival functions for the covariate “longest ultrasound basal dimension”

With the blue colour are the survivals that correspond to lubd small, less than 14.9, and with the red colour are the survivals that correspond to lubd dig, more than 14.9. The covariate longest ultrasound basal dimension defines the longest basal tumour dimension as measured by ultrasonography and as it is shown in the above plot, those patients that have small lubd have more probabilities of survival. Let's take again for example the twentieth year of survival. The group of patients that have big lubd has a mean probability of survival 53% while the group of patients that have small lubd has a mean probability of 60%.

The third plot shows the distribution of the survival functions for the covariate “ultrasound height”

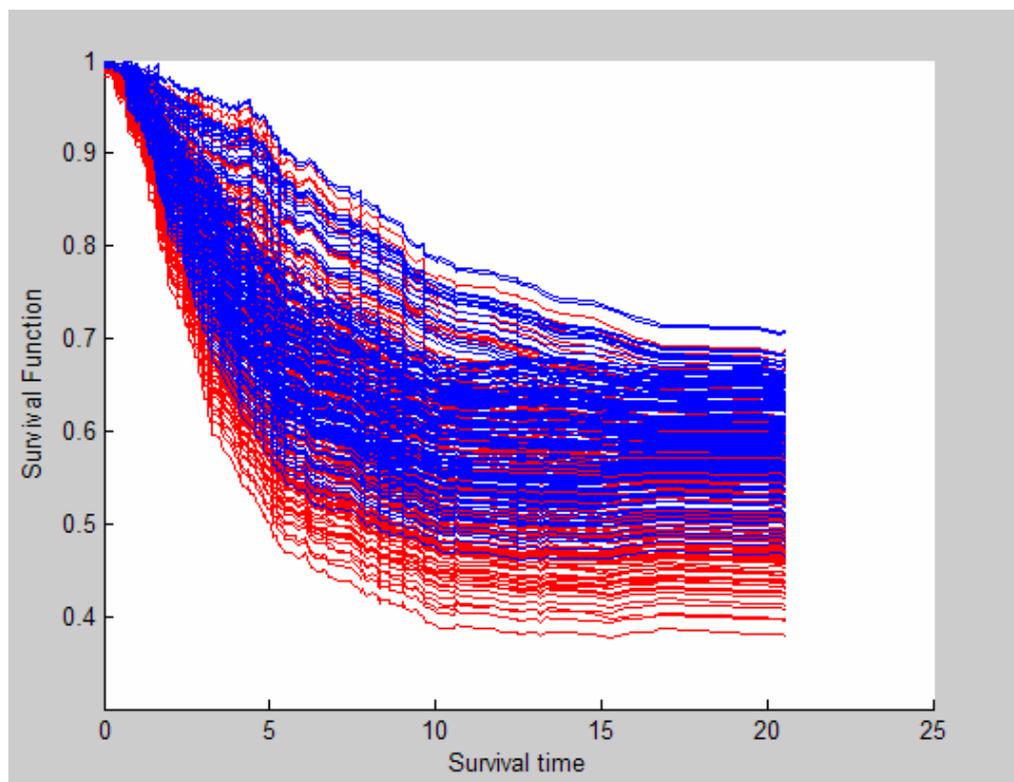


Figure 50: Shows the distribution of the survival functions for the covariate “ultrasound height”

From this plot it becomes really clear the fact that the covariate “ultrasound height” does not influence the disease in a clear way, since the survival functions are overlapped and there is no clear separation so as to come up with effective conclusions. That conclusion was also proven from the analysis of the additive model.

In the final plot, it is shown the distribution of the survival functions for big lubd - epi 1 and for big lubd - epi 0.

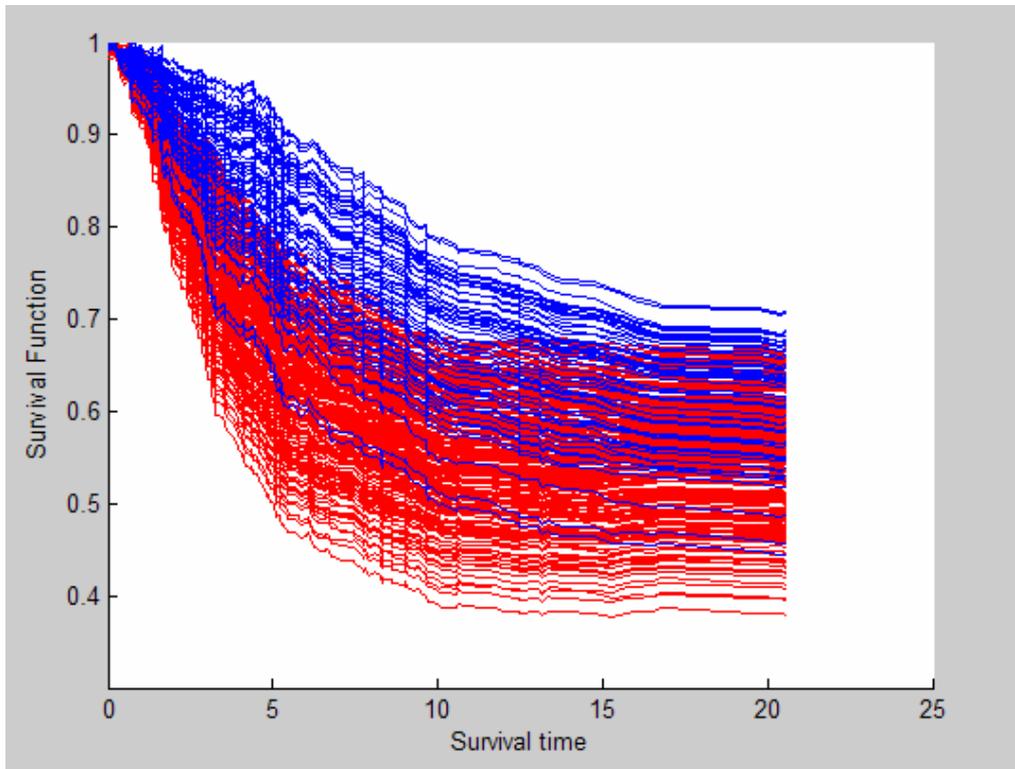


Figure 51: Shows the distribution of the survival functions for big lubd - epi 1 and for big lubd - epi 0.

From this plot it becomes clear that the covariate “Epithelioid cellularity” influences more the disease than the covariate “Longest ultrasound basal dimension”. This is due to the fact that the patients even if they have lubd big, which can be translated to more probabilities of death, when they have epi 0 then they have more probabilities of survival. This is shown clearly from the above plot, where the blue line represents those that have epi 0 and big lubd and the red line represents those patients that have epi 1 and big lubd. The first group has survivals that represent bigger probabilities of survival, while the second group has survivals that represent smaller probabilities of survival.

All the codes for the above plots are given to Appendix B.

5.5 Results from the neural networks

As it was mentioned before, in this study we did not use any of the neural networks that were presented in chapter 1, which make a separate survival analysis but we used neural networks in order to train them with an already used survival method.

In order to train the neural network in a more effective way, the data set of the 1743 patients was split into two subsets. Each subset had to be of equal size and for that reason the original data set was divided in a way that gives two sets with the same deviations. This means that the two new sets contained both small and big values so as to have better results, since the two sets contained equally spaced points from the original data. If the data set was simply divided by two then the one set would contain big values and the other one would contain only small values and so the training would not be effective. If someone wanted to put new values to be trained then the results would not be precise.

Each one of these two subsets was used for a different reason. The one was used for training and the other for simulation and that is why the results were very satisfying, as it will be explained below.

While training a neural network, some stopping criterions must be achieved. In this neural network, the stopping condition was the goal parameter to reach zero. When the neural network stopped its training, that goal was not exactly achieved, which was expected since it is rather difficult to achieve the desired value of goal. However, the error, the MSE, was reduced as it seems from the following training results, which came up after the training procedure of the neural network. The code is given to Appendix A, since it concerns the Kaplan- Meier method

```
TRAINLM, Epoch 1875/2000, MSE 1.5029e-006/0, Gradient 0.094557/1e-010
TRAINLM, Epoch 1900/2000, MSE 1.49401e-006/0, Gradient 0.118256/1e-010
TRAINLM, Epoch 1925/2000, MSE 1.48552e-006/0, Gradient 0.139427/1e-010
TRAINLM, Epoch 1950/2000, MSE 1.47769e-006/0, Gradient 0.120236/1e-010
TRAINLM, Epoch 1975/2000, MSE 1.47148e-006/0, Gradient 0.0676012/1e-010
TRAINLM, Epoch 2000/2000, MSE 1.46685e-006/0, Gradient 0.0313849/1e-010
TRAINLM, Maximum epoch reached, performance goal was not met.
```

Figure 52: Shows the last epochs of the training and the error

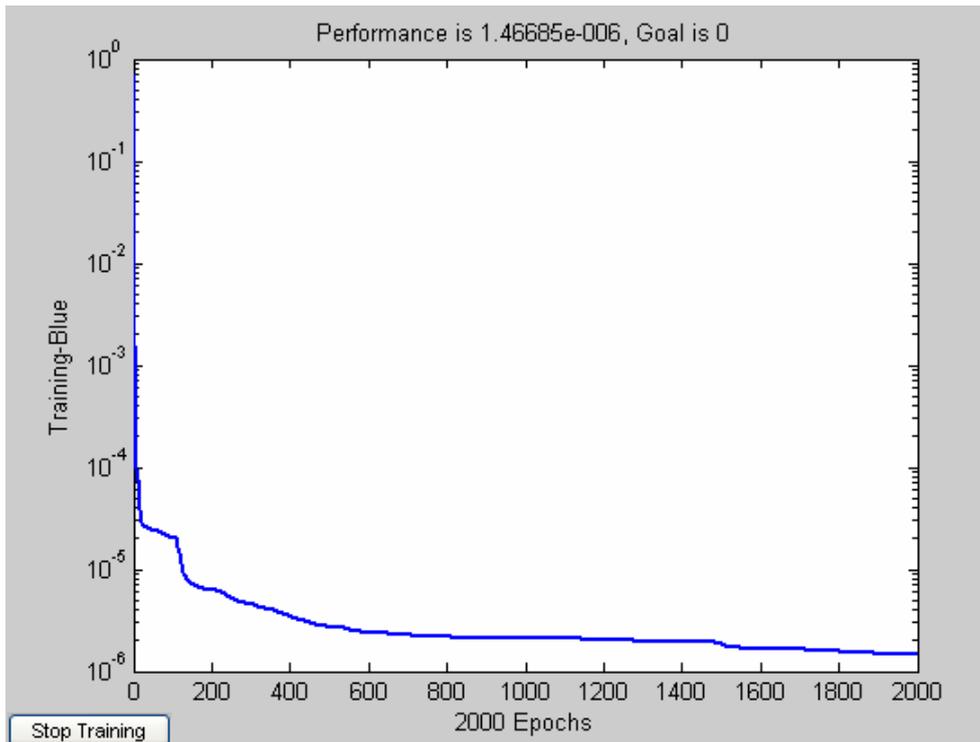


Figure 53: This diagram shows the epochs of the training and the final performance

As it can be seen, while the epochs proceed, the error is diminishing and the final value is 1.46685×10^{-006} which is close enough to the zero value of the goal.

The results that finally came up after the simulation show that the output of the neural network and the target values are very close. This means that the neural network was trained in a very effective way. The neural network was trained for 2000 epochs which are enough since the results was very satisfying.

The results are shown in the diagram that follows. Those results came up from the regression analysis between the network response and the corresponding targets. As it was mentioned before in chapter 4, the performance of a trained network must be checked. One way to do this is by performing a regression analysis between the network response and the corresponding targets, which is something that was done in this study. This regression analysis, as we said in chapter 4, is implemented with the function *postreg*.

The most important parameter of those that the *postreg* function returns is the correlation coefficient between the outputs and targets, or else the R-value. In the following diagram, the R-value is shown which is actually a measure of how well the variation of the output is explained by the targets. If this number is equal to 1, then there is perfect correlation between targets and outputs.

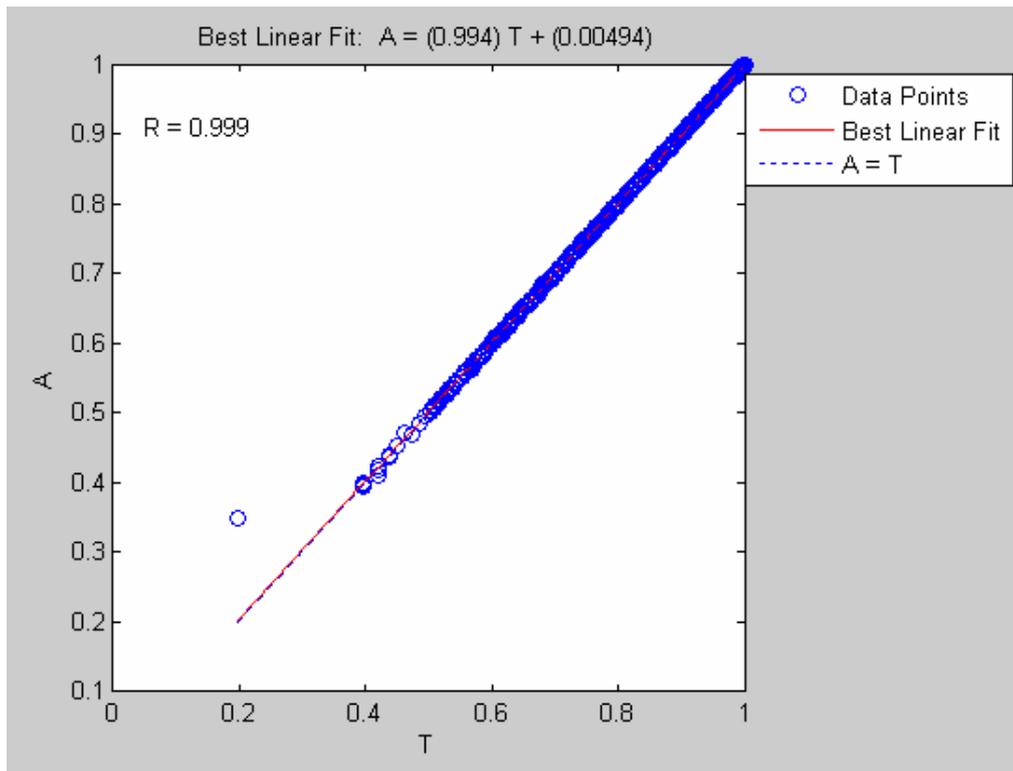


Figure 54: Shows the correlation coefficient, which is one of the results of the training test

Since the correlation coefficient is 0.999, this means that between the target and the output there is a perfect correlation. This can also be shown from the above diagram, where the best linear fit (the data points) approximates the linear line $A=T$ ($y=x$). This means that the target correspond the output, since they both have a range from 0-1. As a result, the output of the neural network and the output from the Kaplan-Meier method are very much alike, as it is shown in the below table.

In order to have a clearer image, they are given some values from the output of the neural network and the Kaplan-Meier method so as to be compared.

Number of element	Output of Kaplan-Meier	Output of neural network
6	0,99942	0,99986
7	0,99942	0,99954
8	0,99942	0,99913
9	0,99884	0,99875
10	0,99826	0,99833
11	0,99709	0,99774
12	0,99709	0,99728
13	0,9965	0,99708
14	0,99592	0,99631
15	0,99592	0,99591
16	0,99533	0,99519
17	0,99414	0,99495
18	0,99355	0,99452
19	0,99296	0,9939

Table 26: Contains some of the values of the outputs of the neural network and the Kaplan-Meier method

As it can be seen, those two outputs are very close and that is why the correlation is so good.

Chapter 6

Comparisons of the results from this study and from other studies

In this last chapter, comparisons will be made. These comparisons are based to the results of this study and also to the results of other studies. These other studies, either make use of the same data set of the ocular melanoma disease, either make use of different data set but of the same disease. All the results of these studies are going to be presented below.

6.1 Comparison of the results of all the methods of this study

In the first part of this chapter, a comparison of the results of the methods that we used in this study will be made. Let us remember which methods we used in this study. From the nonparametric models, the Kaplan-Meier method, the Life-table and the Nelson-Aalen method were used. From the semiparametric models, the Aalen's additive model was used and finally we also used neural networks in order to train a neural network with a known statistical method. This method was the Kaplan-Meier method.

The main scope of the Kaplan-Meier method is to compute the survival function or else the survival curve; this is why another name of this method is the "survival curve" method. Of course, the cumulative survival function can also be derived from this method with the relationship $H(t) = -\log S(t)$.

The main scope of the Nelson-Aalen method is to derive the cumulative hazard function. By having the cumulative hazard function, the survival function can also be derived by the relationship $S(t) = \exp[-H(t)]$. These two methods, the Kaplan-Meier and the Nelson-Aalen are two methods that practically compute the same things but with an inverse way.

The Life-Table method is like the Kaplan-Meier method, with the difference that the former uses time intervals in the procedure of the analysis, while the latter uses all the time points. The Life-table concludes to survival function, to density function and to hazard function.

The Aalen's additive model has as main scope to conclude to plots that give information about the effect of the covariates to the disease. Also individual survival functions can be derived and so results about the survivability of each patient can come up.

Finally the neural network that we used in this study does not count as a separate survival model, since we used the Kaplan-Meier estimator to train it. The results of the neural network will also be compared with the rest results.

Let us first compare the survival functions of the three nonparametric models. The figures that correspond to the survival functions of the Kaplan-Meier method, the Life-table method and the Nelson-Aalen method are the figures 10a, 18a and 28 respectively. If we compare these three figures, we can see that the survival function is almost the same. The survival function of the Kaplan-Meier and the Nelson-Aalen is exactly the same, since these two methods give the same results but in an inverse way. The survival curve of the Life-table differs a little bit, but this is logical since this method uses time intervals. What is important is that these three methods give the same probabilities of survival to the patients. All of them give to the patients a probability of about 20% to be cured. But if we do not take into account the last death, which may be thought as statistical error, as we already said, then the probabilities of the patients to be cured are augmented at about 40%. In addition, the median survival is similar to the three methods, since it varies from 18 years to 19,6 years, which verifies that the three above methods have almost the same survival distribution. The median survival of these three methods is given in the figures 11, 19 and 29.

As far as the cumulative hazard functions are concerned, we can say that they are the same in all the three methods. This is logical, since the cumulative hazard function can be derived from the relationship $H(t) = -\log S(t)$. This relationship uses the survival function, and since the survival function is the same in the three nonparametric methods that we use, the cumulative hazard function is also similar in these methods. The cumulative hazard functions are given in the figures 12a, 25 and 27a.

Lastly, the Life-table method gives density function and hazard function that no other of the two other methods gives. None of the two other methods give any of the functions that we just mentioned, and that is why no comparison can be done.

The results of the Aalen's additive model cannot be compared with the results of the three above methods. This happens due to two main reasons. First because the additive model is a semiparametric model and uses covariates while the other methods are nonparametric. Because of this fact the results are different since

the additive model has as main scope to derive conclusions about the effect of the covariates. Second, because in the additive model we used 743 patients, because of the missing values of the covariates, and not 1743 as in the nonparametric models. Due to this fact the survival functions that we derived from the additive model cannot be compared with the survival functions of the nonparametric models. The results of the additive model will be compared with results from other semiparametric models.

Finally, as we already mentioned in the previous chapter, the results of the neural network are almost the same with the results of the Kaplan-Meier method. This means that the network was trained in an effective way since the output of the neural network has almost the same values with the Kaplan-Meier estimator. These results are shown in table 26.

6.2 Comparisons with relevant studies

6.2.1 Comparison with studies that use the same dataset

6.2.1.1 Comparison with study that uses the Cox model and the Kaplan-Meier method

At this point the results from our study are going to be compared with the results from another study that uses the same dataset of the ocular melanoma disease. Both datasets have 1743 patients. The study that is going to be compared was done by Fotini Agrafioti and uses the Kaplan-Meier method, The Cox-model and log-logistic regression model. [oth-7]. The results from the Kaplan-Meier and the Cox model will only be compared since the log-logistic regression model was not able to be implemented in the dataset of the ocular melanoma disease, and so no results came out.

The study of Fotini Agrafioti's uses the Kaplan-Meier method and the Cox model for a reduced dataset. The dataset was reduced because the Cox model uses covariates and so the data that contained missing values had to be removed. As a result the new subset contained 743 elements. In addition, this reduced dataset was also used as a basis for the Kaplan-Meier estimator of the study just mentioned. We must mention here that in our study we also had to reduce the dataset for the Aalen's additive model which uses covariates, as described in chapter 5.

Having the results from the study of Fotini Agrafioti we can verify if any piece of information was lost due to the elimination of the dataset. We may have not eliminated the dataset for the Kaplan-Meier method but we did it for the Aalen's additive model and that is why we want to verify if this elimination has not biased our data. This can be done by comparing the Kaplan-Meier of our study, with the 1743 patients, with the Kaplan-Meier of the other study, which contains 743 patients. At this point we must also say that the censoring of patients was done in both studies with the same way, which is logical since the same data set is used. This means practically that a patient is considered to be censored if the death indicator is zero.

The plot below shows the survival curves of the two datasets.

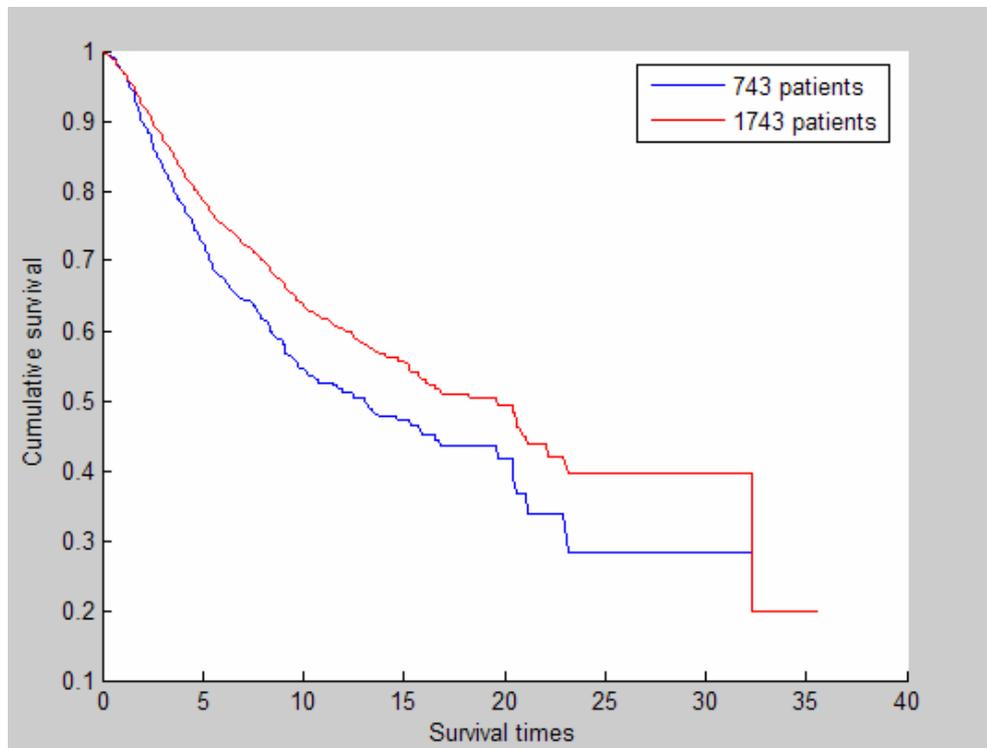


Figure 55: Kaplan- Meier estimator for the dataset of ocular melanoma for the 1743 patients and the 743 patients

As it can be seen, the two survival curves are relatively close to each other. Let us remember the possible statistical error that the last patient of the 1743 patients induces. If the last death is ignored, then the survival curve will not take a step down and so the survivability will reach the 40%. As it can be seen, the survival curve of the 743 patients goes up to 30%. This means that their difference is only 10%, which

may be a difference, but is a small one. In addition, these two survival curves seem to have the same shape.

After comparing the nonparametric models of the two studies, ours and Agrafioti's one, the semiparametric models will be compared. The Cox proportional hazards model and the Aalen's additive model both include the covariates in a way that they make conclusions about their effectiveness on the survivability. This is why these two methods can be compared. Before comparing the results, a short presentation of the Cox model will be made.

The Cox regression model was first introduced in 1972 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. [oth-7]

The covariates are assumed constant in time. 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:

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

The only comparison between the two studies that can make sense is the one referring to the effect of the covariates. In both studies this effectiveness is measured with a regression procedure. In the Cox proportional hazards model, conditional 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:

$$\frac{h(t_i, z_i)}{\sum_{j \in R(t_i)} h(t_i, z_j)} = \frac{h_0(t) \exp(z_i^T \beta)}{\sum_{j \in R(t_i)} h_0(t) \exp(z_j^T \beta)} = \frac{\exp(z_i^T \beta)}{\sum_{j \in R(t_i)} \exp(z_j^T \beta)}$$

Assume we have k failures (deaths) then the partial likelihood function over all failure times is: $L(\beta) = \prod_{i=1}^k \frac{\exp(z_i^T \beta)}{\sum_{j \in R(t_i)} \exp(z_j^T \beta)}$.

The Log partial likelihood is then: $l(\beta) = \sum_{j=1}^k [(z_j \beta) - \log(\sum_{i \in R(t_j)} \exp(z_i \beta))]$

Cox proved that the maximum partial likelihood estimation of β is the solution of the equation: $\frac{\partial l(\beta)}{\partial \beta} = 0$. By the solution of this equation, from the results of the study of

Fotini Agrafioti's the following regression parameters were obtained:

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0.127 \\ 0.033 \\ 0.664 \end{pmatrix}$$

The first value refers to the longest ultrasound basal dimension (lubd), the second to the ultrasound height (uh) and the third to epithelioid cellularity (epi). These coefficients indicate the magnitude of the effects of their corresponding covariates. 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.

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 the analysis of the additive model. The only difference is that the regression parameters is not counted numerically but is represented graphically. The graphs are used to calculate the slope of the distribution. The greater the slope the more does the coefficient affect survivability.

The plot that follows represents the three regression parameters. The slope of the epithelioid cellularity is higher than the longest ultrasound basal dimension which is higher than the ultrasound height. This plot is the same as figure 33, but in order to have a more complete image, it is also given below.

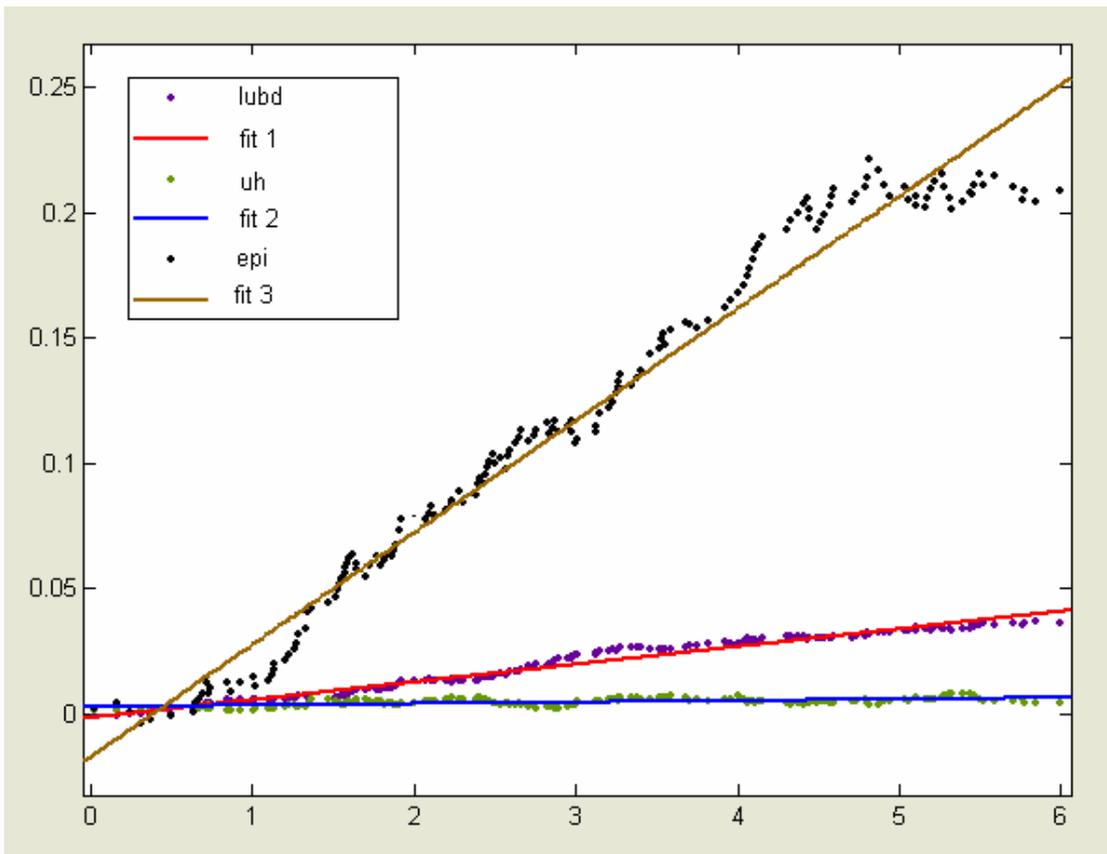


Figure 56: This diagram gives information about which covariate influences more the disease

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 we know that a patient that appears to have epithelioid cellular will have less probability to survive. This result will also be confirmed from the comparisons from other studies that follow.

6.2.1.2 Comparison with study that uses artificial neural networks and Bayes theorem.

The utility of the neural networks in survival analysis has been presented in chapter 1. Neural networks in general are very useful in survival analysis because they can implement in a very effective way the analysis that is required in order to obtain results. This can be done with various methodologies and one of them is presented below

A relevant study was performed by Taktak, Fisher and Damato [p-21]. They modeled the survival after treatment of intraocular melanoma patients. This study introduces a new idea about how the neural networks can be combined with Bayes theorem, which is different from the neural networks that we presented in chapter 1. It actually describes the development of an artificial intelligence system for survival prediction from the data of the intraocular melanoma. After the implementation of the neural network, the output results are compared to the Kaplan Meier estimator and the Cox model.

The dataset contains patients treated in Glasgow and Liverpool between 1969 and 2001. The covariates that are analyzed to check their influence on the disease are the coronal and sagittal tumor location, the anterior tumor margin, the largest basal tumor diameter and the cell type. These covariates and also the time consist the inputs of the neural network. The output of the system is the survival probability. The basic idea is to examine how the survival curve 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 made at 15 years. The entire time period is divided into five time intervals each one trying to contain the same number of events. For each time interval, a separate neural network is made which estimates the survivability at this time interval. Censored patients are only considered until the time intervals when they are active, which means that after that time they are excluded from the study. This also counts for the uncensored patients, meaning that they are also considered until the time interval when they die.

So, 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 represents high chance of survival and 1 low chance for the corresponding time interval. The data in each time interval are divided into training and test sets with a specific method in order to eliminate bias.

The main idea of this study is that the Artificial Neural Network is used to transform the output to a survival function by using the Bayes theorem. If the ANN output at a time interval is above a certain level, which is called Γ_i , this indicates a low chance of survival. An output lower than this level indicates high chance of survival and then, this record is presented to the subsequent network for the next time interval. The probability equations that use this study are 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 basic definitions are includes in the following table:

T/d_i	The number of patients who died from the tumor and had score value $\geq \Gamma_i$
t/d_i	Number of patients who died from the tumor and had score value $< \Gamma_i$
T/n_i	All patients with score $\geq \Gamma_i$
t/n_i	All patients who had score $< \Gamma_i$

Table 27: Contains the basic equations of the Taktak study

In this particular study the syndrome S is represented by a high Artificial Neural Network score, which is greater or equal to Γ_i . The probability of death given the presence and non presence of the syndrome S can be defined with the help pf Bayes theorem as it is shown from the following equations.

$$P(D/S)_i = \frac{P(D)_i \times P(S/D)_i}{P(S)_i}, \quad P(D/\bar{S})_i = \frac{P(D)_i \times P(\bar{S}/D)_i}{P(\bar{S})_i}$$

The parameters of the above theorem are computed by the following way:

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

$$P(S)_i = \frac{\text{patients with score } \geq \Gamma_i}{\text{total patients}} = \frac{T/n_i}{n_i}$$

$$P(\bar{S}/D)_i = \frac{\text{patients died with score } < \Gamma_i}{\text{total deaths}} = \frac{t/d_i}{d_i}$$

$$P(\bar{S})_i = \frac{\text{patients with score } < \Gamma_i}{\text{total patients}} = \frac{t/n_i}{n_i}$$

Finally, the survival function SF is given by:

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

The results of the Artificial Neural Network, when compared to the Kaplan Meier and the Cox model, prove that the Artificial Intelligence 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 increased the difference became lower than 20%. The Kaplan Meier survival curve of this study is shown in the figure that follows.

The results from this study cannot easily be compared to the ones from our study that are presented in chapter 5. The main reason is that the two studies may use the same datasets, but actually our dataset is a subset of Taktak's dataset. This is because the dataset of the Taktak uses more patients and more covariates. This is the main reason that comparisons cannot be made, since the results are different. This is shown from the following figure, where the survival after the ten years is about 60%, while in our study, the survivability reaches 40%, if the last death is ignored. In addition, the study of the TakTak finishes at 15 years, while our study finishes at 33 years.

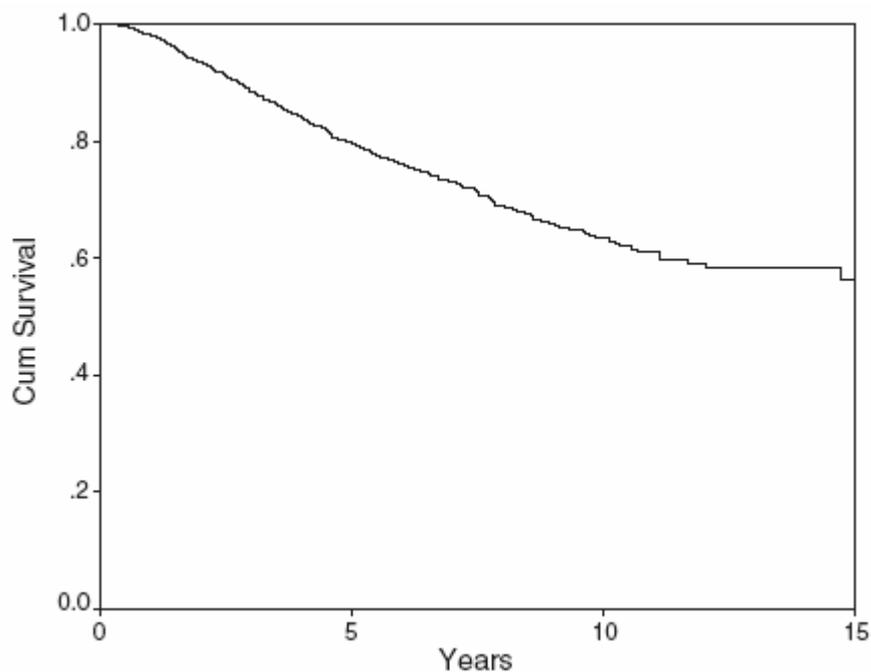


Figure 57: Kaplan Meier survival curve for 2331 patients of intraocular melanoma [p-21].

6.2.2 Comparisons with studies that use different datasets

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

A relevant study on prognostic factors is made from Isager, Ehlers and Overgaard [p-22]. The main purpose of this study is to evaluate prognostic factors for the survival of patients treated by enucleation for choroidal and ciliary body melanomas. The nonparametric Kaplan-Meier method and the semiparametric Cox model are used in order to plot the survival curves and find the influence of the 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 also for death from other causes. In melanoma survival, patients that died from other causes are considered censored. The prognostic factors are estimated by the Cox model and also by the Kaplan Meier with the log rank test. Parameters that are known to be prognostic are included in the Cox proportional hazard analysis. The covariates that are examined are the tumor location, the largest basal diameter, the epithelioid cellularity and the extrascleral extension. The cause of death was melanoma in 56% of the population and non melanoma in 44%.

The study, instead of giving the survival curve, they give instead the survival probabilities in the cumulative melanoma specific survival function, which are summarized in the following table.

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

Table 28: Contains the survival probabilities of the study of Isager

At this point, we will give again the survival curve of the Kaplan-Meier method from our study so as to compare the results.

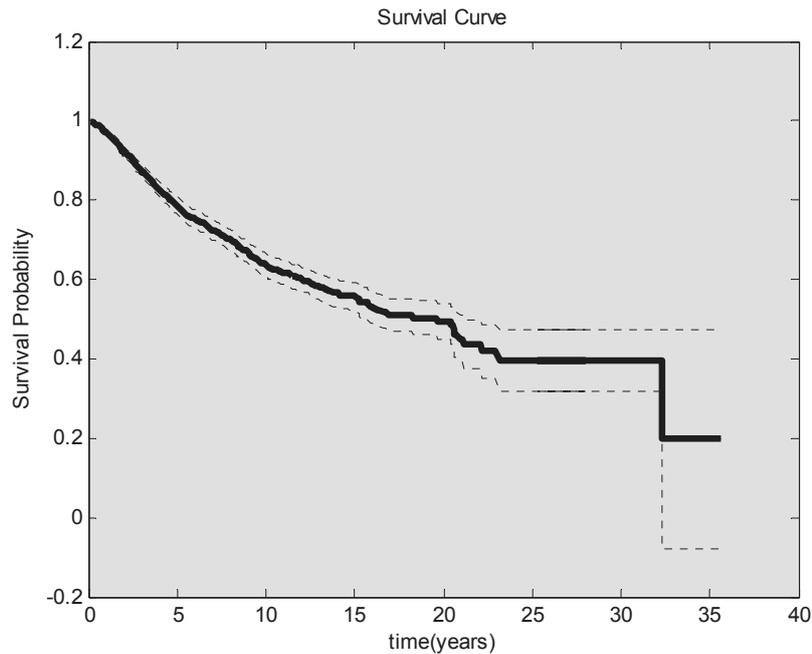


Figure 58: The Kaplan Meier survival curve for the ocular melanoma dataset of our study

The survival probabilities from the survival curve of our study are:

- 5 years → 78%
- 10 years → 60%
- 15 years → 55%
- 20 years → 43%
- 25 years → 40%

If we compare the above survival probabilities with the table 28, we will see that there are some similarities and especially at the end of each study, where the survivability reaches 40%. This means that the results from our study and from this study are close enough. Any difference in the survivability may be due to the fact that the two studies examine different dataset.

Finally we will also compare the results from the Cox model with the results from the additive model of our study. The results in the Isager's study obtained with the Cox model prove that the epithelioid cellularity is associated with the worst survivability. The survivability was found to deteriorate when the patient has epithelioid cells. As shown in the following figure, the population was divided into two groups, the first one contains patients with spindle cell type (non epithelioid) and the second one contains patients with epithelioid cells. This is according to the results

from our study, where we proved that the covariate epithelioid cellularity is the covariate that affects more the evolution of the disease. Moreover, we showed that the patients that have epithelioid cells have less probabilities of survival. The former results are shown in figure 56 which we repeated above so as to be able to have a clearer image. The later results are shown in the following figure (figure 60) which we also repeat so as to have a more concrete idea.

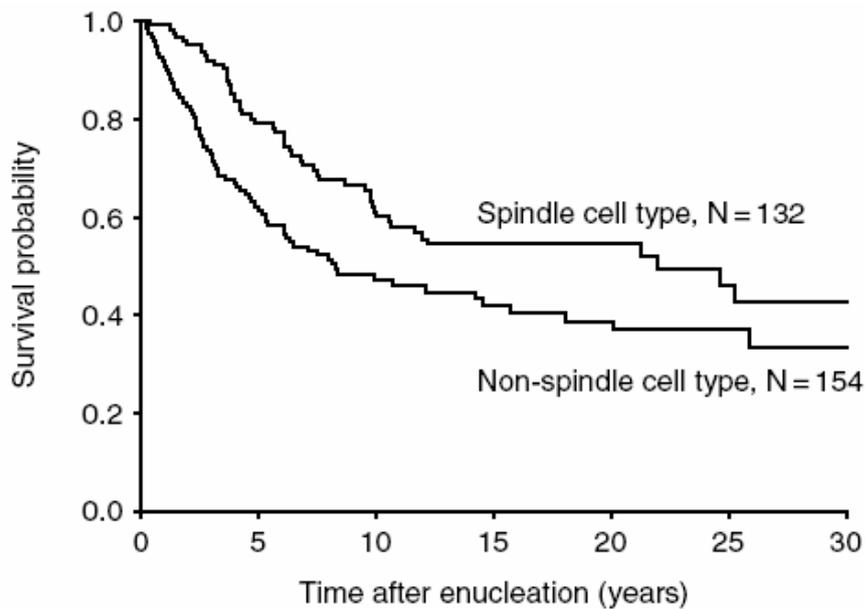


Figure 59: Survival curves for patients with epithelioid and non epithelioid cells as they derived from the Isager's study. [p-22]

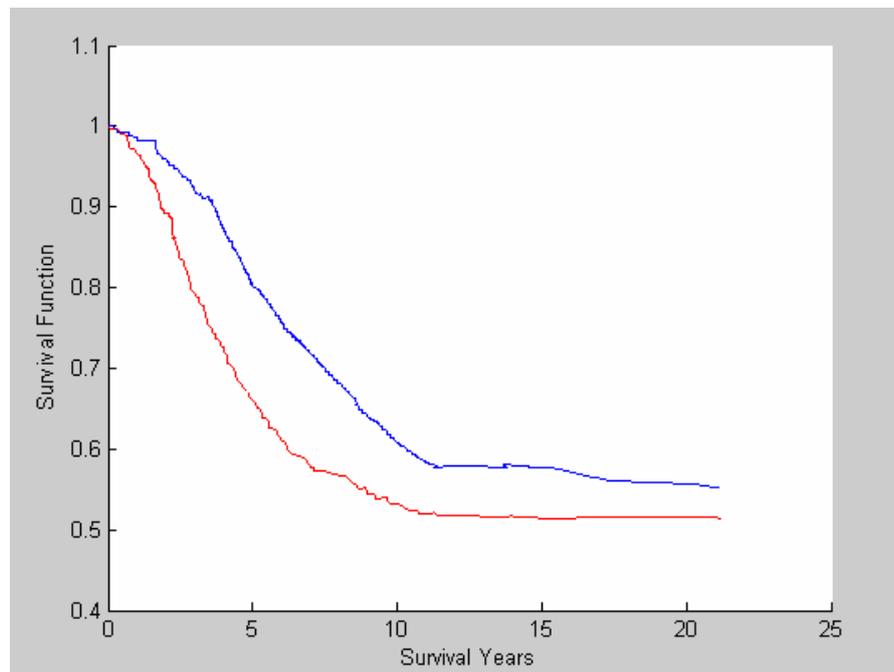


Figure 60: The survival function for the entire population with the mean value of epi 1 and epi 0.

6.2.2.2 Comparison with a study based on prognostic indicators in choroidal and ciliary body melanoma.

This study from Teemu Makitie [oth-8] is based on survival analysis of intraocular melanoma. This study was done in association with the Ophthalmology department and that is why there are more clinical and medical than statistical results. However, we can compare the results from the study of Makitie with the results from our study, since the Kaplan-Meier method and the Cox model were implemented in the former.

This study uses 167 patients who were enrolled from the years 1972 up to 1982 and all have choroidal or uveal melanoma. In this dataset, there are patients that died from other causes and so they are considered censored. At the end of the follow-up time, 37 of the 167 patients were alive without evidence of metastatic melanoma and 130 patients were had died. Of the 130 deaths, 80 were due to metastatic uveal melanoma, 9 were caused by carcinomas and 41 were deaths from other causes. The follow-up time was 22 years and the median survival was about 18-26 years. The covariates that were taken under consideration in the analysis of the Cox model are the epithelioid cells, the large tumor size, the high microvascular density and finally the ezrin immunoreactivity. Since the dataset is different, no comparisons concerning the probabilities of survival can be done. However, comparisons that concern the covariates that influence more the disease and comparisons that concern the gender of the patients and how it affects the survivability of the patients can be done. The figure that shows how the gender influences the survivability of the patients is the following.

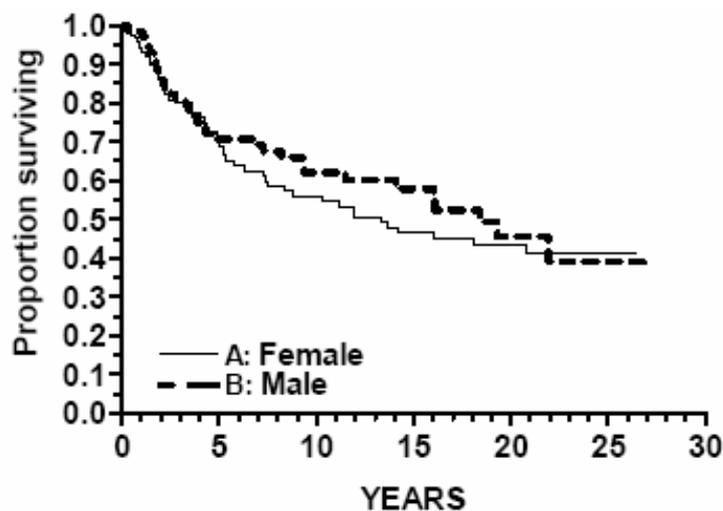


Figure 61: Diagram that shows how the survivability is affected by the gender in the study of Makitie [oth-8].

From the above figure it is shown that there is no difference between the two sexes, which means that the ocular melanoma affects the same the survivability of the two sexes. There are some studies that prove that women have better survival than men and other studies that claim that do not found any relationship between the two sexes. The above figure belongs to the second category of the studies. In our study, as shown in the following figure the females have greater survivability than the males, which means that our study belongs to the first category of the studies just mentioned.

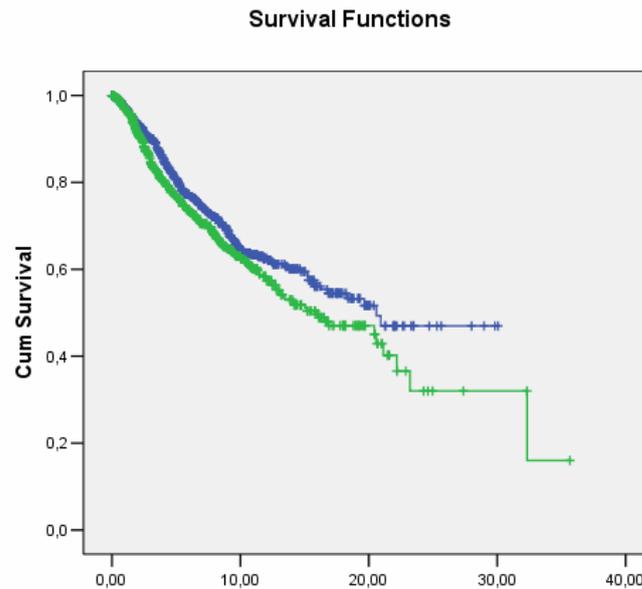


Figure 62: Survival functions of the male and female functions

The study of Makitie, besides the Kaplan-Meier method, also uses the Cox model. From the implementation of the Cox model the results that came up were that the covariate epithelioid cells influences more the disease and after that, the covariate largest basal diameter follows in influence. These results are confirmed with ours since we also concluded in the same results. Finally this study also concluded that the patients that have epithelioid cells have less probabilities of survival, as is shown in the following figure. This result is also verified from our study, as shown in figure 60.

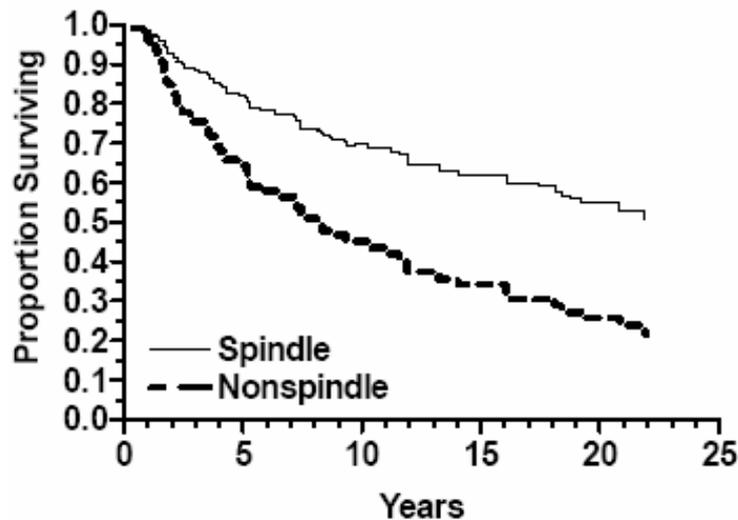


Figure 63: Survival curves for patients with epithelioid and non epithelioid cells, spindle, as they derived from the Makitie's study. [oth-8]

6.2.2.3 Comparison with a study which models survival of uveal malignant melanoma

This study was done from the Ocular Oncology Service and especially from the Department of Ophthalmology and from the Division of Biostatistics of the University of Minnesota by Tero Kivela and Patricia M. Grambsch. The dataset that was used consisted of 167 patients treated from 1972 to 1981 and the simulation was limited to 133 (80%) patients for whom there was complete data. The median follow-up of survivors was at about 17 years. The survival function was analyzed using the Kaplan-Meier product-limit method and the confidence intervals were calculated according to Greenwood method.

There has been a selection of categorical and continuous variables. Some of them were significant associated with the survival of the population and some of them were not. The variables that were chosen were the gender, the height, the largest basal diameter of the tumor, and the presence of epithelioid cells, which could be either absent, either present.

Since the dataset is different, no comparisons that concern the probabilities of survival can be done. However, comparisons that concern the covariates that influence more the disease and comparisons that concern the gender of the patients and how it affects the survivability of the patients can be done. The figure that shows how the gender influences the survivability of the patients is the following:

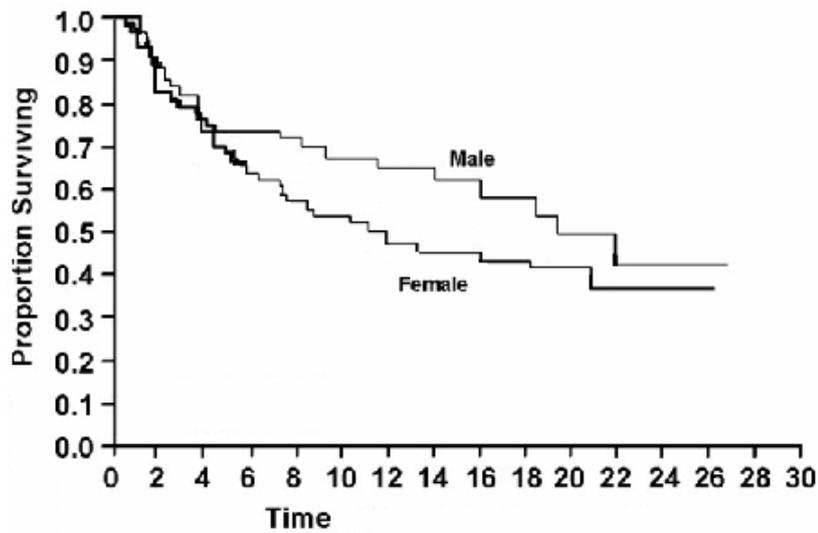


Figure 64: Diagram that shows how the survivability is affected by the gender in the study of Kivela [p-23].

This figure shows that the males have more probabilities of survival than the women, which is opposite to the results from our study. All these different results, from all the different studies that we presented, which concern the sex show that there is not a basic rule about the influence of the sex in the disease. This is because either the males will have more probabilities of survival, either the females or none of them.

This study, besides the Kaplan-Meier method, also uses the Cox model. From the analysis made by the Cox model the results that came up were that the covariate epithelioid cells influences more the disease. These results are confirmed with ours since we also concluded in the same fact. Finally this study also concluded that the patients that have epithelioid cells have less probabilities of survival, as is shown in the following figure. This result is also verified from our study, as shown in figure 60.

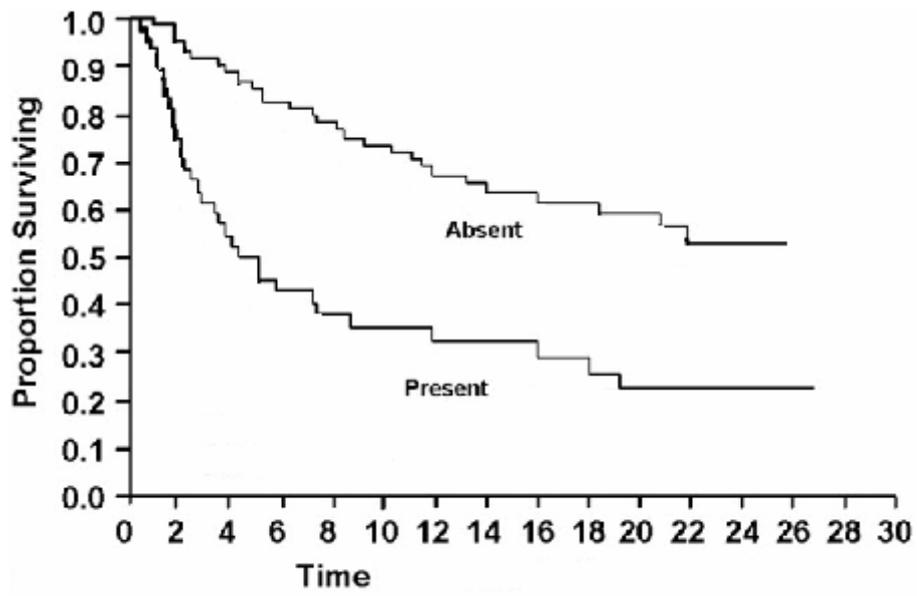


Figure 65: Survival curves for patients with presence of epithelioid cells and no presence of epithelioid cells as derived from the Kivela's study. [p-23]

Chapter 7

Conclusions and future work

7.1 Conclusions

In this study we concentrated on nonparametric methods in order to perform survival analysis. In addition, in a big part of this thesis we performed survival analysis by using a semiparametric model, the Aalen's additive model. Finally, we also made use of neural networks in order to test the Kaplan-Meier method. The tools that were used during the analysis in this study were the SPSS, and the Matlab.

As far as the nonparametric models are concerned, we concluded to survival functions that were very informative since very interesting results came up. We also concluded to cumulative hazard functions but these functions are not in general as informative as the survival functions. What is also very important is that all the nonparametric methods, the Kaplan-Meier, the Life-Table and the Nelson-Aalen, concluded to the same results.

The survival functions, that were almost the same, gave to the patients a probability of 20% to be cured. If we do not take under consideration the last death that takes place at 32, 33 years, which is the right thing to do, the probabilities of survival become 40%, which a pretty good percentage. The median survival varies from 18 to 19, 6 years in all the three methods.

From the three nonparametric methods, the one that is most commonly used is the Kaplan-Meier method. It is a very useful method and is also used by doctors in order to conclude to informative results. The main advantage of this nonparametric method is that in order to be used no knowledge of a parametric distribution is required and that is why sometimes is used as a guide so as the appropriate model for the analysis to be find. The main disadvantage is that it cannot estimate the effect of the covariates in the disease.

That is why we also used in this thesis the additive model, which has as main purpose to discover the effect of each one of the covariates. The covariates that were used in the analysis of the additive model were the "epithelioid cellularity", the "longest ultrasound basal dimension" and the "ultrasound height". The plots of the cumulative regression functions gave information about the effect of each one of the covariates. This was done by observing their slope so as to see when the slope changes. The main result that came up from the analysis of this model is that the

covariate “epithelioid cellularity” influences more the evolution of the disease with a big difference from the other two covariates. The covariate that influences more the disease after the “epithelioid cellularity”, is the “longest ultrasound basal dimension” and the last one is the “ultrasound height” that appears to have no significance effect.

In addition, the results that came up from the survival functions are also very informative. Because the “epithelioid cellularity” was the covariate that influences more the disease of the ocular melanoma we wanted to see how the survivability of the patients is influenced by the presence or the absence of the epithelioid melanoma cells. The results showed that the patients that are characterized by the presence of the epithelioid cells had less probabilities of survival than those patients that did not have that type of melanoma cells. This result was also verified by a number of other studies that were presented in chapter 6.

The additive model is a very powerful one because it concludes in separate regression functions for each one of the covariates and in this way we are able to see how the influence of the disease changes through the observation time.

Finally, the neural network that was used verified the results of the Kaplan-Meier method, since the results from the neural network and the results from the Kaplan-Meier were quite the same. Neural networks can be very effective in the survival analysis because they are very flexible and they can be used in very different ways.

7.2 Future work

The future work that is going to be proposed here concerns mainly the additive model because there are not many additional things that can be done with the nonparametric methods. So, the further work that can be done with the additive model includes:

- Use of the covariate age, which can be obtained from the date of birth and the date of death, so as to see how the survivability of the patients is influenced by their age.
- Use of martingale residuals, which are plotting methods for goodness of fit, so as to see if the additive model actually fits the data.

- Development of testing procedures so as to see if the plots of the cumulative regression functions show an overall influence of the covariate on survival or show a partial influence in a specific time period.
- Development of a direct method to interpret the cumulative regression functions, rather than the indirect way that we used in this thesis which is the interpretation of the slopes of the plots.

As further work we also suggest the use of other semiparametric methods, like the Cox model and the Cox-Aalen, which combines the very useful Cox model and the additive model by multiplying these two models. In addition, the proportional excess hazard model combines the Cox model and the additive model by adding them. The Cox model is the model that is the most used in regression analysis and that is why the results from this model would be very interesting. Comparisons with the Cox-Aalen, the proportional excess hazard model and the Aalen's additive model would be very interesting and informative.

Finally, the use of other aspects of neural networks, like the ones we presented in the introduction of this thesis, would also be very useful and interesting.

Appendix A

In appendix A all the codes that concern the results form the nonparametric models (Kaplan-Meier, Life-Table, Nelson-Aalen) are given and also the code for the implementation of the neural network is given. Also besides the codes, some figures that show the implementation of the SPSS program and the curve fitting tool of Matlab are given.

Kaplan-Meier estimator

1. **The following code is computing the Kaplan –Meier estimator, the Cumulative Hazard Function and the plots of the survival function and the Cumulative Hazard Function are also derived.**

```
sum1=ones(979,1); % initialize sum

for j=2:979
    tmp(j)=1-(dead(j)/risk(j));
    sum1(j)=sum1(j-1)*tmp(j) %computation of the cumulative probability
end
figure(1)%plot of survival curve
plot(time,sum1);
xlabel('time(years)')
ylabel(' Cumulative Survival ')
title('Survival curve')

figure(2)
H=-log(sum1);%computation and plotting of the cumulative hazard function
plot(time,H);
xlabel('time(years)')
ylabel(' Cumulative Hazard ')

clear j tmp
```

2. **The following code computes the 95% Confidence Intervals**

```
function standard_error()

lower = zeros(1,979);%initializing lower and upper bound
upper = zeros(1,979);

for t = 1:979
    w = [1:t];
    ni = risk(w);
    di = dead(w);
    ri = ni - di;
    sum = 0;
    for j = 1:length(w)
```

```

sum = sum + di(j)/(ni(j)*ri(j)); %computation of the basic sum
end
sum = sqrt(sum);

lower(t) = sum1(t) - 1.96*sum1(t)*sum;%computation of lower bound

upper(t) = sum1(t) + 1.96*sum1(t)*sum;%computation of upper bound

end

line1 = plot(time,sum1,'k',time,lower,':k',time,upper,':k');

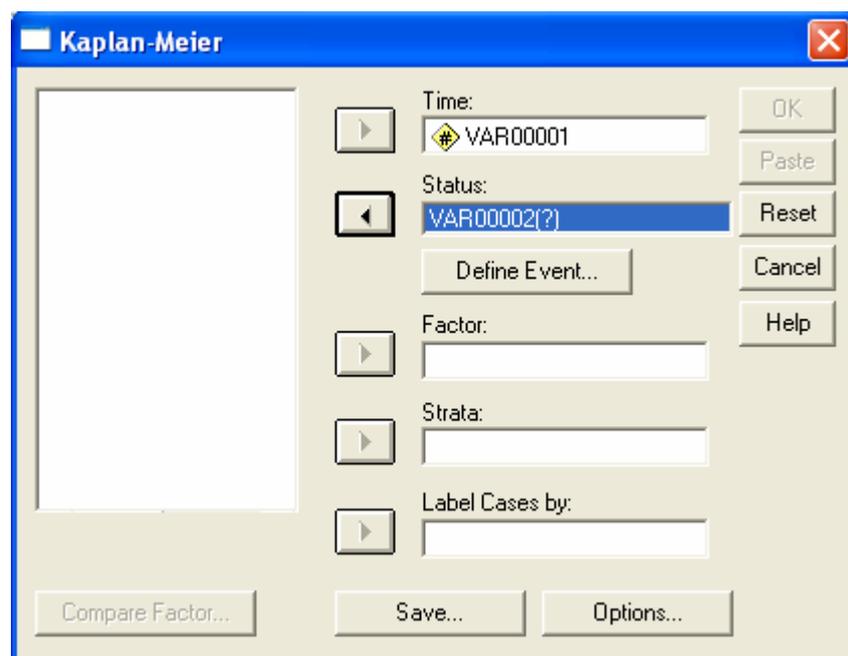
line2 = line(time,sum1,'LineWidth',3,'Color',[0 0 0]);
xlabel('time(years)')
ylabel('Survival Probability')
title('Survival Curve')

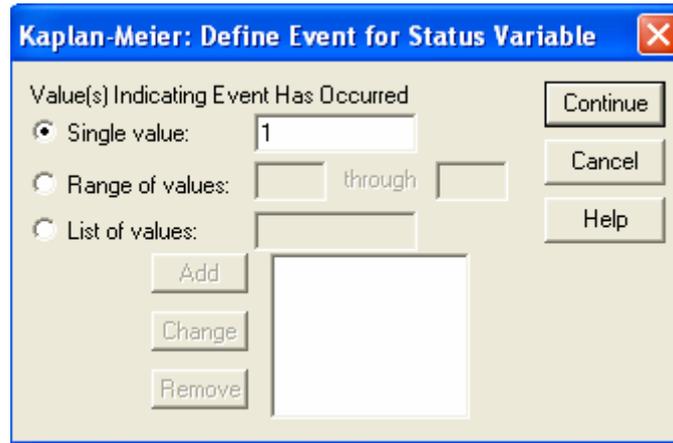
```

3. SPSS program steps

In order to obtain the survival function from the Kaplan-Meier method using the SPSS program the following steps have to be done.

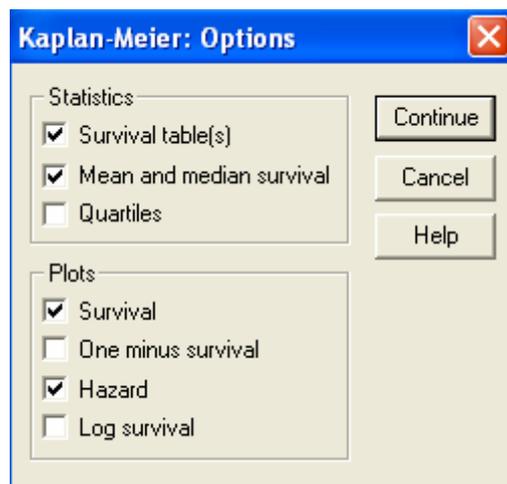
1. From the menu of the SPSS select: Analyze→Survival→ Kaplan-Meier and the following window will appear.
2. In the time field the data that contain the survival years are put and in the status field it should be placed the variable which indicates whether the event was a failure or a censored observation. Click on the Define event button to define codes was used to identify those two situations.





The figure shown above is the SPSS dialog box where failures and censoring are distinguished. In this data set, a value of 1 indicates a failure and 0 represents censoring.

3. Finally by clicking the options button the following window appears where it can be selected what plots will appear.



4. The following figure computes the exponential fitting

Fit Editor

New fit Copy fit

Fit Name: fit 1

Data set: sum1 vs. time Exclusion rule: anna

Type of fit: Custom Equations Center and scale X data

Custom Equations

exp(-b*x)

New equation...
Copy equation
Delete equation

Fit options... Immediate apply Cancel Apply

Results

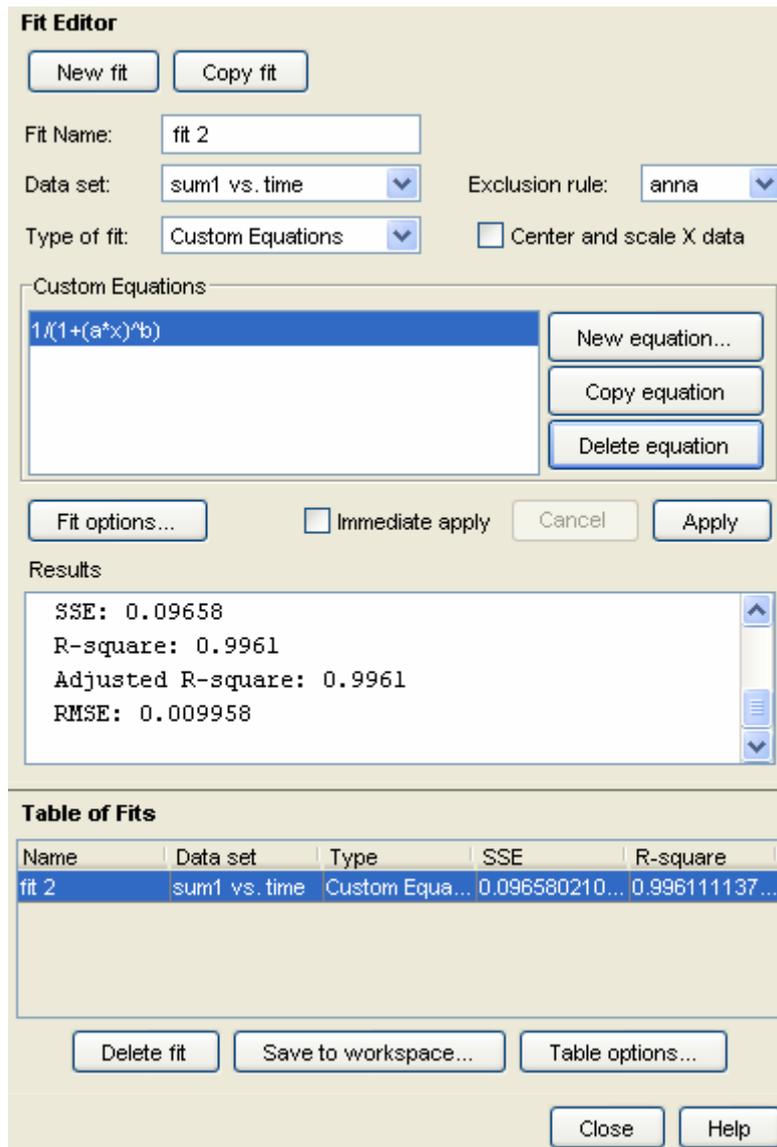
SSE: 0.4506
R-square: 0.9819
Adjusted R-square: 0.9819
RMSE: 0.0215

Table of Fits

Name	Data set	Type	SSE	R-square
fit 1	sum1 vs. time	Custom Equa...	0.450575442...	0.981857298...

Delete fit Save to workspace... Table options...

5. The following figure computes the log-logistic fitting



Life-table estimator

1. The following code computes the survival function

```
S=zeros(36,1);
S(1)=1;
for i=2:36
    S(i)=p(i-1)*S(i-1);
end
plot(time,S)
xlabel('time(years)')
ylabel('Survival function')
clear i
```

2. The following code computes the density function

```

f=zeros(36,1);
for i=1:36
    f(i)=(S(i)*q(i))/b(i)
end

plot(time,f)
xlabel('time(years)')
ylabel('probability density')
title('probability density function')
clear i

```

3. The following code computes the hazard function

```

h=zeros(36,1)
for i=1:36
    h(i)=(2*q(i))/(b(i)*(1+p(i)))
end

plot(time,h)
xlabel('time(years)')
ylabel('hazard rate')
title('hazard function')
clear i

```

4. The following code computes the log-logistic fit

```

function log_logistic_fit

g = fitype('1/(1+a*(x^c))');
F=FITOPTIONS('METHOD','NonLinearLeastSquares','StartPoint',[0.1,0.1]);
[FITTEDMODEL,GOODNESS,OUTPUT]=fit(time,S,g,F);

%-----

a = 0.04005;
c = 1.127;

for i=1:36
    y(i)=1/(1+a*((time(i))^c));
end

figure(1)
hold on
plot(time,y);
xlabel('Survival Years');
ylabel('Cumulative Survival');
plot(time,S,'*');
hold off;

```

FITTEDMODEL =

General model:

$$\text{FITTEDMODEL}(x) = 1/(1+a*(x^c))$$

Coefficients (with 95% confidence bounds):

$$a = 0.04005 \quad (0.02924, 0.05085)$$

$$c = 1.127 \quad (1.034, 1.219)$$

GOODNESS =

Sse: 0.034562

Rsquare: 0.9764

Dfe: 34

Adjrsquare: 0.9757

Rmse: 0.031883

Nelson-Aalen

1. The following code computes the cumulative hazard and survival function with the Nelson-Aalen method.

```
for i = 1:length(dead)
    sum = 0;
    for j = 1:i
        sum = sum + dead(j)/risk(j);
    end
    result(i) = sum;
end
```

```
plot(time,result)
xlabel('time (years)')
ylabel('cumulative rate')
```

```
S=exp(-result);
xlabel('time (years)')
ylabel('survival probability')
```

Neural Networks

1. The following code trains and simulates the neural network

```
function network_kaplan
load net_kaplan;

train_p=[];
train_t=[];

sim_p=[];
sim_t=[];

tmp=[];

for i=1:length(p)
    tmp(i)=0;
end

for i=1:2:length(p)
    tmp(i)=1;
end
m=1;
n=1;
for i=1:length(p)
    if tmp(i)==1
        sim_p(:,m)=p(:,i)
        sim_t(m)=t(i)
        m=m+1;
    end

    if tmp(i)==0
        train_p(:,n)=p(:,i)
        train_t(n)=t(i)
        n=n+1;
    end
end

net=newff(minmax(train_p),[2          5          5          5          5
1], {'tansig','tansig','tansig','tansig','tansig','purelin'}, 'trainlm','learngdm','mse');
%creation of a feedforward backpropagation neural network. It has 2 inputs,
%4 hidden layers with 5 neurons each and one output. The transfer functions
%for the hidden layers are the tansig and the transfer function for the
%output is the purelin.

net.trainParam.epochs=2000;% initialization to 2000 epochs
net.trainParam.goal=0;% the training finishes when it reaches error 0.This is the
performance goal
net=train(net,train_p,train_t);%the neural network is trained
```

```

Y=[];%network outputs
m=1;
for i=1:490
%   SIM_VECTOR=[p(1,i);p(2,i);p(3,i)];
   Y(m)=sim(net,[sim_p(1,i);sim_p(2,i)]);
   a_sim(m)=Y(m);
   t_sim(m)=sim_t(m);
   m=m+1;
end

figure(1)
[m(2),b(2),r(2)]=postreg(a_sim,t_sim);%computation of the correlation coefficient

% figure(2)
% hold on
% plot(time,t)
% plot(time,Y,'r')
% xlabel('time(years)')
% ylabel('Cumulative Survival Function')
% title('Kaplan-Meier vs Neural Network output')
% hold off

```

With the newff function the following network is created :

```
net =
```

Neural Network object:

architecture:

numInputs: 1

numLayers: 6

biasConnect: [1; 1; 1; 1; 1; 1]

inputConnect: [1; 0; 0; 0; 0; 0]

layerConnect: [6x6 boolean]

outputConnect: [0 0 0 0 0 1]

targetConnect: [0 0 0 0 0 1]

numOutputs: 1 (read-only)

numTargets: 1 (read-only)

numInputDelays: 0 (read-only)

numLayerDelays: 0 (read-only)

subobject structures:

inputs: {1x1 cell} of inputs

layers: {6x1 cell} of layers

outputs: {1x6 cell} containing 1 output

targets: {1x6 cell} containing 1 target

biases: {6x1 cell} containing 6 biases

inputWeights: {6x1 cell} containing 1 input weight

layerWeights: {6x6 cell} containing 5 layer weights

functions:

adaptFcn: 'trains'
initFcn: 'initlay'
performFcn: 'mse'
trainFcn: 'trainlm'

parameters:

adaptParam: .passes
initParam: (none)
performParam: (none)
trainParam: .epochs, .goal, .max_fail, .mem_reduc,
.min_grad, .mu, .mu_dec, .mu_inc,
.mu_max, .show, .time

weight and bias values:

IW: {6x1 cell} containing 1 input weight matrix
LW: {6x6 cell} containing 5 layer weight matrices
b: {6x1 cell} containing 6 bias vectors

other:

userdata: (user stuff)

Appendix B

In appendix B all the codes that concern the implementation of the Aalen's additive models are given

1. The code that implements the counting of the death and censored events in the ordered times.

```
%find maximum time index (N) and maximum sample index (M)
N=round(100*max(time));
M=length(death);

%initialize temporary tables;
ttmp(N)=0;
stmp(N)=0;
mtmp(N)=0;

%initialize mortality (m) and survival (s) vectors
m=[];
s=[];

%find the indexes corresponding to the available times recorded in time
%table

for i=1:M
    ttmp(round(time(i)*100))=1;
end

%Calculate mortality vector
for i=1:M,
    if death(i)==1
        mtmp(round(time(i)*100))=mtmp(round(time(i)*100))+1;
    end
end
% form mortality vector rows ("time" "number of deaths")
for i=1:N
    if ttmp(i)~=0
        tmp=[i/100 mtmp(i)];
        m=[m;tmp];
    end
end
%-----
%Calculate censored vector
for i=1:M,
    if death(i)==0
        stmp(round(time(i)*100))=stmp(round(time(i)*100))+1;
    end
end
```

```

% form censored vector rows ("time" "number of survival")
for i=1:N
    if ttmp(i)~=0
        tmp=[i/100 stmp(i)];
        s=[s;tmp];
    end
end
end
%-----
%clear not needed variables
clear i mtmp stmp tmp ttmp M N

```

2. The code that implements the computation of the Y matrix

```

%number of samples from patients
N=length(data);
%number of existing time indexes
M=length(s);

%create initial Y table (for 1 time sample)
data_all=[ones(N,1) data];

%initialize Y for all time moments
Y(N,4,M)=0;

for k=1:M
    Y(:,k)=data_all;
end

%total_m_s: number of deaths+censored UP TO time moment i (inclusive)
total_m_s=0;

for i=2:M %traverse table Y for all time moments
    m_s=m(i-1,2)+s(i-1,2); %m_s:sum of deaths+censored for moment i ONLY!!!!
    total_m_s=total_m_s+m_s;
    for j=1:total_m_s
        Y(j,4)=zeros(1,4); %replace total_m_s number of lines with zeros for time
        moment (i+1)
    end
end
end

```

3. The code that implements the computation of the X matrix

```

%initialize Y(t) 1<=t<=560
yt(743,4)=0;
%initialize number of time moments
M=length(s);

%traverse Y for all time moments
for i=1:M,
    yt(:,i)=Y(:,i);
end

```

```

if det(yt'*yt)==0
    %if det(yt'*yt)==0 then break.'i' holds the time moment where
    %det(yt'*yt)~0;
    end_X=i-1;
    break;
end
end
end

%initialize X table.The 3rd dimension is of length i-1, since we want only
%the time moments where det(yt'*yt)==0
X(4,743,end_X)=0;

for i=1:end_X
    %get Y for time moment i
    yt(:,i)=Y(:,i);
    %X(t)=(Y'(t)*Y(t))^-1 * Y'(t)
    X(:,i)=inv(yt'*yt)*yt';
end

clear end_X yt i ans

```

4. The code that implements the computation of the cumulative regression functions and creates the plots for each covariate.

```

tmp=size(X);% we get all the dimensions of matrix X
MaxX=tmp(3);% we make use only the third dimension of matrix X
xt=[];
k=0;

for i=1:MaxX % we get only the ordered event times where m is different than zero
    if m(i,2)>0
        k=k+1;
        xt(:,i)=X(:,k);%we make use only the X values that correspond to ordered
event times
    end
end

At(4,1,k)=0;% the dimension of A matrix

sum=zeros(4,1); %initialize sum
k=0;
p=0;
for i=1:MaxX
    if m(i,2)>0
        k=k+1;
        Ii=zeros(743,1); %initialize I

        for j=1:length(death)
            if (death(j)==1)

```

```

        if (time(j)==m(i,1))% I is a vector of zeros except for a one corresponding
to the individual

```

```

        % who experiences an event at time Tk

```

```

        Ii(j)=1;

```

```

            p=p+1;

```

```

        end

```

```

    end

```

```

end

```

```

sum = sum + xt(:,i)*Ii; %A(t)=Σ X(Tk)*Ik for Tk<=t

```

```

At(:,k)=sum;

```

```

end

```

```

end

```

```

%.....

```

```

k=0;

```

```

time_all=[];

```

```

for i=1:MaxX % we get only the ordered event times where m is different than zero

```

```

    if m(i,2)>0

```

```

        k=k+1;

```

```

        time_all(k)=m(i);%we make use only the time values that correspond to ordered
event times

```

```

    end

```

```

end

```

```

% Z(length(At(1,1,:)))=0;

```

```

Z1(length(At(2,1,:)))=0;

```

```

Z2(length(At(3,1,:)))=0;

```

```

Z3(length(At(4,1,:)))=0;

```

```

% for j =1:length(At(1,1,:))% we take the first line of all the sheets of A

```

```

%   Z(j) = At(1,1,j);

```

```

% end

```

```

% subplot(2,2,1);

```

```

% plot(time_all,Z); % we plot the components of A against time which result in the
cumulative regression plots.

```

```

% hold on

```

```

%.....plots.....

```

```

for j =1:length(At(2,1,:))

```

```

    Z1(j) = At(2,1,j);

```

```

end

```

```

% subplot(2,2,2);

```

```

plot(time_all,Z1);

```

```

xlabel('time(years)');

```

```

ylabel('cumulative regression');

```

```

title('covariate lubd')

```

```

hold on

```

```

for j =1:length(At(3,1,:))

```

```

    Z2(j) = At(3,1,j);
end
% subplot(2,2,3);
plot(time_all,Z2);
xlabel('time(years)');
ylabel('cumulative regression function');
title('covariate uh')
hold on

for j =1:length(At(4,1,:))
    Z3(j) = At(4,1,j);
end
% subplot(2,2,4);
plot(time_all,Z3);
xlabel('time(years)');
ylabel('cumulative regression function');
title('covariate epi')
hold on
clear MaxX xt k i sum j tmp

```

5. The code that implements the computation of the survival functions for each patient.

```

function survival(patient)

load final;
tmp=size(X);% we get all the dimensions of matrix X
MaxX=tmp(3);% we make use only the third dimension of matrix X
xt=[];
k=0;

for i=1:MaxX % we get only the ordered event times where m is different than zero
    if m(i,2)>0
        k=k+1;
        xt(:,i)=X(:,k);%we make use only the X values that correspond to ordered
event times
    end
end

Pt(1,1,k)=0;
Z=[];
Z=data_all(patient,:);
Z=Z';
sum=ones(1,1); %initialize sum
k=0;

for i=1:MaxX
    if m(i,2)>0
        k=k+1;

```

```

Ii=zeros(743,1); %initialize I
for j=1:length(death)
    if (death(j)==1)
        if (time(j)==m(i,1))% I is a vector of zeros except for a one corresponding
to the individual
            % who experiences an event at time Tk
                Ii(j)=1;
            end
        end
    end
end

sum = sum *(1-(xt(:,i)*Ii)*Z);
Pt(:,k)=sum;

%    Ht=-log(Pt);

    end
end

k=0;
time_all=[];
for i=1:MaxX % we get only the ordered event times where m is different than zero
    if m(i,2)>0
        k=k+1;
        time_all(k)=m(i);%we make use only the time values that correspond to ordered
event times
    end
end

for j =1:length(Pt(1,1,:))% we take the first line of all the sheets of A
    R(j) = Pt(1,1,j);
end

% for j =1:length(Ht(1,1,:))% we take the first line of all the sheets of A
%    H(j) = Ht(1,1,j);
% end

figure(1)
hold on
plot(time_all,R)
xlabel('Survival Years');
ylabel('Survival Function');
grid on
hold off

% figure(2)
% plot(time_all,H);
% xlabel('Survival Years');
% ylabel('Cumulative Hazard');

```

6. The code that implements the computation of the survival functions for the covariate epi which is divided into epi 0 and epi 1

```
function survival_epi_0_1
load survival_epi0_1.mat;

survival0=[];
survival1=[];
m=1;
n=1;
for i=1:743
    if data_all(i,4)==1
        survival1(:,m)=survival(i)
        m=m+1;
    end
    if data_all(i,4)==0
        survival0(:,n)=survival(i)
        n=n+1;
    end
end

figure (1)
hold on;
for i=1:464
    if survival1(:,i)<=1
        plot(time_all,survival1(:,i),'r');
    end
end

for i=1:279
    if survival0(:,i)<=1
        plot(time_all,survival0(:,i));
    end
end
xlabel('Survival time');
ylabel('Survival Function')
hold off;
```

7. The code that implements the computation of the survival functions for the covariate lubd which is divided into lubd big($\geq 14,9$) and lubd small($< 14,9$)

```
function survival_lubd

load survival_epi0_1.mat

lubd0=[];
lubd1=[];
```

```

m=1;
n=1;
p=1;
k=1;

for i=1:743
    if data_all(i,4)==1
        if data_all(i,2)>=14,9
            lubd1(:,m)=survival1(:,k);
            m=m+1;
            k=k+1;

        end

        if data_all(i,2)<14,9
            lubd0(:,n)=survival1(:,k);
            k=k+1;
            n=n+1;

        end

    end

    if data_all(i,4)==0
        if data_all(i,2)<14,9
            lubd0(:,n)=survival0(:,p);
            n=n+1;
            p=p+1;
        end

        if data_all(i,2)>=14,9
            lubd1(:,m)=survival0(:,p);
            m=m+1;
            p=p+1;

        end

    end

end

figure (1)
hold on;
for i=1:360
    if lubd1(:,i)<=1
        plot(time_all,lubd1(:,i),'r');
    end
end

for i=1:383
    if lubd0(:,i)<=1
        plot(time_all,lubd0(:,i));
    end
end

```

```

    end
end
xlabel('Survival time');
ylabel('Survival Function')
hold off;

```

8. The code that implements the computation of the survival functions for the covariate uh which is divided into uh big(≥ 10) and uh small(< 10)

```

function survival_uh

load survival_epi0_1.mat
uh0=[];
uh1=[];
m=1;
n=1;
p=1;
k=1;

for i=1:743
    if data_all(i,4)==1
        if data_all(i,3)>=10
            uh1(:,m)=survival1(:,k);
            m=m+1;
            k=k+1;
        end
        if data_all(i,3)<10
            uh0(:,n)=survival1(:,k);
            k=k+1;
            n=n+1;
        end
    end

    if data_all(i,4)==0
        if data_all(i,3)<10
            uh0(:,n)=survival0(:,p);
            n=n+1;
            p=p+1;
        end
        if data_all(i,3)>=10
            uh1(:,m)=survival0(:,p);
            m=m+1;
            p=p+1;
        end
    end

end

figure (1)

```

```

hold on;
for i=1:194
    if uh1(:,i)<=1
        plot(time_all,uh1(:,i),'r');
    end
end

for i=1:549
    if uh0(:,i)<=1
        plot(time_all,uh0(:,i));
    end
end
xlabel('Survival time');
ylabel('Survival Function')
hold off;

```

9. The code that implements the computation of the survival functions for the covariate epi and lubd. The data set is divided into lubd big and epi 1, and to lubd big and epi 0

```

function survival_lubd_epi

load survival_epi0_1.mat;
% s0=[];
% s1=[];

figure (2)
hold on;
for j=1:743
    for i=1:464
        if data_all(j,2)>=14,9
            if survival1(:,i)<=1
                plot(time_all,survival1(:,i),'r');
            end
        end
    end
end
for j=1:743
    for i=1:279
        if data_all(j,2)>=14,9
            if survival0(:,i)<=1
                plot(time_all,survival0(:,i));
            end
        end
    end
end
xlabel('Survival time');
ylabel('Survival Function')
hold off;

```

References

Web sources:

1. http://www.cancer.org/docroot/CRI/content/CRI_2_4_3X_How_is_eye_cancer_staged_74.asp?sitearea=
2. http://bhj.org/april2005/htm/case_intraocular_197.htm
3. <http://www.jhu.edu/wctb/coms/booklet/book2.htm>
4. http://www.fairview.org/healthlibrary/content/ea_imelanom_oph.htm
5. <http://www.cancerbacup.org.uk/Cancertype/Eye/Melanomaoftheeye>
6. <http://www.mayoclinic.com/health/ocular-melanoma/AN00956>
7. <http://www.bccancer.bc.ca/PPI/TypesofCancer/EyeorOrbital/OcularMelanoma/default.htm>
8. http://bjo.bmjournals.com/cgi/reprint/83/10/1190?maxtoshow=&HITS=&hits=&RESULTFORMAT=&fulltext=ocular+cancer&searchid=1134490897572_1413&stored_search=&FIRSTINDEX=0&resourcetype=1&journalcode=bjophthalmol
9. <http://www.nlm.nih.gov/medlineplus/ency/article/001022.htm#top>
10. <http://www.stlukeseye.com/anatomy/AngleStructures.asp>
11. <http://www.stlukeseye.com/anatomy/OpticNerve.asp>
12. <http://www.stlukeseye.com/anatomy/Cornea.asp>
13. <http://www.stlukeseye.com/anatomy/Macula.asp>
14. <http://www.umdj.edu/eyeweb/faqs/vitrelect.html>
15. http://www.uveitis.org/patient/glossary/m_s.html#M
16. <http://www.stlukeseye.com/anatomy/Fovea.asp>
17. http://www.cancerguide.org/scurve_km.html
18. <http://staff.pubhealth.ku.dk/~ts/surv/4.Dag1.pdf>
19. <http://www.cceb.upenn.edu/localio/EPI521/2006/v2part7.pdf>
20. <http://courses.umass.edu/be640/s05/topics/surv.pdf>
21. <http://www.graphpad.com/www/book/survive.htm>
22. <http://www.cmh.edu/stats/model/survival/kaplan.asp>
23. http://www.cancerguide.org/scurve_basic.html
24. <http://etd.lsu.edu/docs/available/etd-04092004-081738/unrestricted/05chapter4.pdf>
25. <http://www.maths.soton.ac.uk/teaching/units/math6021/Chapter1.pdf>

26. <http://www.maths.soton.ac.uk/teaching/units/math6021/Chapter2.pdf>
27. http://www.ats.ucla.edu/stat/sas/seminars/sas_survival/default.htm
28. http://en.wikipedia.org/wiki/Survival_analysis
29. http://mercury.bio.uaf.edu/courses/wlf625/my_notes/lec9_covariates.htm
30. <http://www.stat.umu.se/~goran.brostrom/courses/eventhistory/rep4.pdf>

Books:

1. Elisa T.Lee and John Wenyu Wang, “*Statistical Methods for Survival Data Analysis*”, 3rd Edition 2003, John Wiley & sons, inc, publication
2. Ettore Marubini and Maria Grazia Valsecchi, “*Analysing survival data from clinical trials and observational studies*”, John Wiley & sons, inc, publication
3. Kevin Gurney (1997), “*An introduction to neural Networks*”, CPR Press

Papers:

1. Aalen, O.O. (1989), “*A linear regression model for the analysis of life times*”. *Statistics in Medicine* **8** 907-925.
2. Aalen, O.O. (1993), “*Further results on the non - parametric linear regression model in survival analysis*”. *Statistics in Medicine* **12** 1569-1588
3. Alicia M.Howell and John P.Klein, “*A SAS Macro for the additive hazards regression model*”
4. Anna Torner. (2003) “*Proportional hazards and additive regression analysis of survival for severe breast cancer*”
5. Tyler Smith, and Besa Smith, “*Survival Analysis And The Application Of Cox's Proportional Hazards Modelling Using SAS*”
6. Svetlana Borovkova (2002) “*Analysis of Survival data*”
7. L. Bottaci, P.J. Drew, J.E. Hartley, M.B. Hadfield, R. Farouk, P.W.R. Lee, I.M.C. Macintyre, G.S. Duthie, and J.R.T. Monson. “*Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions.*” *The Lancet*, 350:469–472, 1997.
8. L. Ohno-Machado. “*Sequential use of neural networks for survival prediction in aids.*” *Journal of the American Medical Informatics Association*, 3:170–174, 1996.

9. M. De Laurentiis and P. Ravdin. “*Survival analysis of censored data: neural network analysis detection of complex interactions between variables.*” *Breast Cancer Research and Treatment*, 32:113–118, 1994.
10. P.M. Ravdin and G.M. Clark. “*A practical application of neural network analysis for predicting outcome of individual breast cancer patients*”. *Breast Cancer Research and Treatment*, 22:285–293, 1992.
11. P. Lapuerta, S.P. Azen, and L. LaBree. “*Use of neural networks in predicting the risk of coronary artery disease*”. *Computers and Biomedical Research*, 28:38–52, 1995.
12. D. Faraggi and R. Simon. “*A neural network model for survival data*”. *Statistics in Medicine*, 14:73–82, 1995.
13. W.N. Street. “*A neural network model for prognostic prediction.*” In *Proceedings of the Fifteenth International Conference on Machine Learning (ICML)*, pages 540–546, Madison, Wisconsin, U.S., 1998. Morgan Kaufmann.
14. D.R. Mani, J. Drew, A. Betz, and P. Datta. “*Statistics and data mining techniques for life- time value modelling*”. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 94–103, San Diego, CA, U.S., 1999.
15. S.F. Brown, A. Branford, and W. Moran. “*On the use of artificial neural networks for the analysis of survival data*”. *IEEE Transactions on Neural Networks*, 8:1071–1077, 1997.
16. L. Ohno-Machado, M.G. Walker, and M.A. Musen “*Hierarchical Neural Networks for Survival analysis*”
17. B.D Ripley and Ripley “*Neural Networks as Statistical methods in survival analysis*”
18. P.J.G. Lisboa, H. Wong, P. Harris, R. Swidnell “*A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer*” *Artificial Intelligence in Medicine*, 28 :1-23, 2003
19. Torben Martinussen and Thomas Scheike “*Goodness of fit for Cox’s regression Model, Extensions of Cox’s regression Model*” (2004)
20. Aalen, O.O “*A model for nonparametric regression analysis of counting processes*”, *Lecture notes in Statistics*, 2, Springer, New York, 1980, 1-25
21. Azzam F G TakTak, Anthony C Fisher and Bertil E Damato, “*Modelling survival after treatment of intraocular melanoma using artificial neural*

- networks and Bayes theorem*”, Physics in Medicine and Biology, 49, (2004), 87-98
22. Peter Isager, Niels Ehlers and Jens Overgaard, “*Prognostic factors for survival after enucleation for Choroidal and ciliary body melanomas*”, Acta Ophthalmologica Scandinavica, 82, (2004), 517-525
23. Tero Kivela and Patricia M. Grambsch, “*Evaluation of Sampling Strategies for Modeling Survival of Uveal Malignant Melanoma*”, Investigative Ophthalmology and Visual Science. 2003;44:3288-3293

Other sources:

1. online support of the Biopatern group
2. Curve fitting Toolbox of Matlab
3. Neural Network Toolbox
4. Ammar Belatreche, “*Chapter 3: Artificial Neural Networks*”, University of Ulster, Magee
5. Mei-Cheng Wang “*Summary notes for Survival Analysis*”, Johns Hopkins University, Department of Biostatistics, (2005)
6. Jaap Murre, “*Kernthema Lecture Connectionism*” University of Amsterdam
7. Fotini Agrafioti “*Survival analysis of ocular melanoma patients using semiparametric regression models and neural networks*”, (2006)
8. Teemu Makitie, “*Prognostic indicators in choroidal and ciliary body melanoma*”, University of Helsinki, (2001)