



Stockholms  
universitet

# Surviving non-proportional hazards with flexible parametric modeling

Martin Andersson

Masteruppsats 2024:9  
Matematisk statistik  
Juni 2024

[www.math.su.se](http://www.math.su.se)

Matematisk statistik  
Matematiska institutionen  
Stockholms universitet  
106 91 Stockholm

# Surviving non-proportional hazards with flexible parametric modeling

Martin Andersson\*

June 2024

## Abstract

In this thesis we have investigated the difference in output between the base Cox regression model (a proportional hazard model) and the flexible parametric model (a more general survival model). The comparison was performed on a linked data-set from the Swedish Prescribed Drug and the National Patient Register. The data-set was intended to statistically test if treatments prescribed to patients with atopic dermatitis or psoriasis had a significant effect on developing venous thromboembolism. This data-set was used to create a plasmode simulation to test how the Cox model performed on a data-set with proportional hazards and one with simulated crossing hazard rate functions. Additionally, a Cox model and a flexible parametric model was fitted to the data to compare the results, specifically how the flexible parametric model captures the hazard rates using natural splines.

Moreover, the question on how to deal with non-proportional hazards is discussed. In addition to being able to visualize the hazard rates, the flexible parametric model can also model interaction between the covariates and time using natural splines. We also investigated how to test for and identify non-proportional hazards.

The results indicate that the base Cox model can be seen as a special case of the more general flexible parametric model. As such, if the results of the base Cox model are the main interest, the flexible parametric model can supply those, along with several other results the Cox model cannot. While the flexible parametric model captures non-proportional hazards, testing for these kinds of hazards is not easy. The tests tend to have low power and some need to be performed visually. However, cases with visually non-crossing hazards but with tests positive for non-proportional hazards do not necessarily make the results of a proportional hazards model unusable. Crossing hazards on the other hand are likely to be very obvious during visual tests and can be modeled accordingly when spotted. The proportional hazards model, either using the Cox method or the flexible parametric model is therefore still generally usable, even if the hazards are not exactly proportional.

---

\*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.  
E-mail: [martin.hans.erik.andersson@gmail.com](mailto:martin.hans.erik.andersson@gmail.com). Supervisor: Johannes Heiny, Marie Linder.

## Acknowledgements

I would like to thank my supervisors Marie Linder and Johannes Heiny for all their help, advice and constructive criticism.

I would also like to thank Julia Eriksson and David Hägg for helping me with the technical aspects when working on location at KI.

Thanks to Paul Lambert for letting me use his illustrations of splines.

All programming was performed in the R-language by RStudio Team (2020).

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
1.1	Survival analysis in biostatistics . . . . .	6
1.2	Cox regression . . . . .	7
1.3	Flexible parametric model (FPM) . . . . .	9
1.4	Identifying non-proportional hazards . . . . .	10
1.5	The dermatology data-set . . . . .	10
1.6	Plasmode simulation . . . . .	12
1.7	Structure of the thesis . . . . .	12
<b>2</b>	<b>Data</b>	<b>14</b>
2.1	Dermatology data-set on venous thromboembolism . . . . .	14
2.2	Visualizing the data-set . . . . .	14
<b>3</b>	<b>Theory</b>	<b>20</b>
3.1	Survival functions, hazard rates and hazard ratios. . . . .	20
3.1.1	Kaplan-Meier estimate of the survival function. . . . .	22
3.2	Cox regression . . . . .	22
3.3	Flexible parametric model . . . . .	23
3.3.1	Natural splines . . . . .	23
3.3.2	Modelling the hazard rate using splines . . . . .	25
3.4	Tests for non-proportional hazards . . . . .	26
3.4.1	Log-minus-log survival plots . . . . .	26
3.4.2	Schoenfeld residuals . . . . .	26
3.4.3	Schoenfeld test . . . . .	27
3.5	Simulation . . . . .	27
3.5.1	Plasmode simulation . . . . .	27
3.5.2	The Weibull distribution . . . . .	28
3.5.3	Simulating crossing hazards . . . . .	29
3.5.4	Simulating the piecewise exponential distribution . . . . .	29
3.5.5	Details of the simulation . . . . .	31
3.6	Comparing the models . . . . .	31
3.6.1	Estimate of $\beta$ ( $\hat{\beta}$ ) . . . . .	31
3.6.2	Akaike's information criterion (AIC) . . . . .	32
3.6.3	Concordance . . . . .	32
3.6.4	Schoenfeld test $p$ -value . . . . .	32
3.6.5	Percentage of uncensored observations . . . . .	33

<b>4</b>	<b>Results</b>	<b>34</b>
4.1	Simulation . . . . .	34
4.2	Kaplan-Meier plots . . . . .	35
4.3	Log-minus-log survival plots . . . . .	37
4.4	Cox model . . . . .	38
4.4.1	Schoenfeld test . . . . .	39
4.4.2	Schoenfeld residuals . . . . .	39
4.5	Flexible parametric model . . . . .	41
4.5.1	Hazard graphs . . . . .	42
<b>5</b>	<b>Discussion and conclusions</b>	<b>44</b>
5.1	Hypothesis tests in survival analysis . . . . .	44
5.2	Simulation . . . . .	45
5.3	Cox Regression variants . . . . .	46
5.4	Data sets with obviously non-proportional hazards. . . . .	46
5.5	Model selection . . . . .	47
5.6	Complications of survival analysis . . . . .	48
5.6.1	The frailty effect . . . . .	48
5.6.2	Immortal time bias . . . . .	48
5.6.3	Causality . . . . .	48
<b>6</b>	<b>References</b>	<b>49</b>
<b>7</b>	<b>Appendix</b>	<b>50</b>
7.1	Kaplan Meier graphs . . . . .	50
7.2	Log-minus-log survival plots . . . . .	57
7.3	Schoenfeld residual graphs . . . . .	63
7.4	Hazard rate graphs . . . . .	71
7.5	Code . . . . .	77

# 1 Introduction

## 1.1 Survival analysis in biostatistics

When developing and distributing medicine, monitoring any side effects caused by this medicine is important. To this end, two registers are of great importance. First, the Prescribed Drug Register, kept by the National Board of Health and Welfare is a register of prescription medicines dispensed in Sweden to Swedish citizens. Second, the National Patient Register keeps a record of diagnoses from Swedish inpatient and specialist outpatient care (care that did or did not require a stay at a hospital).

All Swedish citizens are assigned a unique personal identifier at birth or immigration which is unchanged through their life. This identifier can be used to link the data in the Prescribed Drug Register and the National Patient Register to create data sets relating to diagnoses and dispensed medicine. Researchers can therefore request data to study the side effects of dispensed drugs. Should a specific side effect be suspected then this data can be used to test statistically whether the medicine has a significant effect on that specific type of outcome.

An important factor in these statistical tests is that time is a factor in the outcome of side effects. Since different people are likely to suffer side effects at a different rate, it is important to take the time component into account when building a statistical model. While this might sound like a simple case of fitting the data to some chosen time distribution, there is a complication in the form of censored data, specifically right censored observations. An observation is right censored if we only know that the outcome did not take place before some point in time. Observations become right censored because the person in question moved away, died or most likely, because the observation period of the study ran out. In these cases we can no longer get information on when they suffered an outcome or if they ever would.

While it would be possible to fit a time distribution using right censored data, the variance of such an estimate would depend on the ratio of censored to uncensored observations. A large number of censored observations would likely cause a large variance in our estimator making such an estimator unreliable. In the data-set we have analyzed in this thesis, approximately 97% of the observations are censored. Trying to estimate a time distribution from this data-set would most likely be unreliable.

For data sets relating to time with right censored observations survival analysis is commonly used. While it is called “survival” analysis, it is used to model any type of time until event data. Some examples include time until illness, time until an object breaks or time until death. We will refer to the event of interest as “outcomes”. Depending on the area it is applied to, survival analysis and the concepts within it can go by different names even if their definitions are the same. For the example the “survival function” in survival analysis might be called the “reliability function” depending on the field of use.

Any right censored observation contains less information than a non-censored observation does but it still contains some relevant information. For that reason, the methods we utilize have to be able to take right censored observations into account. For example, when estimating the survival function (probability that an observation takes longer than time  $t$  to take place) the Kaplan-Meier estimator is used. The Kaplan-Meier estimator is designed to extract more information from censored observations than a naiver estimator would. This is especially helpful in cases where a large majority of observations are censored.

Since estimating the time distribution function for our data is unreliable we instead choose to model the hazard rate. The hazard rate is interpreted as the rate of outcomes per time unit at time  $t$ . The hazard rate will likely change over time. For example, we might expect the rate of objects breaking due to manufacturing errors to shrink as time passes while we expect the rate of people dying of natural causes to increase with age.

The hazard rate is defined as

$$h(t) = \lim_{h \rightarrow 0^+} \frac{P(t < T < t + h | T > t)}{h} = \frac{f(t)}{1 - F(t)}, t \geq 0,$$

but in practice it is easier to work with the equivalent form

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

where  $f(t)$  is the time until event probability density function and  $F(t)$  is the distribution function for  $f(t)$ . The first form describes the hazard rate better than the second but the second is easier to work with in practice. The hazard rate can also be defined for distributions that take negative values. However in such cases the interpretation of it is not quite as clear. In this thesis we will only deal with time distributions with positive values for  $t$ .

The hazard rate is more robust to right censored observations since even if we do not know whether someone suffered the outcome or not, we can still include them as an “individual at risk” up until the point they are censored. The number of individuals at risk is an important measure when estimating the hazard rate as just adding more people to a study would increase the number of outcomes even if the hazard rate was the same for every person. The more people currently at risk, the better the estimate of the hazard rate. This of course means that the estimate gets worse as time passes and more and more people are eliminated either because of censoring or because they suffered an outcome.

A disadvantage of the common hazard rate estimates is they are step functions that only change value at the time of an outcome. As such, it is not continuous as we likely expect the real hazard rate to be, as long as no catastrophic events take place. Since the hazard rate can theoretically take any shape as long as it is positive, trying to fit the hazard rate to a standard continuous function is difficult. That is where the Cox regression model is useful.

## 1.2 Cox regression

Cox regression is the most commonly used model in survival analysis and relies on the proportional hazards assumption. In practice, we expect that observations with different covariates would have different hazard rates. The proportional hazards assumption would have us assume that while their hazard rates would be different, they would be proportional to each other. All the hazard rates would therefore have the same shape, just different scales.

The hazard rate for an observation with covariate vector  $x$  is modeled as

$$h(t|x) = h_0(t)e^{\beta x}$$

Where  $h_0(t)$  is defined as the base hazard rate, that is the hazard rate for some value for  $x$  that we use as a reference point. This would mean that the hazard ratio, the hazard rate of one observation divided by the hazard rate of another, would simply be constant. This means that we can eliminate the hazard rate from the model by focusing on the hazard ratio instead. The Cox regression model focuses entirely on this estimate of the hazard ratio

$$\frac{h(t|x)}{h_0(t)} = \frac{h_0(t)e^{\beta x}}{h_0(t)} = e^{\beta x}$$

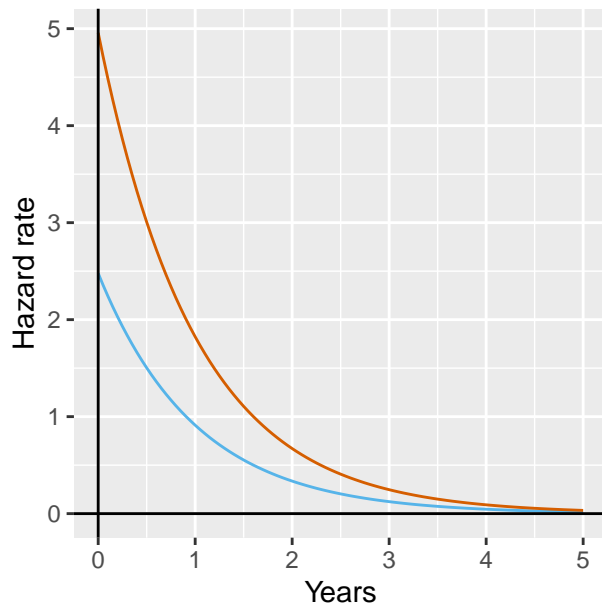
and lets us avoid many of the difficulties with trying to model the hazard rates directly.

Proportional hazards are easy to interpret. For example, one can use the hazard ratio to make statements such as “smokers suffer lung cancer at a rate of eight times that of non-smokers”. This ease of implementation and interpretability has made the Cox regression model a staple of survival analysis since its reveal in 1958. Proportional hazards are illustrated in Figure 1 along with an example of non-proportional hazards.

It is important to note that the Cox model is not the only proportional hazards model. The thing that signifies the Cox model is that we do not model the shape of the hazard rates. The Cox regression can be extended to model the hazard rates but there is no single, generally agreed upon method that this is done.



**Proportional hazards**



**Crossing hazards**

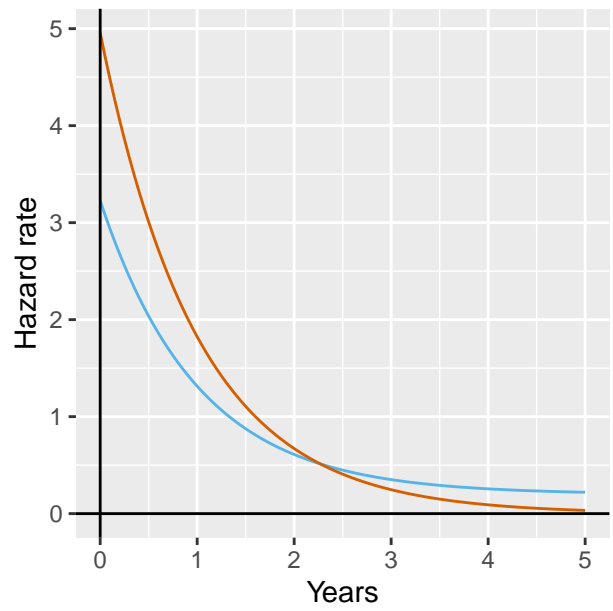


Figure 1: Figure illustrating potential proportional and non-proportional (crossing) hazards.

While the proportional hazards assumption might seem quite strict and to not hold up in general, it is still quite useful. While hazard rates might not always be exactly proportional to each other, in such cases the hazard ratio can be interpreted as the mean hazard ratio over the time interval. So, while smokers might not be eight times as likely to suffer lung cancer compared to non-smokers at all points in time, it is still true on average. Consequently, the conclusion would be the same, smoking is significantly correlated with lung cancer.

The worst kind of breach of the proportional hazards assumption is the case of “crossing hazards”. This happens when the hazard rates for two groups of observations cross each other at some point in time. If that happens then we end up in a case where one group fares better in the short term but worse in the long term. An example of this could be aggressive and conservative treatments of cancer. An aggressive treatment could for example, be surgery, which could lead to death during or after the surgery due to complications. However, if one survives the surgery, then it might have a better long-time effect than a more conservative treatment that is unlikely to kill you early on but will eventually fail. Crossing hazards are illustrated in Figure 1.

In the case of crossing hazards, the mean hazard ratio is likely to be small as the positive part and the negative part of the hazard ratio cancel each other out when you calculate the mean hazard ratio. Even if the effect of these crossing hazards is immense on the interpretation of the effect of the covariate on the hazard rate, a proportional hazards model is likely to just report a very small hazard ratio that might not be significantly different from zero if such a hypothesis test is applied. In these cases, it is less straightforward to say that one group fares strictly better than another and just reporting a hazard ratio with a proportional hazards model would be irresponsible.

The drawbacks of the Cox regression do not end there. Since the Cox method does not model the hazard rates by design, we do not get any direct information about their shapes. Knowing how the hazard rate changes with time is certainly interesting and worthy of being illustrated in the results of a model. Such results could be used to tell when the rate of outcomes is at its highest or when the danger has likely passed. In addition, since the model is non-parametric we are unable to estimate survival times for a group of observations among other things.

There are methods for mitigating crossing hazard curves using the Cox model like for example, dividing the time interval into several smaller intervals and applying a separate Cox model in each time interval. At that point you might start wondering if maybe a different type of model would be preferable rather than trying to force a Cox shaped model for every survival analysis data-set. There are many different variants on the Cox model, but we will narrow our view down to only the base Cox model. Variants of the Cox model are mentioned Section 5.3 of the thesis.

The motivation for this thesis was that in practice, survival analysis tends to default to using the Cox regression model. Perhaps there is another model that could be used instead that does not suffer from the same drawbacks that the Cox regression does. To compare, we have chosen to use the Flexible parametric model.

### 1.3 Flexible parametric model (FPM)

Unlike the Cox model, FPM does model the hazard rate for the observations. To avoid the problem of needing to assume a shape for the hazard rate, the FPM models the hazard rate using splines. The idea is to divide the interval into several smaller intervals and fit a third-degree polynomial curve within each interval. By using piece wise polynomial functions, any shape of the hazard function can be approximated, meaning that we do not need to make any assumptions that alter the results of the model. To avoid over-fitting some restrictions are applied to the piece-wise polynomials to make sure they are essentially continuous. The concept of splines are illustrated in Figure 2.

In addition, to modelling the hazard rates, FPM can also model interaction between the covariates and time. This interaction is one cause of non-proportional hazards and potentially crossing hazards, so by modelling it, we have a method of accounting for non-proportional hazards. The interaction is again modeled using splines as to avoid making assumptions about exactly how this interaction behaves.

FPM specifically models the logarithm of the cumulative hazard rate but this can then be used to model the hazard rates. The model is

$$\ln(H(t|x)) = s(\ln(t)) + \beta x + \sum_{j=1}^D s(\ln(t))x_j,$$

where the  $s$  functions are the splines,  $x_j$  are the covariates we model with interaction with time and  $D$  is the total number of these covariates. The fact that we need to specifically choose these covariates complicates the model. As such we would need some method for choosing these covariates. This is brought up in the next section.

Any variable that is chosen to not have an interaction with time is assumed to have proportional hazards. If no covariates are chosen to interact with time then we simply just get a proportional hazards model like in the Cox models case. However, unlike the Cox model case, we do model the hazard rates and have parametric estimates available for use. For that reason, the proportional hazards FPM model can be seen as the base Cox model but with the extra bonus of being able to visualize the hazard rates among other statistics.

Packages for the FPM are available in most statistical software.

## 1.4 Identifying non-proportional hazards

How do we know if our hazards are proportional or non-proportional? The results from a proportional hazards model will tell us that all hazards are proportional even if the original data does not support the assumption. There are tests that can be performed, either on the data or on the results of a model to test the assumption of proportional hazards. Some common tests include Kaplan-Meier curves, log-minus-log survival plots, Schoenfeld residuals and the Schoenfeld test. The first three methods are graphical tests that need to be judged visually and the last is a statistical hypothesis test. All four of these tests have some significant drawbacks. Visual tests tend to be subjective and prone to visual illusions. A hypothesis test is more objective but it tends to have low power. It also only tests for a certain type of non-proportional hazards.

Identifying non-proportional hazards is not easy but one could argue that it is not always necessary. As was stated earlier, the hazard ratio can in cases of non-proportional hazards be interpreted as a mean hazard ratio over the time interval. Therefore as long as we do not have an extreme case of non-proportional hazards like for example, crossing hazard rates then a hazard ratio is still a useful measure. If we do have an extreme case of non-proportional hazards, then such relations would likely be glaringly obvious when we apply graphical tests to the data. As such, the proportional hazards assumption could be considered “good enough” unless the counter evidence is obvious.

We have investigated the tests for non-proportional hazards in this thesis. We have also further discussed the problems with model selection when we want to account for non-proportional hazards in the discussion part of the thesis.

## 1.5 The dermatology data-set

To contextualize our simulations, we have a real-world data-set at our disposal. The data-set links information on drug exposure from the Prescribed Drug Register with information on diagnoses from the National Patient Register, specifically patients diagnosed with atopic dermatitis (AD), psoriasis or both. The outcome of interest is venous thromboembolism (VTE) and the covariate of specific interest is the use of the medicine methotrexate and/or several other “biological” medicines (a medicine whose active substance has been created by a living creature). The time frame is medicines dispensed between the years 2013 and 2022.

Janus kinase inhibitors have recently been introduced as a treatment for AD which is theorized to raise the risk of VTE. Studies on this association has yielded conflicting results and it is impossible to say whether

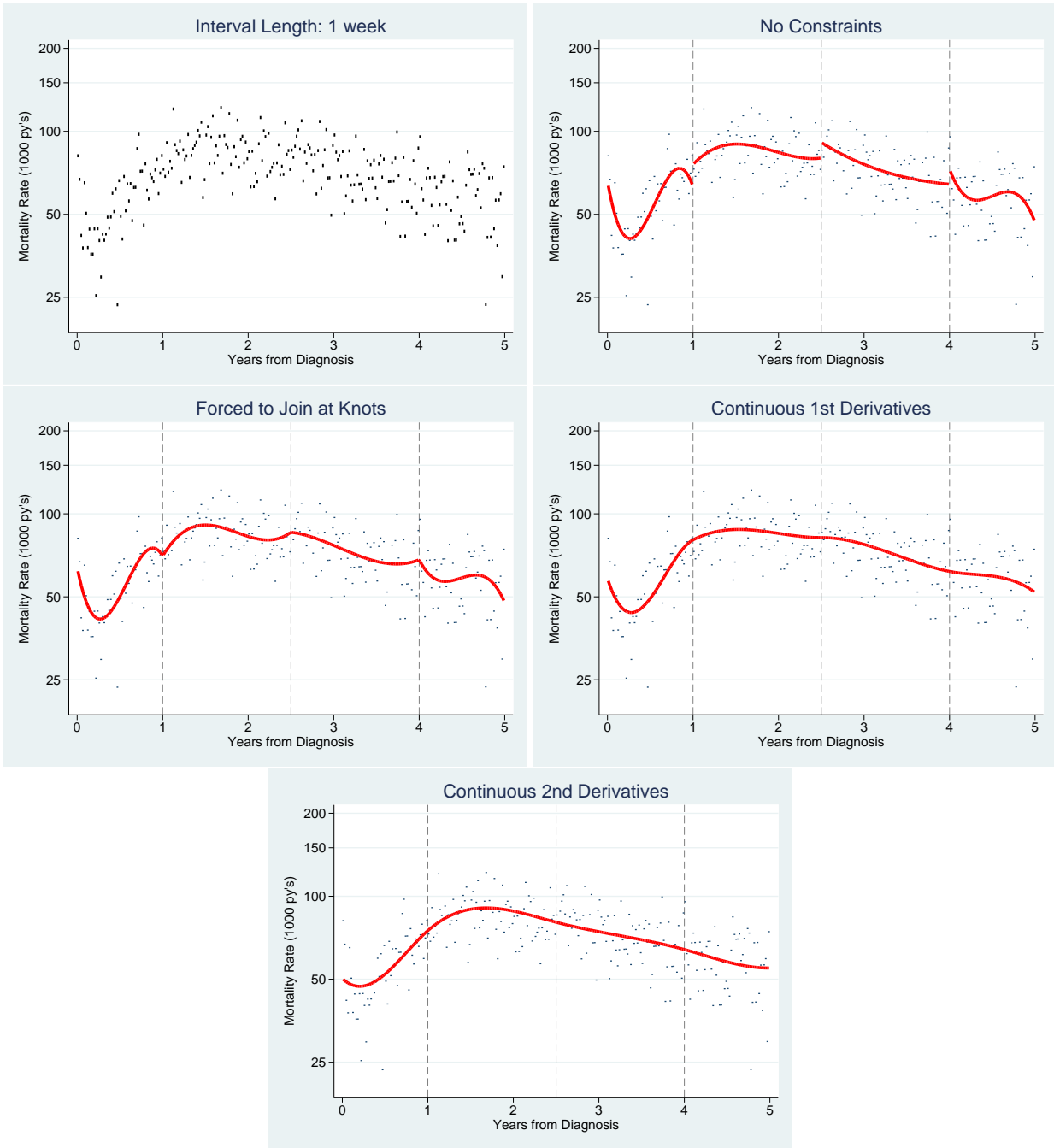


Figure 2: Images illustrating the restrictions we place on cubic splines, implemented one by one to show why they are implemented and what effect they have.

it is AD itself or the treatment that raises the risk of VTE. Therefore testing whether the two types of treatments for AD has a significant effect on the outcome of VTE is of great interest.

## 1.6 Plasmode simulation

To test how well the two models fit data with and without proportional hazards we chose to use simulation. Simulating data from a biological data-set can be tricky, especially if we want to preserve the underlying correlation between covariates. It seems likely that the age of the observed patient is going to have a large correlation with history of illnesses and treatments just because they have had a lot time to potentially suffer from these illnesses. The correlation between different illnesses and treatments is going to be subtler, maybe to such an extent that an entire model could be made to investigate just that.

in addition, we have the extra simulation step for censor times. While some censor times can be almost arbitrary, some of them, especially censoring due to death is also likely to be correlated with the other covariates. Optimally we would need to make some kind of assumptions about the relation between the covariates and the censored by death times.

The way we have chosen to deal with this is using plasmode simulation as described in Schreck et al. (2024). To avoid needing to make assumptions about the distribution of the covariates and censor times, we instead chose to sample the covariates directly from the original data. By sampling the covariates and censor times from the original data-set, the main thing we need to model is the survival times.

As the hazard rates in the original data-set was estimated to be close to constant, we chose to have the simulated survival times also have constant hazard rates. Simulating constant hazard rates is simple as simulating times from the exponential distribution will give us a constant hazard rate. For any two exponentially distributed random variables with two different rate parameters, the hazard rates will be proportional. For that reason, just simulating with exponentially distributed survival times will give us a data-set with proportional hazards.

To simulate non-proportional hazards we have chosen to use the piecewise exponential distribution method. The idea is to choose a midpoint in the interval. For any observation where an event is simulated to not take place before the midpoint, then the hazard rate changes to a higher or lower rate for after the midpoint. If we switch the rate of outcomes for the types of observations at the midpoint, then the hazard rates will theoretically cross each other, simulating non-proportional hazards. The simulated hazard rates are illustrated in Figure 3.

This is relatively intuitive when using the exponential distribution for the simulation as it is memoryless. The idea being that an exponential distribution is an exponential distribution at all points in time as long as an outcome has not yet taken place.

## 1.7 Structure of the thesis

In Section 2 we describe the data set on dermatology that is used to illustrate the results of the two models. In Section 3 we go through the theory for survival analysis and the two models. In Section 4 we show the results of the thesis. Section 5 contains discussion and conclusions drawn from literature and the results of the thesis.

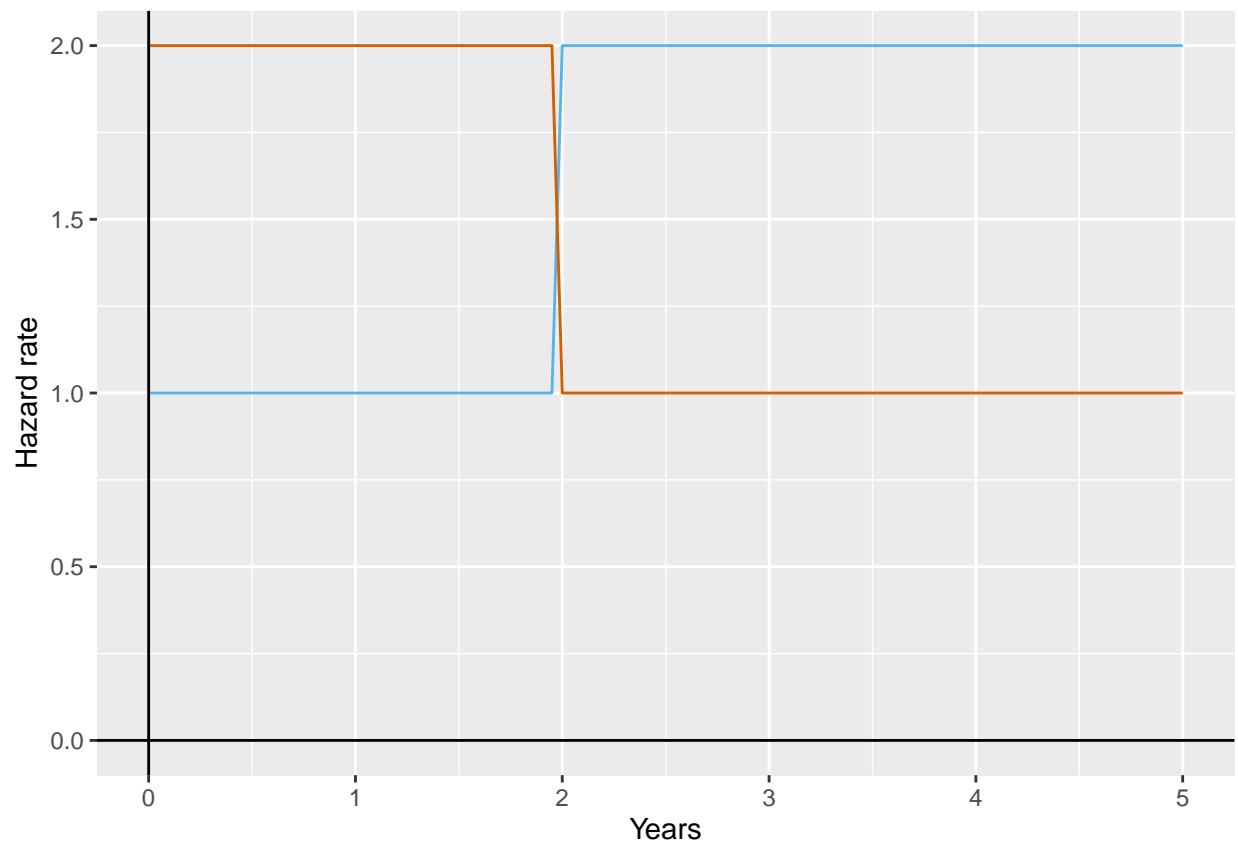


Figure 3: Illustrating a case of crossing hazards where the hazard rates switch values for the two treatments after 2 years has passed.

## 2 Data

### 2.1 Dermatology data-set on venous thromboembolism

The data-set is a linked data-set consisting of data from the Swedish prescribed drug register and the Swedish national patient register. These registers contain information about the dispensation of medicine to Swedish citizens and diagnoses of Swedish citizens respectively. Our data-set contains individuals who suffer from atopic dermatitis (AD), psoriasis or both and have filled prescriptions for the medicine methotrexate or some specified types of biological medicine (a medicine whose active substance has been created by a living creature) during the period between 01-01-2013 and 31-12-2022. All individuals are “new users” meaning that they have not filled prescriptions for that specific treatment for at least one year prior to filling a new prescription.

The data-set consists of 48 591 observations. The information containing dates consists of 4 different types of times.

The first time VTE is the time at which the patient experienced venous thromboembolism (VTE) or NA if this event did not take place. VTE includes the two medical conditions deep vein thrombosis and pulmonary embolism. Deep vein thrombosis is when a blood clot forms in a deep vein (a vein not close to the surface of the skin). Pulmonary embolism is when the arteries that transport blood to your lungs are obstructed.

The last three are what we define as the censor times. Those being the time a person died, emigrated or the study ended. Death or emigration times are NA if they did not take place. All patients have an end of study time, even those who emigrated, died or had VTE. The end of study time is always 31 December 2022, which is the last day in the range of dates that was taken from the registry.

For these three types of censor times, if no case of VTE was observed before the earliest of these times then the observation is considered censored at that point in time.

In total, 3 305 individuals have used both methotrexate and a biological medicine during the period in question. Such observations have been split into two separate observations with the same covariates except for having different “Treatment” covariates and index times. After the split we have 51 895 observations.

Descriptions of the covariates in the data-set can be found in Table 1.

### 2.2 Visualizing the data-set

In this section we visualize statistics relating to the times and covariates. Extra focus was put on visualizing how the statistics vary between age groups as age tends to be very correlated with several other covariates and statistics.

It is important to keep in mind that the times do not run in parallel. Each outcome time is relative to that observations time zero which is the time they filled a prescription for their treatment.

Figure 5 show that the number of outcomes of VTE, death and emigration decrease over time. This is to be expected, as over time, the number of people who are at risk decrease, either due to suffering outcomes of VTE or from being censored. When the number of people at risk decrease, so does also the number of outcomes/censorings even if the rate of outcomes/censorings is the same for the entire interval. As an example, you would get more heads when flipping 100 coins than you would when flipping only 10, even if the probability of heads is 50% in both cases.

Figure 5 also shows the end of time dates for each observation. The events in this graph can be read as happening simultaneously. Any observation that has an end of time date equal to one year, joined the study one year before it ended. The figure shows that the end of study times behave relatively consistent over time, though maybe waning over time. There appears to be some sort of pattern in the highs and lows of the distribution.

Table 4 shows that age is likely to be correlated with several of the other covariates.

Table 1: Table containing descriptions of the covariates in the dermatology data set.

Covariate	Description
<b>ageGroup</b>	The persons age group. The original data did not define any age groups. We have chosen to use the age groups 0-44, 45-59, 60-74 and 75+.
<b>Treatment</b>	The type of medicine that the patient filled a prescription for. Either "Methotrexate" or "Biological".
<b>AD</b>	History of atopical dermatitis, also known as eczema, a chronic disease that causes inflammation/redness/irritation of the skin
<b>AD_Drugs</b>	History of drugs for the aforementioned atopical dermatitis
<b>Anticoagulant</b>	History of anticoagulant drugs which makes your blood less likely to clot.
<b>Asthma</b>	History of asthma which causes inflammation in the throat which in turn can cause coughing, wheezing and shortness of breath.
<b>Atrial fibrillation</b>	History of atrial fibrillation which is irregular heart beats.
<b>CD</b>	History of Crohns disease, a disease that causes inflammation of the bowel.
<b>COPD</b>	History of chronical obstructive pulmonary disease which is a chronic inflammatory lung disease that obstructs airflow from the lungs.
<b>Cancer</b>	History of cancer.
<b>Diabetes</b>	History of diabetes.
<b>fractures</b>	History of bone fractures.
<b>heart failure</b>	History of heart failure.
<b>History BIOLOGIC</b>	History of biological medicine before enrolling in the trial.
<b>History Methotrexate</b>	History of being treated with methotrexate before enrolling in the trial.
<b>Hyperlipidemia</b>	History of hyperlipidemia (High blood fat).
<b>Hypertension</b>	History of hypertension (High blood pressure).
<b>MACE</b>	History of MACE (major adverse cardiovascular events) in which medicine causes major problems with the cardiovascular system (Heart and blood)
<b>Melanoma skin cancer</b>	History of skin cancer with melanoma, that being cancer that starts in the melanocytes that create the pigment that gives the skin its color.
<b>Obesity</b>	History of obesity.
<b>PSO</b>	History of psoriasis which is signified by an overactive immune system that causes skin cells to multiply at an overly fast rate.
<b>RA</b>	History of reumatoid arthritis which is an autoimmune disease which causes your immune system to attack itself and causing inflammation.
<b>Rhinitis</b>	History of rhinitis which is a reaction in your nose which leads to congestion, runny nose, sneezing and itching.
<b>Skin cancer</b>	History of skin cancer without the aforementioned melanoma.
<b>Stroke</b>	History of stroke which is caused by cell death in the brain due to poor blood flow.
<b>UC</b>	History of ulcerous cholitis which is a chronic condition where the large intestine and rectum become inflamed.
<b>Uveitis</b>	History of uveitis which is an inflammation in the uvea of the eye.
<b>VTE</b>	History of venous tromboembolism. Venous thromboembolism is either a blood clot that is not close to the surface of the skin or in the arteries connected to the lungs.
<b>sex</b>	Sex of patient, 1 for male, 2 for female.



Figure 4 Shows that ages resemble a normal distribution but skewed slightly to the older side. Figure 4 also Shows that the age group 75+ has a relatively low representation in the data-set. However, the 75+ age group has a very large effect on the outcomes of the regression models due to their relatively high rate of outcomes of VTE. The boxplot in Figure 4 shows that we do have many outliers in the 75+ age group. Those individuals are likely to have more extreme rates of VTE or death than the other 75+ aged people.

Table 3 shows that the older age groups have filled prescriptions for biological medicine relatively less often than the younger age groups. This could be because doctors are less likely to prescribe biological medicine to the elderly due to health concerns and prefer to instead prescribe the safer methotrexate.

Table 5 shows that people in the 75+ age group suffer VTE and death relatively more frequently than the younger age groups. Higher rates of death are to be expected as this age group is close to the age where we would start to expect natural causes of death to set in. Most importantly, people in the age group 75+ suffer outcomes of VTE about six times more often than people aged 0-44.

The information in Table 3 and Table 5 might help explain the results of the regression models. The Cox model estimated that observations using methotrexate suffered outcomes of VTE at approximately 1.27 the rate as those on biological medicines. This is the opposite of what you might expect since methotrexate is considered the safer of the two types of treatments. However, the fact that people aged 75+ suffer VTE at a much higher rate than younger people and that they also get prescribed methotrexate relatively more often than younger age groups could be causing the estimate that methotrexate is more dangerous than biological medicines. A way to improve the model would be to take the relative frequency of which medicines are prescribed to which age group into account and weigh the outcomes correspondingly.

For every type of graph we have created one for each covariate. Only the graphs for treatment, age group, sex and history of VTE have been included in the results. Treatment is included as it is the main covariate of interest. Sex and age group are demographic covariates and should always be included. According to the Schoenfeld test, history for VTE had non-proportional hazards at a 5% significance level. As such, showing how this covariate appears in graphs is of interest. As for the graphs of the remaining covariates, they are all included in the appendix.

Table 2: Table of means

Covariate	Mean value
age	50.98
AD	12.5 %
femaleRatio	51.4 %
PSO	81.1 %
VTE	3 %
Hypertension	19.1 %
MACE	4.7 %
COPD	2.6 %
Fractures	23.6 %
Skin_cancer	2 %
Melanoma_skin_cancer	1 %
Cancer	5.6 %
Asthma	8.5 %
CD	3.6 %
Diabetes	7.7 %
Obesity	6.7 %
RA	6.4 %
Rhinitis	1.7 %
UC	3.2 %
Uveit	7.1 %
Atrial_fibrillation	4.1 %
Stroke	0.2 %

Covariate	Mean value
Heart_failure	2.5 %
Hyperlipidemia	5.9 %
Anticoagulant	89.7 %
AD_Drugs	91.1 %
History_Methotrexate	88.5 %
History_BIOLOGIC	19.2 %

Table 3: Treatment by agegroup

Treatment	n	0-44	45-59	60-74	75+
Biological	40517	29 %	22 %	15 %	10 %
Methotrexate	11378	71 %	78 %	85 %	90 %

Table 4: Relative frequency by age group table rounded to two decimal places

covariate	0-44	45-59	60-74	75+
AD	21 %	8.71 %	7.28 %	8.44 %
sex	48.99 %	51.16 %	53.03 %	57.45 %
PSO	72.94 %	83.93 %	86.77 %	86.44 %
VTE	1.11 %	2.66 %	4.52 %	7.18 %
Hypertension	2.08 %	14.29 %	35.36 %	54.48 %
MACE	0.36 %	2.8 %	9.08 %	15.84 %
COPD	0.17 %	1.11 %	5.48 %	8.91 %
Fractures	21.1 %	20.32 %	26.83 %	35.35 %
Skin_cancer	0.33 %	1.28 %	3.15 %	8.44 %
Melanoma_skin_cancer	0.26 %	0.94 %	1.33 %	2.75 %
Cancer	2.49 %	3.68 %	9.06 %	14.23 %
Asthma	11.55 %	6.86 %	6.74 %	8.14 %
CD	5.87 %	2.8 %	2.28 %	1.44 %
Diabetes	2.03 %	6.08 %	13.68 %	18.32 %
Obesity	6.84 %	7.4 %	6.62 %	4.31 %
RA	3.62 %	6 %	8.69 %	12.52 %
Rhinitis	1.49 %	1.52 %	1.95 %	2.15 %
UC	4.42 %	2.8 %	2.34 %	1.96 %
Uveit	3.42 %	4.34 %	9.29 %	25.72 %
Atrial_fibrillation	0.32 %	1.8 %	7.29 %	17.97 %
Stroke	0.02 %	0.07 %	0.32 %	0.79 %
Heart_failure	0.17 %	1.2 %	4.2 %	11.09 %
Hyperlipidemia	0.63 %	4.13 %	11.59 %	16.34 %
Anticoagulant	87.73 %	88.59 %	91.93 %	94.7 %
AD_Drugs	87.05 %	91.72 %	94.22 %	95.97 %
History_Methotrexate	82.05 %	90.08 %	92.99 %	95.5 %
History_BIOLOGIC	24.96 %	19.34 %	14.51 %	9.73 %

Table 5: Proportion of time type per age group

ageGroup	n	proportion_VTE	proportion_emigration	proportion_death	proportion_eos
0-44	16712	0.98 %	1.04 %	0.38 %	97.59 %
45-59	14874	2.26 %	0.54 %	1.71 %	95.48 %
60-74	12964	4.5 %	0.41 %	7.64 %	87.46 %
75+	4040	6.34 %	0.07 %	22.97 %	70.62 %

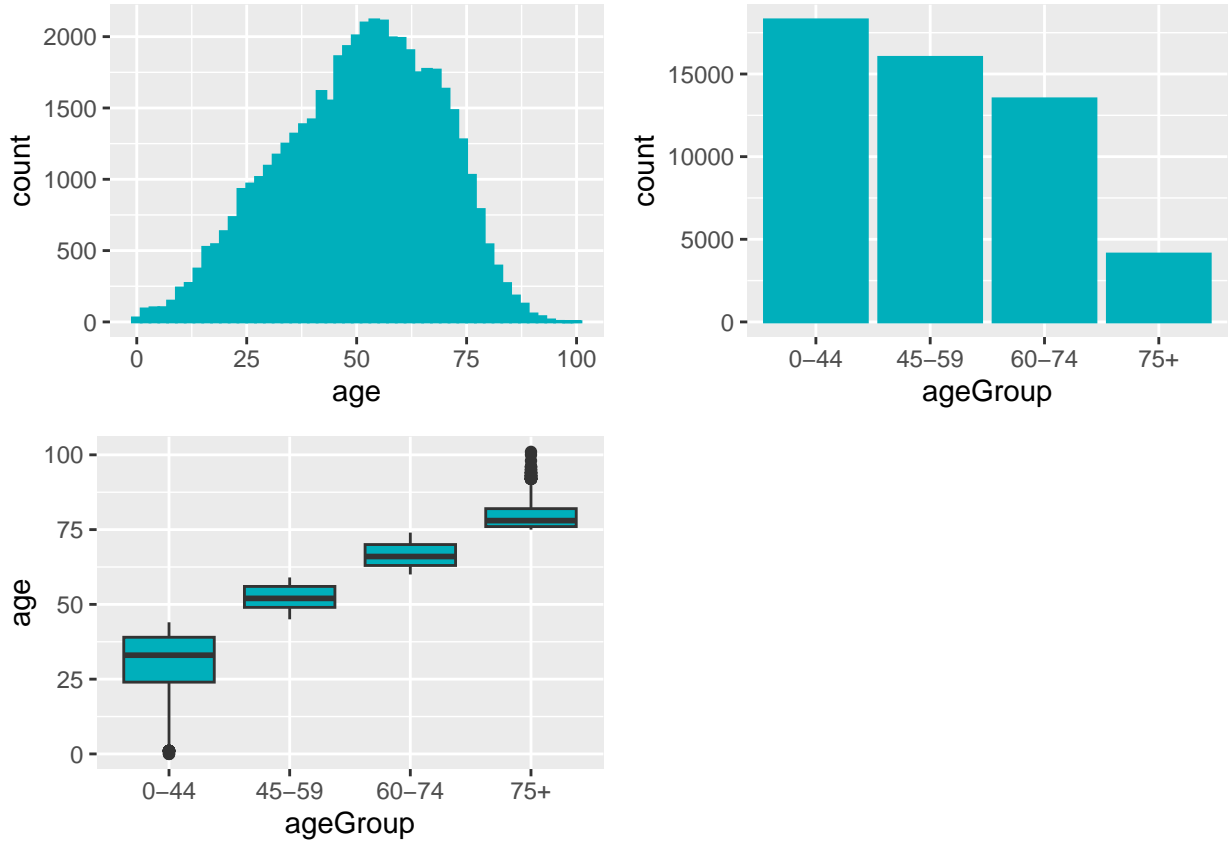


Figure 4: Figures illustrating the distribution of ages in the dermatology data set.

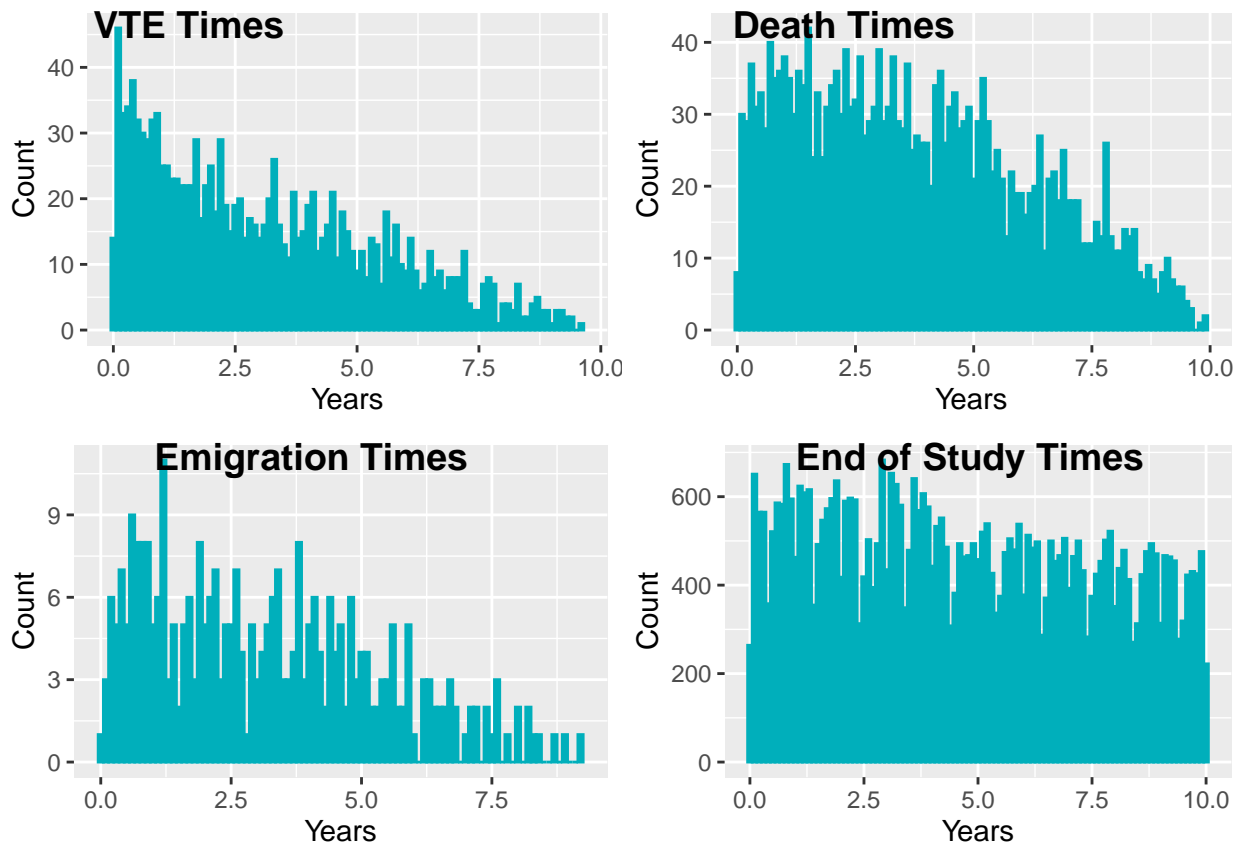


Figure 5: Figures illustrating the distribution of the 4 types of time in the dermatology data set. The times are either outcome of VTE or censoring due to either death, emigration or end of study.

## 3 Theory

### 3.1 Survival functions, hazard rates and hazard ratios.

In the following section we use  $T$  to refer to a random variable with a positive time as support.

The survival function is defined as  $S(t) = P(T > t)$ . In words, the probability that the outcome takes place after time  $t$ . The survival function has different names depending on the area it is used in. For example, it is sometimes referred to as the reliability function when modelling how long a product lasts before breaking. In the case of the data-set we are using, an outcome is defined as an occurrence of venous thromboembolism (VTE) in an observation. Let  $f(t)$  be the probability density function for  $T$ . Then

$$S(t) = \int_t^\infty f(t)dt = \int_0^t f(t)dt + \int_t^\infty f(t)dt - \int_0^t f(t)dt = 1 - F(t),$$

where  $F(t)$  is the cumulative density function of  $f(t)$ . The survival function is a continuous, monotonically decreasing function that takes values between 1 and 0.

The hazard function  $h(t)$  is defined as the rate of outcomes at time  $t$ . If  $h(t_0) = \lambda$  then at time  $t_0$  outcomes of VTE happen at a rate of  $\lambda$  per time unit.

The hazard rate is defined as

$$h(t) = \lim_{h \rightarrow 0^+} \frac{P(t < T < t + h | T > t)}{h}, \quad t \geq 0,$$

where the formula inside the limit can be seen as an approximation of the rate of outcomes at time  $t$  per time unit. When you then take the limit, the result is the instantaneous rate of outcomes at time  $t$  if that limit exists.

Using the definition to calculate the hazard rate is cumbersome. Another equivalent expression of the hazard rate is

$$h(t) = \frac{f(t)}{S(t)}. \tag{1}$$

The two expressions are equivalent as

$$\begin{aligned} h(t) &= \lim_{h \rightarrow 0} \frac{P(t < T < t + h | T > t)}{h} = \lim_{h \rightarrow 0} \frac{P(T > t | t < T < t + h) P(t < T < t + h)}{h P(T > t)} = \\ &= \lim_{h \rightarrow 0} \frac{P(t < T < t + h)}{h S(t)} = \frac{f(t)}{S(t)}, \end{aligned}$$

where the first equality follows from Bayes theorem and the second from the interpretation of the probability density function as the limit of a probability function. From these formulas we can see that the hazard rate is always positive.

For two observations with covariate vector  $x_1$  and  $x_2$ , we define their hazard rates as  $h(t|x_1)$  and  $h(t|x_2)$  respectively. These two hazard rates will likely differ if  $x_1 \neq x_2$ . We define their hazard ratio as

$$\frac{h(t|x_1)}{h(t|x_2)}$$

and interpret it as how much higher or lower  $h(t|x_1)$  is compared to  $h(t|x_2)$  at time  $t$ .

Usually the hazard function in the denominator is the base hazard rate  $h(t|x = \mathbf{0}) = h_0(t)$ , which is interpreted as a observation with the zero vector as covariate vector. Such an observation might not even exist in the data-set but is convenient to use as the base case when we want to model how the change in value of one covariate affects the hazard rate at time  $t$ .

Under the proportional hazards assumption, for every  $x$  there exists a positive value  $c$  so that  $h(t|x) = ch_0(t)$ . Under this assumption

$$\frac{h(t|x_1)}{h(t|x_2)} = \frac{c_1 h_0(t)}{c_2 h_0(t)} = \frac{c_1}{c_2}$$

which does not depend on  $t$ . In both the Cox regression model and the proportional hazards version of the flexible parametric model (FPM) we model this value as  $c = e^{\beta x}$  where  $\beta$  is a vector of parameters.

While the hazard rate is the more intuitive measure, in practice it is common to use the cumulative hazard function. It is defined as the hazard function as

$$H(t) = \int_0^t h(s) ds$$

and interpreted as the hazard that has accumulated up until time  $t$ .

The hazard rate can be obtained from the survival function via

$$h(t) = -\frac{d}{dt} \ln(S(t)),$$

since

$$-\frac{d}{dt} \ln(S(t)) = -\frac{\frac{d}{dt} S(t)}{S(t)} = -\frac{\frac{d}{dt}(1 - F(t))}{S(t)} = \frac{f(t)}{S(t)} = h(t),$$

where we have used the chain rule for differentiation and the definition of the survival function.

Another important relation is

$$S(t) = e^{-H(t)},$$

which is shown by integrating the previous relation from 0 to  $t$ :

$$\begin{aligned} -\int_0^t \frac{d}{ds} \ln(S(s)) ds &= \int_0^t h(s) ds \\ -\ln(S(t)) &= H(t) \end{aligned} \tag{2}$$

$$S(t) = e^{-H(t)}.$$

Under the proportional hazards assumption this becomes

$$S(t) = \exp\left(\int_0^t h(s) ds\right) = \exp\left(\int_0^t h_0(s) e^{\beta x} ds\right) = \exp(H_0(t) e^{\beta x}) = \exp(H_0(t))^{e^{\beta x}} = S_0(t)^{e^{\beta x}} \tag{3}$$

where  $H_0(t)$  is the cumulative base hazard function.

### 3.1.1 Kaplan-Meier estimate of the survival function.

The survival function  $S(t) = P(T > t)$  gives the probability that the outcome takes place after time  $t$ . Given a data-set with no censored times, the survival function is estimated using the empirical distribution. When working with right censored times, it is a bit more complicated. Since we do not know exactly how long a censored observation survived for, we lose information that is necessary to estimate the survival function. The larger the ratio of censored to uncensored observations is, the more difficult it becomes to estimate the survival function. Consequently, we need a specialized estimator to estimate the survival function when censored observations are present.

The Kaplan-Meier estimator, defined in Kaplan and Meier (1958) is commonly used to estimate the survival function in survival analysis. This estimator is a non-parametric method specifically designed to handle right censored observations. The Kaplan-Meier estimator of the survival function at time  $t$  is defined as

$$\hat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i}\right),$$

where  $t_i$  is a time one or more events happened,  $d_i$  is the number of events that have happened up until time  $t_i$  and  $n_i$  is the individuals known to have survived up to time  $t_i$ .

The estimator yields a non-continuous step function since the formula operates on a discrete time scale. This function is usually called the Kaplan-Meier curve and is a good visual aid to assess non-proportional hazards. Should for example, the curves for two different treatments cross each other then we expect the hazards to be non-proportional.

## 3.2 Cox regression

The Cox regression model was first introduced in the article by Cox (1972). The main assumption of the Cox model is the proportional hazards assumption which is that the hazard rate for any covariant vector  $x$  is proportional to the base hazard rate  $h_0(t)$ . We model this assumption as

$$h(t; \beta|x) = h_0(t)e^{\beta x}$$

where  $\beta$  is the regression coefficient vector for our model. The factor  $e^{\beta x}$  is interpreted as the proportionality constant which is always positive, as the hazard rate is by definition. If we were to calculate the hazard ratio for this hazard rate with the base hazard rate we get

$$\frac{h(t; \beta|x)}{h_0(t)} = \frac{h_0(t)e^{\beta x}}{h_0(t)} = e^{\beta x}$$

where the base hazard rates cancel in the ratio. By doing this we eliminate the time variable from the model and entirely focus on the proportionality constant for any covariate vector. The advantage of the model is that we do not need to make any inference about the base hazard rate.

Let  $\beta_i$  be element  $i$  of the regression coefficient vector. The value of  $e^{\beta_i}$  can be interpreted as the change in hazard rate as the corresponding covariate is increased by one unit. This interpretation allows statements like that smokers get lung cancer at 8 times the rate of non-smokers if  $e^{\beta_{smoker}} = 8$ .

For the likelihood function we consider a partial likelihood where we consider the base hazard rate to be a nuisance parameter and  $\beta$  to be the parameter of interest. The likelihood function for the outcome time of observation  $i$  to happen at time  $T_i$  is

$$L_i(\beta) = \frac{h(T_i|X_i)}{\sum_{j:T_j > T_i} h(T_j|X_j)} = \frac{h_0(T_i)e^{\beta x_i}}{\sum_{j:T_j \geq T_i} h_0(T_j)e^{\beta x_j}} = \frac{e^{\beta x_i}}{\sum_{j:T_j \geq T_i} e^{\beta x_j}},$$

where the nuisance hazard rates cancel in the ratio due to the proportional hazards assumption. This leaves only the main parameter of interest  $\beta$  to be estimated. By eliminating the hazard rates, time only enters the likelihood through the denominator as the set of outcomes that took place after observation  $i$  and observation  $i$  itself.

A way to interpret this likelihood is that it increases whenever observations that take place late in the time interval have small hazard ratios which correspond to longer estimated times.

We can then form the complete likelihood function by taking the product of this over all observations:

$$L(\beta) = \prod_{i=1}^n L_i(\beta),$$

and then we maximize it to get our estimate of  $\beta$ . There is no closed form solution for this estimate so instead it is usually optimized numerically using for example, the Newton-Raphson algorithm.

### 3.3 Flexible parametric model

The flexible parametric model (FPM) is defined in article by Royston and Parmar (2002). The main idea is to deal with the problem of the hazard rates possibly following any arbitrary curve by approximating them using spline curves. This way the model should be able to deal with arbitrary hazard rates while at the same time being able to visualize the approximation.

#### 3.3.1 Natural splines

Polynomials are quite flexible when modelling arbitrary functions as they can take many different shapes, many more than other standard mathematical functions can. However, trying to approximate longer, more complicated function using polynomials would require a polynomial of a very large degree which is difficult to model and has come with the problems of potentially overfitting the data. Cubic splines solve this problem by dividing any interval we want to approximate a function for into several smaller intervals and fitting a cubic polynomial in each one. To make sure these piecewise polynomial function create a smooth continuous function, we enforce restrictions on them so that at any point these functions meet, they have the same value, first and second derivative.

Figure 6 illustrate the idea of fitting a cubic spline to data points by enforcing the restrictions on the splines one by one.

The formula for the pieces of a cubic spline is

$$S_j(t) = a_j + b_j(t - t_j) + c_j(t - t_j)^2 + d_j(t - t_j)^3.$$

We define a list of coordinates  $[(t_0, y_0), (t_1, y_1), \dots, (t_n, y_n)]$ ,

where  $t_i$  are the locations of the knots and  $y_i$  is the fitted value at the knot. The package we chose to model FPM places the knots at the evenly spaced percentiles of observed outcomes. For example, with 5 knots, we would have  $t_0$  at 0%,  $t_1$  at 25%,  $t_2$  at 50%,  $t_3$  at 75% and  $t_4$  at the 100% percentile.

We need the piecewise polynomial functions to fulfill all the following restrictions:

The leftmost piece equals the fitted value  $y_0$  at time 0:

$$S_0(t_0) = y_0.$$

The rightmost piece equals the fitted value  $y_n$  at the end of the time interval:



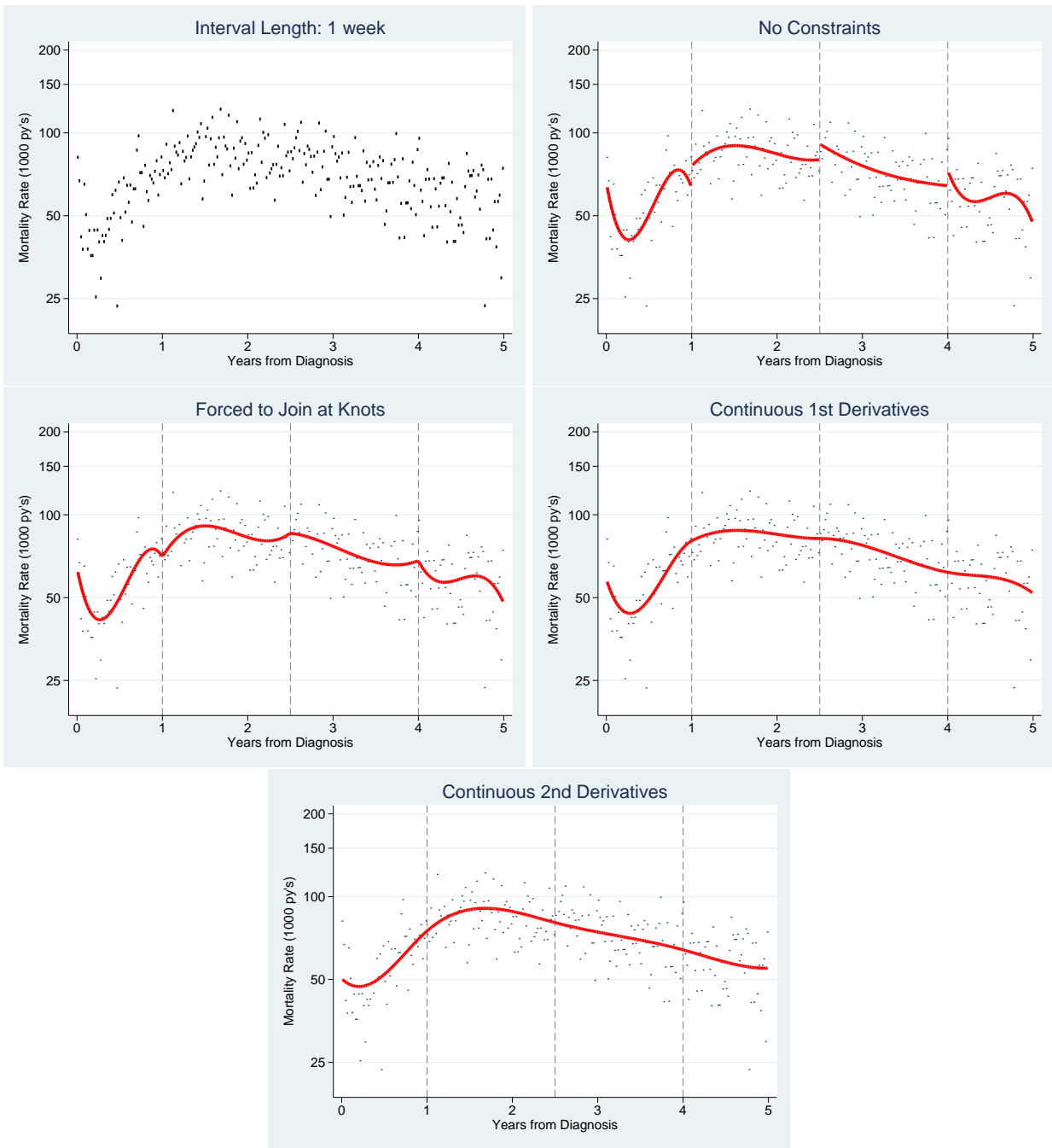


Figure 6: Images illustrating the restrictions we place on cubic splines, implemented one by one to show why they are implemented and what effect they have.

$$S_{n-1}(t_n) = y_n.$$

Each neighboring piece must have equal values at their bordering knot:

$$S_i(t_i) = y_i = S_{i-1}(t_i), \quad i = 1, 2, \dots, n - 1.$$

Each neighboring piece must have equal first derivatives at their bordering knots:

$$S'_i(t_i) = S'_{i-1}(t_i).$$

Each neighboring piece must have equal second derivatives at their bordering knots:

$$S''_i(t_i) = S''_{i-1}(t_i).$$

Finally for natural splines the leftmost and rightmost piece have second derivative 0 at time 0 and time  $t_n$  respectively:

$$S''_0(t_0) = S''_{n-1}(t_n) = 0.$$

which means that the splines are linear outside the interval.

### 3.3.2 Modelling the hazard rate using splines

It was shown in Equation 2 in Chapter 4.1 that

$$S(t) = S_0(t)e^{\beta x},$$

which in combination with Equation 2 also in Chapter 4.1 lets us show that

$$\ln(-\ln(S(t))) = \ln(H(t)) = \ln(H_0(t)) + \beta x.$$

When modelling the base hazard rate in the FPM, we are specifically approximating  $\ln(H_0(t))$  using a spline function. The approximation becomes

$$\ln(H(t|x_i)) = s(\ln(t)|\gamma, k_0) + \beta x_i$$

where  $s$  is the spline function, the  $\gamma$  variables are the parameters for the splines and  $K$  is the number of knots.

Furthermore if we want to model interaction with time we add a second spline to be fitted for the specified covariates,

$$\ln(H_i(t|x_i)) = \eta_i = s(\ln(t)|\gamma, k_0) + \beta x_i + \sum_{j=1}^D s(\ln t|\gamma_j, k_j)x_{ij},$$

where  $j$  is the index for the covariates we want to model as time dependent and  $D$  is the total number of those covariates. The base hazard is always modeled in the flexible parametric model. The number of covariates we want to be modeled using splines needs to be defined beforehand. If no covariates are selected

then the model becomes a proportional hazards model. Then the biggest difference between the flexible parametric model and the Cox model is that we have modeled the base hazard function.

We have used the `rstpm2` implementation of FPM for our models as defined in Lambert and Royston (2009). The log likelihood for this model is defined as

$$\ln L_i = d_i(\ln[s'\{\ln(t_i)|\gamma, k_0\}]) + \eta_i - e^{\eta_i},$$

where  $d_i$  is the indicator function for an uncensored observation and  $\eta$  was defined in the previous equation. This likelihood is then optimized using numerical methods.

## 3.4 Tests for non-proportional hazards

### 3.4.1 Log-minus-log survival plots

Log-minus-log survival plots are a method of testing the proportional hazards assumption. As was shown in Equation 2 of Section 3.1:

$$S(t) = S_0(t)e^{\beta x}.$$

If we then take “log-minus-log” of this we get that

$$\ln(-\ln(S(t))) = \beta x + \ln(-\ln S_0(t)) = \beta x + \ln(H_0(t))$$

where we use  $-\ln$  since  $S(t)$  only returns values between 0 and 1 which means  $\ln(S(t))$  is negative and as such, we can't use it as an argument for  $\ln$  again since the natural logarithm is only defined for positive values. The last equality follows from equation 2 in Section 3.1.

It is worth mentioning that after this transformation, any groups log-minus-log survival curve will differ from the base log-minus-log survival curve by the additive constant  $\beta x$ . Consequently, these two curves are parallel. If they are not parallel then the proportional hazards assumption does not hold.

While we do not know  $S(t)$ , we do have the Kaplan-Meier estimate of the survival function. If we insert that estimate into the log-minus-log transformation we can judge how well the proportional hazards assumption holds by how close to parallel the curves are.

### 3.4.2 Schoenfeld residuals

For our definition of the Schoenfeld residuals we use the one defined by Lin and Zelterman (2002). We define

$$\bar{x}(\hat{\beta}, t) = \frac{\sum_{T_j \geq T_i} e^{\hat{\beta} x} x_i(t)}{\sum_{T_j \geq T_i} e^{\hat{\beta} x}}$$

as the observed weighted mean of  $x$  over all observations that are currently at risk at time  $t$ . The Schoenfeld residual for observation  $k$  at time  $t_k$  is defined as

$$s_k = x_k(t) - \bar{x}(\hat{\beta}, t_k)$$

and is interpreted as how much the covariate value for the outcome at time  $t$  deviates from the estimated weighted mean for all observations currently at risk. If this measure is consistently high or low, then we expect the hazard to be higher or lower at that time than what has been estimated by the model. If the hazards are actually proportional we expect the Schoenfeld residuals to be close to constant.

By dividing the Schoenfeld residuals by their variance we can perform a visual test for proportional hazards. If we smooth the Schoenfeld residuals across the time interval we get an estimate of areas where each  $\beta_i$  is larger or lower than expected. We can then judge visually if the smoothed curve stays close to horizontal. If it does not, then the assumption of proportional hazards is questionable.

The formula for the variance of the Schoenfeld residuals is usually left out of most references as it is quite complicated, but it is available in Grambsch and Therneau (1994) even if the formulation of the Schoenfeld residuals is more general than the one we have used here.

### 3.4.3 Schoenfeld test

As stated by Grambsch and Therneau (1994), the Schoenfeld residuals can be fitted to a standard linear model, that is  $s = a + \beta b$ . This lets us make an asymptotic  $\chi^2$  test for  $b = 0$  for all  $\beta$ . This corresponds to a hypothesis test with null hypothesis of zero slope. If the test is significant for some  $\beta_i$ , then the effect of the corresponding covariate is assumed to not be constant and instead changes over time which contradicts the proportional hazards assumption.

## 3.5 Simulation

The purpose of the simulation is to compare the results of the Cox model and the flexible parametric model (FPM) when applied to data sets simulated to have exactly proportional hazards or non-proportional hazards, specifically the case of crossing hazard rates.

### 3.5.1 Plasmode simulation

A challenge when simulating our data sets is that we do not know the underlying distribution of the covariates. Certain covariates are going to be correlated and just simulating them naively will introduce bias in our model. For example, a lot of medical conditions are highly correlated with age. The simplest solution to the problem is using plasmode simulation. The idea is that we sample the covariates from the ones already in our data-set.

For each sample in our simulation data-set we sample one observation from our original data-set and use the covariates of that observation to simulate a new time using our chosen simulation method. We repeat this until we have a data-set of requested size. The size of the data-set has to be chosen with care as we need the data-set to be large enough for our models to work but also not too large to create redundancy. We also need the number of simulate data sets  $N$  to be large enough for the results to converge.

Schreck et al. (2024) recommend  $m$  out of  $n$  bootstrapping to create the plasmode data-set, where  $m$  is optimally chosen using the algorithm described in Bickel and Sakov (2008). However, this algorithm can take an unreasonably long time on a very large data-set. Therefore if the algorithm is deemed unfeasible, just choosing  $m = 0.9n$  is considered “good enough” and has seen plenty of use in practice, even if this method is not entirely justified.

As for the censoring time we could simulate them ourselves. However, that would require an assumption about the distribution of the censoring times. To avoid such an assumption we choose to sample the censoring times in addition, to the covariates when performing the plasmode simulation. If a person in our original data-set was censored at time  $t$  then they are censored at time  $t$  in the new data-set as well, unless of course they were simulated to have an outcome before that time.

To make the simulation simpler to perform and interpret, we have chosen to reduce the number of covariates. We decided to only use treatment as it is the main covariate of interest and age group, which has a significant effect on the outcome of venous thromboembolism (VTE).

The article “Statistical Plasmode Simulations - Potentials, Challenges and Recommendations” by Schreck et al. (2024) outlines a step-based method for performing a plasmode simulation which we have chosen to summarize.

Several of the steps involve the formulation of the research problem and the sampling of the data-set. As the problem in questions involves comparing two models generally and the data-set is used mostly to illustrate the two models this does not apply in our case.

After that the values for  $m$  and  $N$  must be chosen. We chose  $m = 45000$  as the algorithm was deemed to take too much time to implement and run. We chose  $N = 500$ , as figure 4 on page 1820 in Schreck et al. (2024) shows that the bias and the prediction converges some time around 500 simulated data-sets.

After the simulation is performed a quality check has to be performed to make sure the simulated data sets agree with the original data. We chose to quality check by making sure the number of uncensored observations were on average close to the one observed in the original data set.

Finally the results should be reported. Our results can be found in Section 4.1 of the thesis.

### 3.5.2 The Weibull distribution

The Weibull distribution is a flexible distribution that works well for modeling times. The Weibull distribution has a shape parameter  $k$  and rate parameter  $b$  and probability density function

$$f(x; k, b) = bkx^{k-1}e^{-bx^k}, \quad x \geq 0,$$

and zero otherwise.

The reason we chose this parametrization is because of the resulting hazard rate

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{bkx^{k-1}e^{-bx^k}}{1 - (1 - e^{-bx^k})} = bkx^{k-1},$$

which gives a clear interpretation of how each parameter affects the hazard rate. We can see that when  $k = 1$  the hazard rate takes the constant value  $b$ . When  $0 < k < 1$  the power of  $x$  becomes negative resulting in a hazard rate that decreases with time. Finally when  $k > 1$  we have a positive power of  $x$  and the hazard rate increases with time. The parameter  $b$  adjusts the rate of outcomes for the specified shape parameter  $k$ .

The Weibull distribution we fitted to the observed data had a shape parameter equal to 1, meaning the hazard rate is constant. A Weibull distribution with shape parameter 1 is the exponential distribution. While fitting right censored data to a time distribution is likely to give an estimate with a very large variance, the result was deemed realistic when compared with the results of the Cox and FPM model fitted to the data.

The probability density function of the exponential distribution is

$$f(x; \lambda) = \lambda e^{-\lambda x},$$

where this parametrization of the exponential distribution uses the rate parameter as opposed to the mean parameter.

The exponential distribution is memoryless. This means that an exponentially distributed random variable with parameter  $\lambda$  has the same distribution at all points in time until an event actually takes place. Let  $T$  be exponentially distributed with non-negative rate  $\lambda$ ,  $s$  be some positive point in time and let  $t$  be a positive number. Then

$$\begin{aligned} P(T > s + t | T > s) &= \frac{P(T > s + t \cap T > s)}{P(T > s)} = \frac{P(T > s + t)}{P(T > s)} = \\ &= \frac{1 - F(s + t)}{1 - F(s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t), \end{aligned}$$

where  $F(x)$  is the cumulative density function of  $T$ . This demonstrates the memoryless property, that at any point in time in which an event has not yet taken place, the amount of time that has passed up until that point contains no information on when the outcome will take place.

### 3.5.3 Simulating crossing hazards

Recall Equation 1 from chapter 4.1 that  $h(t) = \frac{f(t)}{1-F(t)}$ . As such, for the exponential distribution we get that

$$h(t) = \frac{\lambda e^{-\lambda t}}{1 - (1 - e^{-\lambda t})} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

Therefore an exponential distribution with rate parameter  $\lambda$  has a constant hazard rate equal to  $\lambda$ .

Therefore, given two exponential distributions with parameters  $\lambda_1$  and  $\lambda_2$ , their hazard ratio is

$$\frac{h_1(t)}{h_2(t)} = \frac{\lambda_1}{\lambda_2}$$

which is constant, fulfilling the criteria for proportional hazards.

To simulate non-proportional hazards we have chosen to use the truncated piecewise exponential method defined in “Understanding the Cox Regression Models with Time-Change Covariates” by Zhou (2001).

To simplify the explanation, assume we only use treatment as covariate. Let  $\lambda_1$  be the hazard rate for treatment 1 and  $\lambda_2$  be the hazard rate for treatment 2. We divide the complete time interval into two parts at some positive point  $m$ . Then for each observation in our simulated data-set we simulate a time from an exponential distribution with parameter  $\lambda_1$  or  $\lambda_2$  depending on which treatment they were given.

Then if the simulated time was larger than  $m$ , we simulate a second time from the exponential distribution except this time we switch the parameters so that observations with treatment 1 have the rate of  $\lambda_2$  and vice versa. The simulated times for these observations are then returned as  $m$  plus the second simulated time (First simulated time is truncated at  $m$ ). Any simulated time larger than the censor time is then censored and the censor time is returned instead and the observation is marked as censored.

The idea is that in the first interval the hazard ratio is  $\frac{\lambda_1}{\lambda_2}$  and in the second it is  $\frac{\lambda_2}{\lambda_1}$ . Consequently, the hazard ratio is not constant and specifically is larger than 1 in one interval and smaller in the other, meaning that one treatment is better in the short term and the other in the long term. Theoretically this will cause the Kaplan-Meier estimate of the survival curves to cross if the values of  $\lambda_1$ ,  $\lambda_2$  and  $m$  are chosen properly. The hazard rates for this type of simulation is illustrated in Figure 7.

### 3.5.4 Simulating the piecewise exponential distribution

To simulate the piecewise exponential distribution we use the inverse cumulative distribution function method. The method utilizes that if  $F(x)$  is the cumulative distribution function for the random variable  $X$ , then  $F(X) = U$  where  $U$  is standard uniformly distributed. By inverting the cumulative distribution function we get that  $F^{-1}(U) = X$  which means we can simulate from  $X$  by simulating from a standard uniform distribution and inserting it into the inverse cumulative distribution function.

Recall from Section 3.1 that  $F(t) = 1 - e^{-H(t)}$ . Some quick calculations give that  $F^{-1}(t) = H^{-1}(-\ln(1-t))$  so the only thing we need now is  $H(t)$  to calculate  $H^{-1}(t)$ . To get  $H(t)$  we first need  $h(t)$ . We define the hazard rate that we want to achieve with our simulated times. We want the hazard rate to be constantly  $\lambda_1$  up until time  $m$  and then constantly  $\lambda_2$  from that point onwards. We write this hazard rate as

$$h(t) = \begin{cases} \lambda_1 & , 0 \leq t \leq m \\ \lambda_2 & , m < t \end{cases}$$

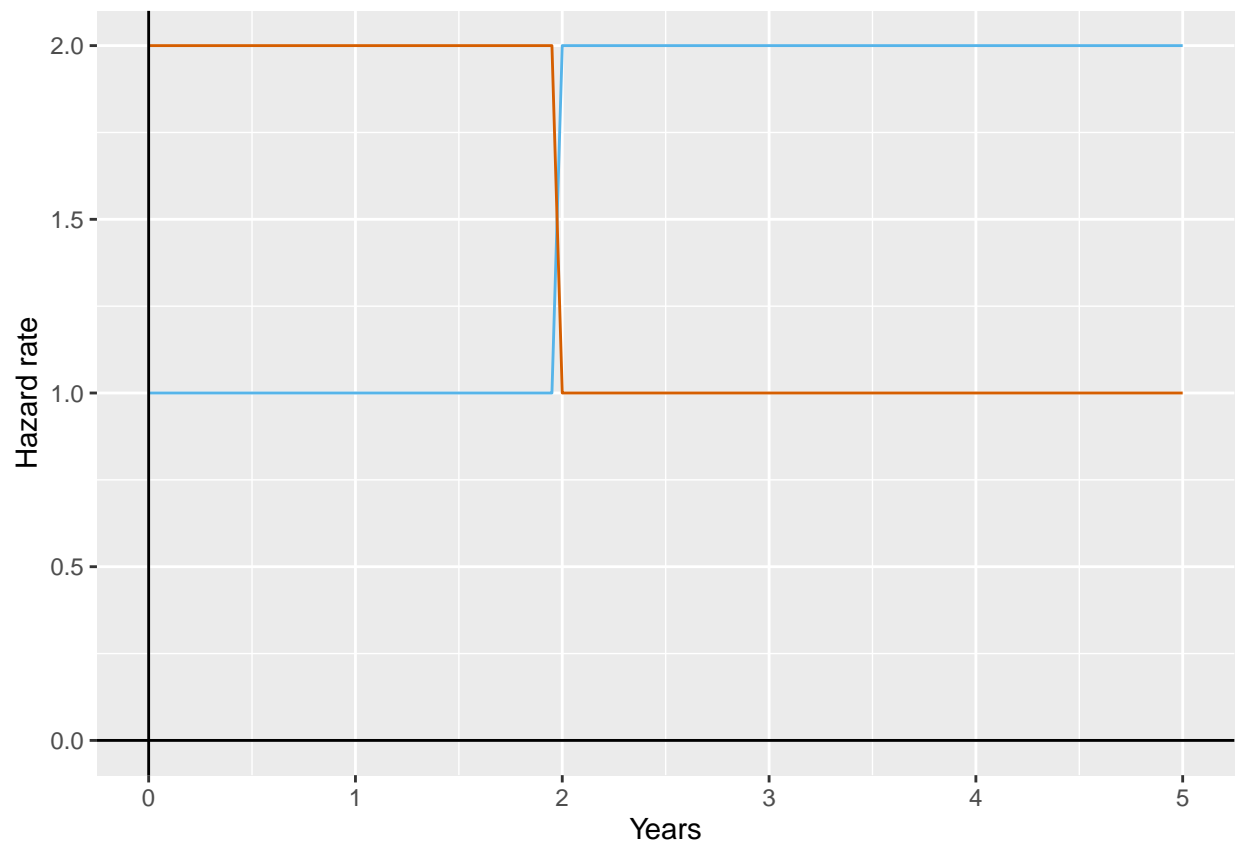


Figure 7: Hazard function for piecewise exponential distribution where the hazard rates switch after 2 years has passed.

where  $I$  is the indicator function. The cumulative hazard rate is then

$$H(t) = \int_0^t h(s)ds = \begin{cases} \lambda_1 t & , 0 \leq t \leq m \\ \lambda_1 m + \lambda_2(t - m) & , m < t \end{cases}$$

We then invert it using standard methods

$$H^{-1}(x) = \begin{cases} \frac{x}{\lambda_1} & , 0 \leq x \leq \lambda_1 m \\ m + \frac{x - \lambda_1 m}{\lambda_2} & , \lambda_1 m < x \end{cases}$$

Putting it all together we can simulate from our specified piecewise exponential distribution by simulating a standard uniform distribution and inserting it into  $F^{-1}(U) = H^{-1}(-\ln(1 - U))$ .

### 3.5.5 Details of the simulation

We choose to use the two covariates treatment and age group for the simulation. We simulate the effect of treatment to be proportional in one simulation and to have crossing hazards in the second. For the age group we multiply the rate of outcomes with a constant so that the 75+ age group suffers outcomes at 6 times the rate of the 0-44 age group and similarly for the other age groups based on the values seen in the data visualization.

For the simulation of proportional hazards we choose the parameters so that their hazard ratio matches the one observed in the original data-set. The parameters are then scaled so that the number of uncensored observations is close to the one observed in the original data-set.

In the case of non-proportional hazards for the methotrexate treatment the we simulate using the piecewise exponential distribution with parameters  $\lambda$  for the first period and  $2\lambda$  for the second with midpoint after two years. For biological treatment we use the same midpoint and parameter values just with  $2\lambda$  in the first interval and  $\lambda$  in the second. For the midpoint we chose to place it at the two-year mark.

For the value of  $\lambda$  we calibrated it so that the number of uncensored observations is close to the one observed in the original data-set.

## 3.6 Comparing the models

To compare the models we have chosen to use the summary statistics described below.

### 3.6.1 Estimate of $\beta$ ( $\hat{\beta}$ )

Both models estimate  $\beta$ . In the case of proportional hazards,  $\beta$  when multiplied with the covariate vector  $x$  and exponentiated ( $e^{\beta x}$ ) is interpreted as the hazard ratio between an observation with covariates  $x$  and an observation with zeroes for all covariates. In the case of non-proportional hazards, this value can be interpreted as the mean hazard ratio across the time interval. The FPM is capable of modelling a proportional hazards model without any interactions between the covariates and time. Therefore it would be interesting to see if this estimate differs significantly between the two models when we assume the proportional hazards assumption to hold. Therefore, using the FPM to model proportional hazards will give the same results as the Cox model, just with the extra features from the parametric model like hazard curves.



### 3.6.2 Akaike's information criterion (AIC)

AIC is commonly used to compare different models. AIC is a measure based on the likelihood of the model that penalizes a large number of estimated parameters. As including more estimated parameters will lead to larger likelihood values this penalty gives us a good way to compare models with different numbers of estimated parameters.

AIC as defined in Akaike (1974) as

$$AIC = 2k - 2 \ln(\hat{L}),$$

where  $k$  is the number of estimated parameters in the model and  $\hat{L}$  is the value that maximized the likelihood function for that model.

Using AIC to compare the Cox model with FPM did not turn out to be that useful. The FPM had consistently lower AIC than the Cox model, even when their estimates of  $\beta$  were equal. The difference is likely that the modelling of the splines is included in the likelihood for the model and as such, the AIC for the two models are not directly comparable with this method. AIC would be better used for comparing models from the same families.

### 3.6.3 Concordance

The concordance statistic can be summarized as how often the model would be correct when guessing.

So, for any pair of observations the model has estimated which one has a larger hazard. If the observation with the larger hazard is found to have a longer time than the one with the smaller hazard then the model in a sense "guessed wrong". However, if the observed time matches the estimated hazard then the model "guessed right". By doing this for all pairs of observations in the data-set we can get a probability that the model guessed correctly which is known as the concordance.

This is slightly more complicated when dealing with right censored times. For example, an observation that is censored after 1 year and an observation that is observed to suffer an outcome after 2 years cannot be compared as we do not know for sure whether the censored observation took place before the uncensored one. The solution is simply to only compare observations which are directly comparable.

A value between 50% and 55% means that the model is not significantly better than just guessing randomly. A value between 60% and 70% is typical for hazard models. Should the concordance be lower than 50% then the model is somehow worse than just guessing randomly and needs to be reconsidered.

Concordance as defined in Therneau and Atkinson (1AD) and adapted to the Cox model is

$$\sum_{i=1}^n \left( \sum_{t_j > t_i} \text{sign}(e^{\beta x_j} - e^{\beta x_i}) \right).$$

Under the proportional hazards assumption an observation with covariate vector  $x_1$  has a lower hazard than an observation with covariate vector  $x_2$  if and only if  $\hat{\beta}x_1 < \hat{\beta}x_2$ . As a lower hazard rate corresponds to a longer survival time, this has to be considered when defining the concordance formula as one side of it is reversed from normal definition. If the proportional hazards assumption does not hold then neither does this method of calculating the concordance.

### 3.6.4 Schoenfeld test $p$ -value

The Schoenfeld test was defined in Section 3.4.3 of the thesis. The Schoenfeld test is a hypothesis test for non-proportional hazards. As such for each simulation we can perform this test and record the  $p$ -value. This

gives us an idea of the power of this test. The null hypothesis of the test is for proportional hazards. As the first set of simulations has exactly proportional hazards, we know that the null hypothesis is true in this case. As such we can calculate the probability of type-I error, that is the probability of rejecting the null hypothesis when it is actually true. In the second set of simulations the hazards are not proportional which means that we know the null hypothesis is false. In this case we can calculate the probability of type-II error, the probability of not rejecting the null hypothesis when it is actually false.

### **3.6.5 Percentage of uncensored observations**

This statistic was used to calibrate the simulation parameters. We wanted the percentage of unobserved observations to be close to the one observed for the dermatology data-set, that being 2.67%. To accomplish this we scaled the parameters, ran the simulation and estimated the mean percent of uncensored observations.

## 4 Results

### 4.1 Simulation

Table 6 and 7 show the results for the simulation with a 95% empirical confidence interval. The concordance for the Cox model actually increases for the crossing hazards simulation. As the concordance is a binary measure and the crossing hazards is also binary with a single jump in hazard rates at year 2, perhaps that means the model is still relatively good at guessing which time should be larger than the other.

The table also includes the p-value for the treatment from the Schoenfeld test. A low p-value corresponds to more evidence for non-proportional hazards while a high p-value corresponds to low evidence. The results are easier to interpret in Table 8 where we have calculated the probability of type I and type II error when using the Schoenfeld test to make inference about the simulations.

The last rows shows the difference in the estimate of  $\beta$  for the two models. The difference is negligible and are equal up to around 2 decimal places. This shows that the two models will give the same estimate of the hazard ratio should this be the only statistic of interest.

Next is the AIC for the Cox model and the flexible parametric model (FPM). FPM has lower AIC than the Cox model in both cases. This is not entirely a fair comparison since the estimates of  $\beta$  are the same for both models as was shown in the same tables. The difference in AIC for the two models is likely caused by the FPM modelling the hazard rates while the Cox model does not.

Table 8 shows the simulated probability of type I and type II error when using the Schoenfeld test. In the case of proportional hazards we can only have type I errors. The probability of type I error in this case is relatively close to the significance levels of the tests.

The probability of type II errors in the case of crossing hazards is a bit worse. With a 26% chance of type II error at a significance level of 1%, the test will fail to reject the null hypothesis of proportional hazards one out of four times. Since this case involves a very extreme case of non-proportional hazards, the probability of type II error is larger than one might expect. One would assume that the probability of type II error would increase if the non-proportionality of the hazards were less extreme.

Table 6: Simulation results for the proportional simulation with mean and empirical 95% confidence interval.

Type	lowQuantile	mean	highQuantile
Cox Concordance	0.650	0.663	0.679
Schoenfeld p value	0.027	0.483	0.976
Cox AIC	24521.973	25869.510	27147.326
FPM AIC	14622.174	15278.028	15912.404
Difference in beta estimate	-0.001	0.000	0.001

Table 7: Simulation results for the crossing hazards simulation with mean and empirical 95% confidence interval.

Type	lowQuantile	mean	highQuantile
Cox Concordance	0.654	0.669	0.682
Schoenfeld p value	0.000	0.020	0.173
Cox AIC	25567.616	26959.031	28304.204
FPM AIC	14968.842	15644.808	16281.475
Difference in beta estimate	-0.002	0.000	0.001

Table 8: Simulated probability of Type I and Type II error when using the Schoenfeld test at 5% and 1% significance levels.

Hazards	Error type	Probability of 5% significance	Probability of 1% significance
Proportional	Type-I	4.4 %	0.8 %
Non-Proportional	Type-II	9 %	26.8 %

## 4.2 Kaplan-Meier plots

A Kaplan-Meier plot visualizes the Kaplan-Meier estimate of the survival function  $S(t) = P(T > t)$ . When creating a Kaplan-Meier plot for a covariate, we plot the Kaplan-Meier estimate for all observations with and without that covariate separately. This gives us a non-parametric method of judging the effect of a covariate on the rate of outcomes. In addition it is also possible to use it to evaluate the proportional hazards assumption. However, it is much easier to judge the log-minus-log transformation of the Kaplan-Meier estimates, as then the graphs are parallel under the proportional hazards assumption. In the Kaplan-Meier estimates the relation between them is that they have different exponents which is unintuitive to judge visually. The log-minus-log transformed estimates can be found in figure 9.

The graphs are all on a scale from 0 to 1. Since a very large number of observations are censored, most curves do not move significantly over time. While it might be tempting to zoom in further on the graphs to see the shapes of the curves more clearly, doing so would likely make the differences between them appear more significant than they are in their effect on the outcomes. The fact that the difference in survival curves for age group in Figure 8 and the one for history of VTE are very apparent even at this scale, shows that their effects on the probability of survival are significant. The differences for treatment and sex on the other hand in the same figure are almost unnoticeable.

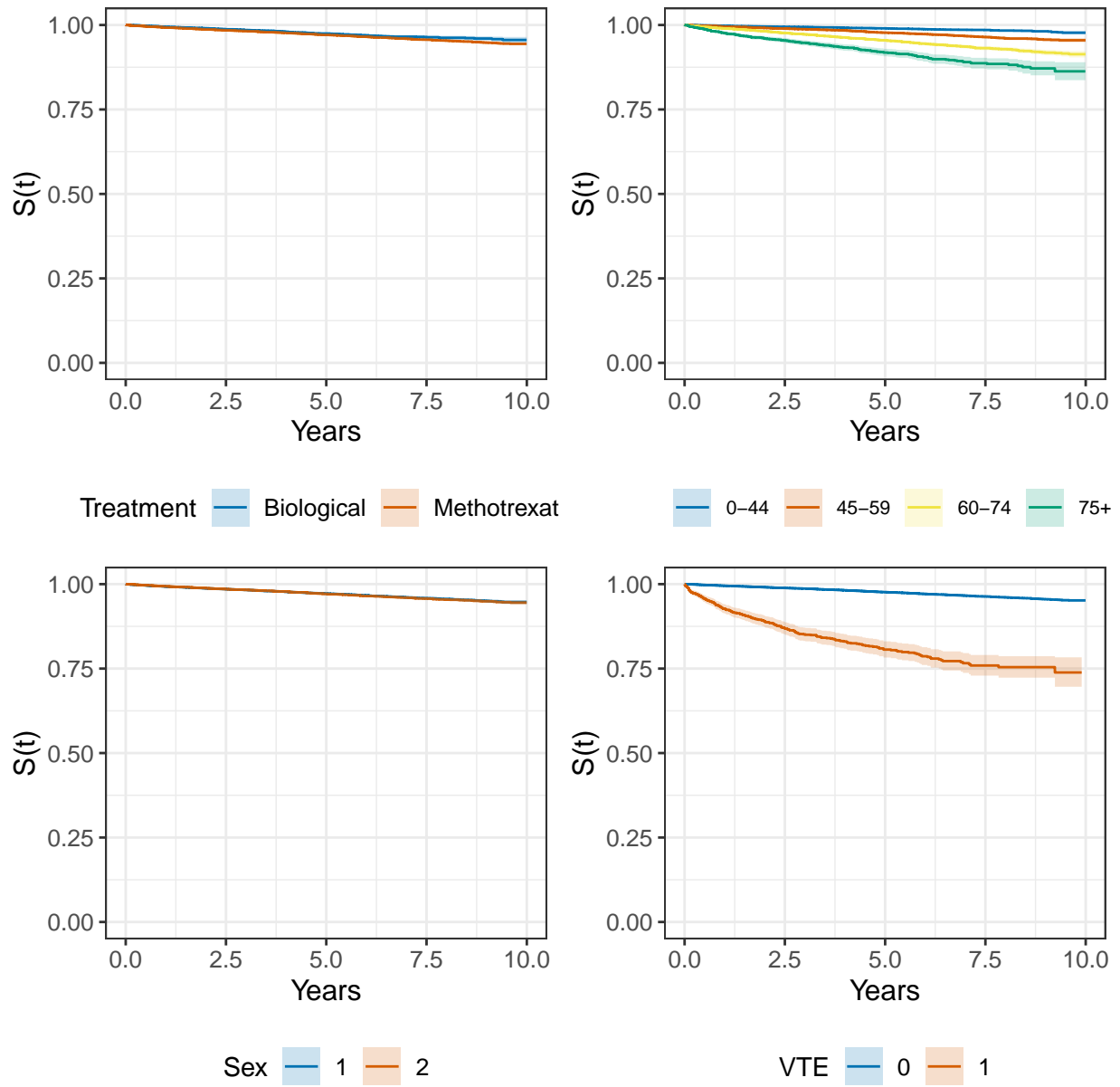


Figure 8: Kaplan-Meier estimate of the survival function for the covariates agegroup, treatment, sex and history of VTE.

### 4.3 Log-minus-log survival plots

If the hazards for a variable are proportional then the log-minus-log survival plots should be parallel to each other. Making this judgement visually is quite difficult. According to the Schoenfeld test, only VTE has non-proportional hazards and as such, we would expect its log-minus-log survival plot in Figure 9 to not be parallel. That plot is not symmetrical early on but does gets closer in the later part of the interval. The other plots are also mostly symmetrical closer to the end of the time interval and slightly off at the beginning.

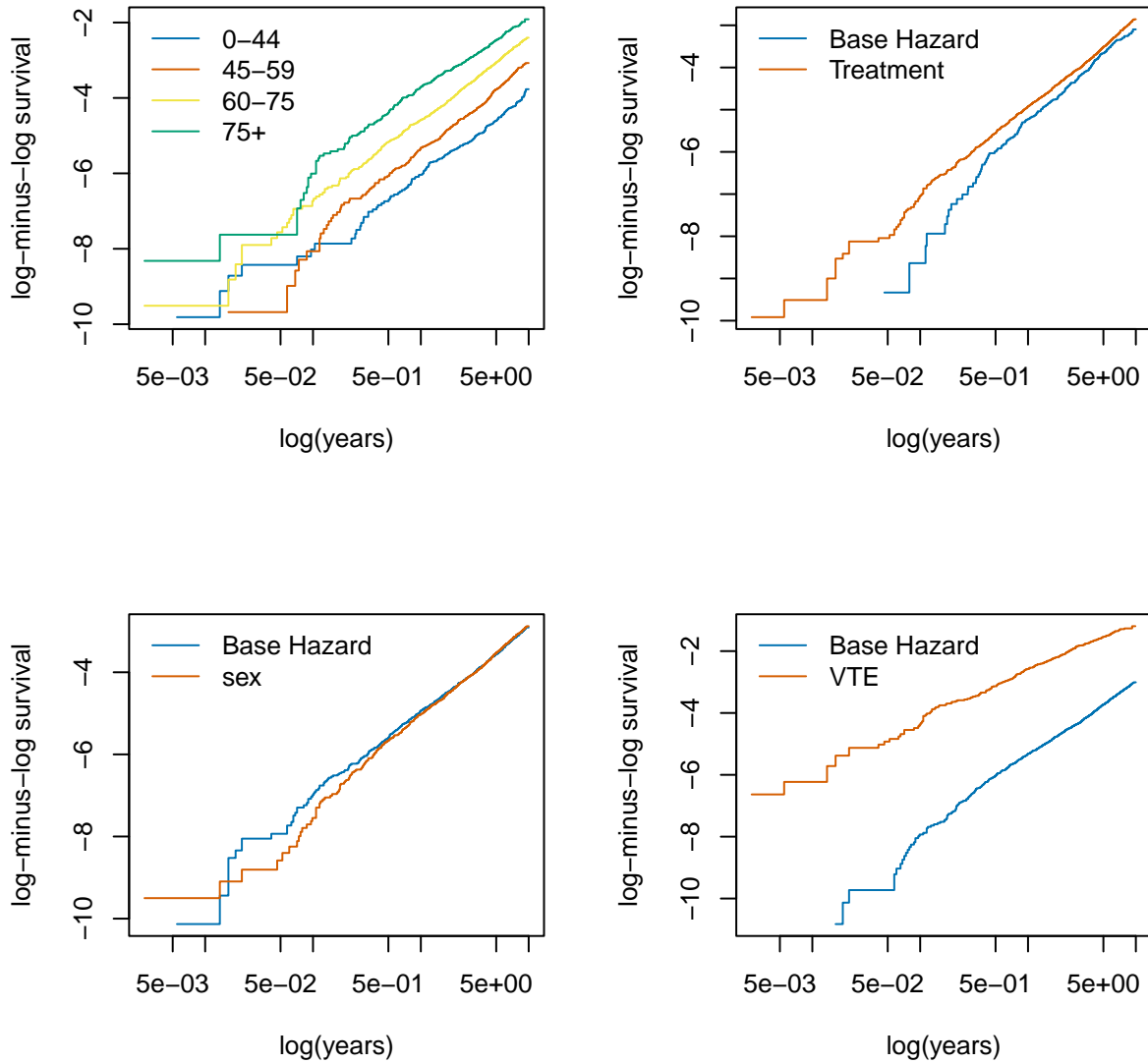


Figure 9: Log-minus-log transformation of the Kaplan-Meier estimate of the survival function for the covariates agegroup, treatment, sex and history of VTE.

## 4.4 Cox model

Table 9: Summary of the Cox regression model on the dermatology data-set.

Covariate	beta estimate	exp(beta estimate)	std.error	z-statistic	p.value
TreatmentMethotrexate	0.2416014	1.2732866	0.1323885	1.8249428	0.0680097
ageGroup45-59	0.7277413	2.0703990	0.0934402	7.7883146	0.0000000
ageGroup60-74	1.3035907	3.6824955	0.0919487	14.1773709	0.0000000
ageGroup75+	1.7733687	5.8906640	0.1108481	15.9981920	0.0000000
AD	-0.0124496	0.9876275	0.1113426	-0.1118138	0.9109711
sex	-0.1236040	0.8837298	0.0554613	-2.2286527	0.0258370
PSO	0.0257622	1.0260969	0.0839597	0.3068407	0.7589647
VTE	1.7787932	5.9227044	0.0733280	24.2580200	0.0000000
Hypertension	0.0905573	1.0947842	0.0687393	1.3174022	0.1877039
MACE	0.0433765	1.0443311	0.1054362	0.4114010	0.6807785
COPD	0.3693187	1.4467486	0.1124827	3.2833368	0.0010259
Fractures	0.2210865	1.2474313	0.0597623	3.6994314	0.0002161
Skin_cancer	-0.3397898	0.7119200	0.2289222	-1.4843024	0.1377287
Melanoma_skin_cancer	0.6686748	1.9516492	0.3086998	2.1661002	0.0303035
Cancer	0.2206578	1.2468967	0.0936871	2.3552646	0.0185095
Asthma	0.1424743	1.1531235	0.0954634	1.4924494	0.1355814
CD	-0.0950237	0.9093514	0.1677054	-0.5666105	0.5709789
Diabetes	-0.0968573	0.9076856	0.0924823	-1.0473058	0.2949586
Obesity	0.2147095	1.2395017	0.1018473	2.1081518	0.0350179
RA	0.0453026	1.0463445	0.0911233	0.4971575	0.6190780
Rhinitis	0.4162649	1.5162875	0.1635467	2.5452355	0.0109204
UC	0.2835701	1.3278619	0.1502459	1.8873727	0.0591102
Uveit	0.1419451	1.1525133	0.0875851	1.6206531	0.1050920
Atrial_fibrillation	-0.3237390	0.7234390	0.1225996	-2.6406211	0.0082754
Stroke	-0.5679746	0.5666720	0.5876443	-0.9665278	0.3337801
Heart_failure	0.0310426	1.0315295	0.1372846	0.2261187	0.8211091
Hyperlipidemia	0.1193047	1.1267131	0.0972779	1.2264313	0.2200364
Anticoagulant	0.2620749	1.2996238	0.1082680	2.4206131	0.0154944
AD_Drugs	0.3439973	1.4105748	0.2139021	1.6081995	0.1077915
History_Methotrexate	-0.4756267	0.6214954	0.2119963	-2.2435611	0.0248607
History_BIOLOGIC	0.2072559	1.2302973	0.1261367	1.6431046	0.1003613

Table 9 shows the results of the full cox model. The column "exp(beta estimate)" shows the estimated hazard ratio between the presence of that covariate and base values for all other covariates and the base hazard rate that has base values for all covariates.

For example, a person on methotrexate is estimated to have a hazard rate that is 1.27 times as large as a person with the same covariates as the first expect for that they are using a biological medicine instead. This is the opposite of what one might expect as methotrexate is seen as the safer treatment. However, the reason for this is likely caused by correlations between the covariates. Specifically, Table 3 and 5 in Section 3.2 show that people in age group 75+ suffer outcomes of VTE relatively around 6 times that of people in age group 0-44 and that increased age leads to relatively more outcomes of VTE. Simultaneously, older age groups are prescribed biological medicines relatively less often than younger age groups. If we do not account for this then it is possible that methotrexate is estimated to be more risky than biological medicines simply because older age groups use methotrexate and suffer VTE more often than younger age groups.

We can see that the model estimated that people in age group 75+ suffer VTE at a 5.9 times higher rate than those in age group 0-44. This estimate is close to the results in Table 5 in Section 2.2 of the thesis.

The p value is for the z-statistic and is a hypothesis test with null hypothesis  $\beta = 0$  which corresponds to a hazard ratio of 1, that is the presence or absence of the covariate has no effect on the hazard rate.

The variables with a significance level of 5% are age group, sex, VTE, COPD, Fractures, Melanoma skin cancer, cancer, Obesity, Rhinitis, Atrial fibrillation, Anticoagulant and history of methotrexate.

The variables with a significance level of 1% are age group, VTE, COPD, Fractures and atrial fibrillation. The effect of treatment is not significantly different from zero, even at a 5% significance level.

#### 4.4.1 Schoenfeld test

Table 10 shows the p-values for the Schoenfeld test. The only covariate with a p-value below 5% is history of VTE.

Table 10: Schoenfeld test table.

Covariate	chisq	df	p
Treatment	0.0975076	1	0.7548413
ageGroup	6.8286394	3	0.0775649
AD	2.0124466	1	0.1560135
sex	0.3867011	1	0.5340385
PSO	0.5401185	1	0.4623836
VTE	38.0958109	1	0.0000000
Hypertension	2.2449131	1	0.1340544
MACE	1.0234135	1	0.3117107
COPD	1.6145751	1	0.2038499
Fractures	1.6193453	1	0.2031831
Skin cancer	0.0305406	1	0.8612693
Melanoma skin cancer	0.1691179	1	0.6808969
Cancer	0.4605059	1	0.4973876
Asthma	0.8854886	1	0.3467028
CD	0.1604009	1	0.6887877
Diabetes	1.5845345	1	0.2081087
Obesity	0.1969608	1	0.6571853
RA	0.1069514	1	0.7436420
Rhinitis	0.0000974	1	0.9921260
UC	0.0126569	1	0.9104247
Uveit	0.0844033	1	0.7714166
Atrial fibrillation	0.0731022	1	0.7868724
Stroke	0.5621549	1	0.4533933
Heart failure	0.0261710	1	0.8714835
Hyperlipidemia	1.6742459	1	0.1956908
Anticoagulant	0.7923508	1	0.3733902
AD Drugs	0.0501182	1	0.8228577
History Methotrexate	0.5161045	1	0.4725081
History Biologic	0.0604406	1	0.8058010

#### 4.4.2 Schoenfeld residuals

If the line in the middle of a Schoenfeld residual is not horizontal then that signifies non-proportional hazards. Treatment, history of VTE and sex in Figure 10 show signs of non-proportionality. The graph for the age group covariate is relatively horizontal.



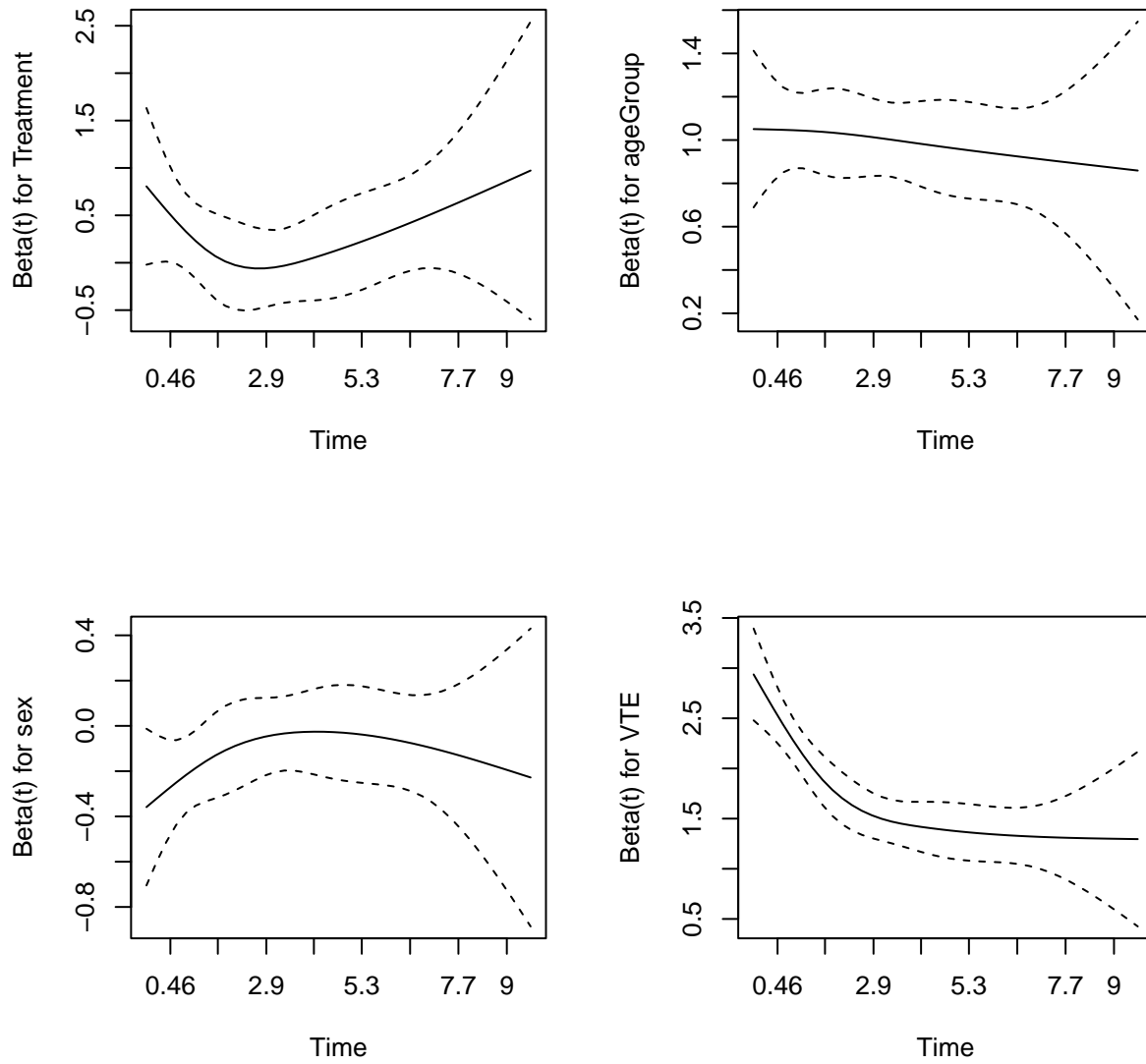


Figure 10: Schoenfeld residuals for the Cox regression estimates of the parameters for the covariates agegroup, treatment, sex and history of VTE.

## 4.5 Flexible parametric model

We chose to use the Schoenfeld test at 5% significance to decide on which covariates should be modeled with interaction with time. As such we only model the covariate history of VTE using splines.

Table 11: Results of the main effects from the flexible parametric model with history of VTE modeled with interaction with time.

Covariate	beta estimate	exp(beta estimate)	std.error	z-statistic	p.value
(Intercept)	-15.1388124	0.0000003	1.0679978	-14.1749473	0.0000000
ageGroup45-59	0.7272433	2.0693682	0.0934256	7.7841965	0.0000000
ageGroup60-74	1.3090555	3.7026747	0.0918932	14.2454015	0.0000000
ageGroup75+	1.7704949	5.8737594	0.1110230	15.9470966	0.0000000
TreatmentMethotrexate	0.2375675	1.2681605	0.1326386	1.7910888	0.0732790
AD	-0.0006422	0.9993580	0.1111702	-0.0057767	0.9953908
AD_Drugs	0.3270701	1.3868986	0.2137208	1.5303610	0.1259274
Anticoagulant	0.2631261	1.3009907	0.1082812	2.4300255	0.0150978
Asthma	0.1374818	1.1473808	0.0954737	1.4399971	0.1498682
Atrial_fibrillation	-0.3126731	0.7314890	0.1225703	-2.5509697	0.0107424
CD	-0.0871333	0.9165549	0.1676046	-0.5198743	0.6031512
COPD	0.3726476	1.4515727	0.1125519	3.3108962	0.0009300
Cancer	0.2196949	1.2456966	0.0937365	2.3437497	0.0190910
Diabetes	-0.1014934	0.9034872	0.0924550	-1.0977599	0.2723094
Fractures	0.2174595	1.2429150	0.0597934	3.6368467	0.0002760
Heart_failure	0.0076039	1.0076328	0.1376991	0.0552208	0.9559625
History_BIOLOGIC	0.2058580	1.2285787	0.1263975	1.6286561	0.1033858
History_Methotrexate	-0.4605639	0.6309278	0.2117855	-2.1746719	0.0296547
Hyperlipidemia	0.1150271	1.1219038	0.0972495	1.1828032	0.2368871
Hypertension	0.0876871	1.0916465	0.0688997	1.2726773	0.2031326
MACE	0.0423895	1.0433007	0.1054393	0.4020272	0.6876640
Melanoma_skin_cancer	0.6385877	1.8938043	0.3079767	2.0734938	0.0381263
Obesity	0.2157304	1.2407678	0.1018010	2.1191386	0.0340788
PSO	0.0261050	1.0264487	0.0838514	0.3113241	0.7555542
RA	0.0494879	1.0507329	0.0911408	0.5429829	0.5871416
Rhinitis	0.4087134	1.5048804	0.1635849	2.4984796	0.0124727
Skin_cancer	-0.3300561	0.7188834	0.2282743	-1.4458753	0.1482122
Stroke	-0.5623505	0.5698680	0.5880178	-0.9563494	0.3388957
UC	0.2917329	1.3387454	0.1503712	1.9400857	0.0523693
Uveit	0.1394111	1.1495966	0.0876465	1.5906069	0.1116981
VTE	6.2619661	524.2486390	1.2293446	5.0937434	0.0000004
sex	-0.1196859	0.8871990	0.0554266	-2.1593590	0.0308223

Table 12: Results of the parameters for the spline functions from the flexible parametric model with history of VTE modeled with interaction with time.

Covariate	beta estimate	exp(beta estimate)	std.error	z-statistic	p.value
nsx(log(time), df = 5)1	9.100221	8.957269e+03	1.0338922	8.801905	0.0000000
nsx(log(time), df = 5)2	9.687869	1.612085e+04	1.0470284	9.252728	0.0000000
nsx(log(time), df = 5)3	8.064648	3.180038e+03	0.6769513	11.913188	0.0000000
nsx(log(time), df = 5)4	16.900224	2.186121e+07	2.0581885	8.211213	0.0000000

Covariate	beta estimate	exp(beta estimate)	std.error	z-statistic	p.value
nsx(log(time), df = 5)5	7.019304	1.118009e+03	0.4197688	16.721833	0.0000000
nsx(log(time), df = 5)1:as.numeric(VTE)	-4.120530	1.623590e-02	1.2151113	-3.391072	0.0006962
nsx(log(time), df = 5)2:as.numeric(VTE)	-4.234376	1.448880e-02	1.2319478	-3.437139	0.0005879
nsx(log(time), df = 5)3:as.numeric(VTE)	-3.390966	3.367610e-02	0.8022721	-4.226703	0.0000237
nsx(log(time), df = 5)4:as.numeric(VTE)	-7.924312	3.618000e-04	2.4228461	-3.270662	0.0010730
nsx(log(time), df = 5)5:as.numeric(VTE)	-2.802450	6.066130e-02	0.5142143	-5.449965	0.0000001

Table 11 shows the results of the flexible parametric model (FPM) where we modeled history of VTE with interaction with time because of its p-value in the Schoenfeld test. The variables in Table 12 with names that include “nsx” are the parameters used for the splines that FPM uses to model the hazard rates and interaction between VTE and time. Consequently, their exact values are difficult to interpret.

Otherwise the estimates for  $\beta$  are quite close to those estimated in the Cox model in Table 9.

#### 4.5.1 Hazard graphs

These figures show the hazard rates estimated by FPM. The hazard rates for agegroup, treatment and sex in Figure 11 are exactly proportional by assumption of the model. Only history of VTE has a non-proportional hazard rate and this is because we specifically defined it to be modeled that way. As such, this does not necessarily mean that any of these covariates have or do not have proportional hazards in reality.

The hazard for history of VTE is very large at the beginning of the time interval before getting closer to proportional at the end. The plots all have the same values for the  $y$  axis. This makes it difficult to see the shape or difference between the hazard rates for most covariates. However, this also shows how small an effect most covariates have in comparisons to the most significant ones like age group and history of VTE.

The hazard rate for history of VTE most drastically deviates from proportional hazards in the very beginning but after that it flattens out. This agrees with the results of the log-minus-log plot for history of VTE in figure 9. This could be interpreted as an observation that has history of VTE is relatively very likely to suffer another outcome of VTE shortly after beginning any of our two treatments. However, if it does not happen early on then the danger recedes after a year or so. The rate then shrinks even further after four years.

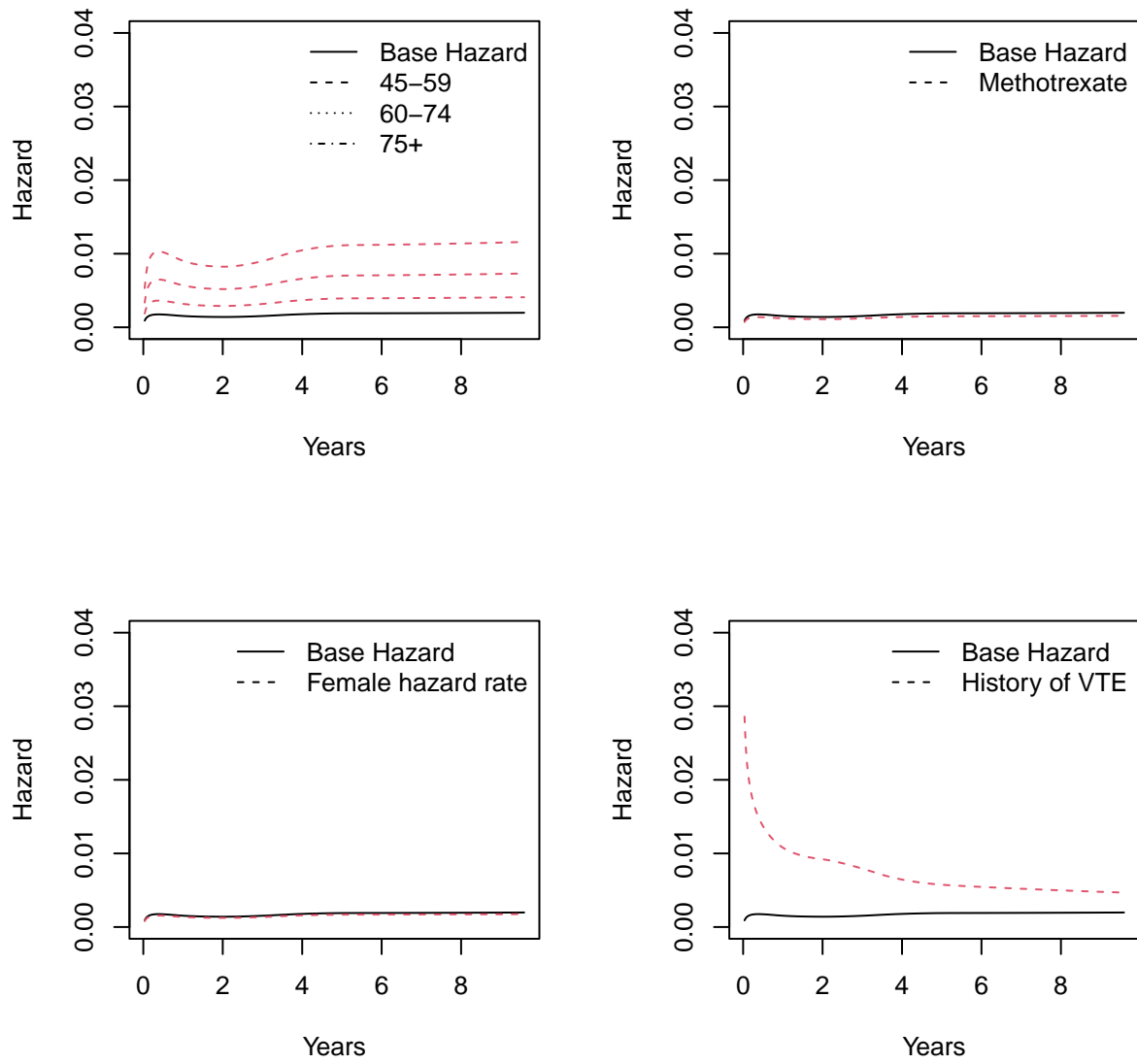


Figure 11: FPM estimated hazard curves for the covariates agegroup, treatment, sex and history of VTE.

## 5 Discussion and conclusions

### 5.1 Hypothesis tests in survival analysis

While the flexible parametric model (FPM) does give us more options for illustrating results, it also complicates model selection. In addition, to needing to choose which covariates should be included in the model, we also need to choose which of these covariates should vary with time. Any choice made is going to change the results of the model and therefore any decision must be made with care.

To make that decision, a statistical test would be the easy answer. The Schoenfeld test which we have used in this thesis is specifically a hypothesis test with proportional hazards as null hypothesis. The problem with using this test is that it has a reputation for being sensitive to small and large sample sizes. Specifically, having a too large data-set can lead to significant results even if proportional hazards is still “good enough” to describe the data.

More visual tests like log-minus-log plots and Schoenfeld residual plots can be used instead. One could argue that making judgement based on visual graphs is likely to at least give an idea of if proportional hazards is “good enough” or not. However, such a subjective judgement is going to differ between different people. The concept of “good enough” is certainly vague enough to be argued about. For log-minus-log plots, one has to judge visually if two lines are parallel to each other. Such a test is likely to be vulnerable to many kinds of visual illusions. There are many visual illusions related to parallel lines so trying to trust your eyes on such matters is risky. Schoenfeld residuals only involve a single line but the judgement must still be made visually.

While the Schoenfeld test is a test for proportional hazards, it is not an omnibus test as there is no such thing. The Schoenfeld test only tests for a single kind of interaction which is interaction between the covariate and time. In addition, there are more kinds of interaction that can cause non-proportional hazards like for example, interaction between the covariates. While it is possible to test for these interactions as well, you might need to test every single pair of variables for such an interaction if you do not already have some previous knowledge of such an interaction. The final complication is that again, tests of interaction in survival analysis have low power and are unlikely to give significant results even when an alternative hypothesis is true.

One could also argue that hypothesis tests are arbitrary. While the significance level for the p-value which makes us accept the alternative hypothesis can be made arbitrarily small, ultimately the choice is arbitrary. While 5% significance levels are commonly used, this makes the choice no less arbitrary.

A critique of testing for proportional hazards can be found in the article “Why test for proportional hazards?” by Stensrud and Hernán (2020). They bring up the case that hypothesis testing is arbitrary but also that statistical tests in survival analysis are usually of low power. The power of a test is defined as the probability of rejecting the null hypothesis when the alternative hypothesis is true. A low powered test means that the test is unlikely to give us significant results even when our alternative hypotheses are true. That being the case, relying on these tests is unlikely to be effective in the long run. On the other hand, Sjölander and Dickman (2024) in their comments on Stensrud and Hernán (2020) argues that while making decisions only based on  $p$ -values is not optimal, having some measure of certainty for the model is useful.

Rather than having a single powerful test for hypotheses, an alternative would be to instead have a very large number of trials. So instead of using a single trial as an argument for why something might be true it might be preferable to have a very large number of trials that say approximately the same thing. If our tests have low power it might still be useful to reveal the results to see if other trials had the same kinds of patterns and behaviors. By doing so, certain hypotheses might instead accumulate evidence over time and be significant in a different way.

This does not mean that we should not perform any hypothesis tests or report any confidence intervals. While using them to make hard conclusions is frowned upon, having some measurement of certainty is certainly useful.

When choosing which covariates should be modeled to have interaction with time, we chose to use the Schoenfeld test with a 5% significance level. The result was that we only modeled history of venous thromboembolism (VTE) as non-proportional. Had we instead chosen to use the log-minus-log plots or Schoenfeld residuals to make our decision then we might have modeled several covariates as time dependent. Several of the visual tests showed some deviation from proportionality, but how much deviation is enough to warrant additional parameters to model them. The more parameters we estimate the higher the risk of overfitting becomes and as such, only modeling a few select covariates is to prefer.

Another method would be to use experience as a basis when choosing the non-proportional covariates. For example, age tends to have a non-linear effect on a lot of medical conditions. Given this, one could assume that this would also apply in our case and that we should always model age as non-linear. This would also be subjective and there is no guarantee that age would be non-proportional in our case.

No method is perfect, at least not in the case of the dermatology data-set. Given that all the tests return different conclusions, perhaps we should just conclude that the proportional hazards assumption is good enough. Maybe we should save modelling non proportional effect to the cases where the visual tests show such obvious signs of non-proportionality that it would be difficult to imagine that anyone could disagree with such a judgement.

Specifically in our case we should instead first focus on taking into account the negative correlation between age and use of biological medicine which could have caused methotrexate users to be estimated to have a larger hazard rate than non-methotrexate users.

## 5.2 Simulation

A drawback with the simulation is that we can only extract summary statistics from it. As has been shown throughout the thesis, a lot of the results from a survival analysis are graphical in nature. It is not feasible to create a graph for each of the hundreds or thousands of repetitions that are necessary to get good results from a simulation. Consequently, we have to settle for summary statistics. The question is then if these summary statistics are good enough to convince anyone who has been using one model to instead start using another.

If someone considers the proportional hazards assumption to be good enough then a summary statistic about the AIC of a different model likely would not hold much sway. In addition, even when the models returned close to the same  $\beta$  parameters, the AIC for the Cox model and FPM were significantly different. As such using AIC to compare the two models was not so useful.

Simulation might not be the best way of showing why a more modern model might be preferable. Consequently, focusing the comparison on the kind of results we can only get from a different model might be a better way to argue for its use.

For our simulation we kept the censor times the same as in the original data-set as to keep the correlation between the covariates and the censor times. Simulating new censoring times would have introduced more variance into the simulation which could have in turn led to more realistic results. However, the way that the results of a survival model are influenced by the outcome of the censoring distribution is a well deep enough to warrant its own investigation. The supplementary material in the article by Stensrud and Hernán (2020) contains a single table showing how different outcomes of the censoring distribution changes the estimate of the hazard ratio without changing the survival difference. The plasmode simulation method of keeping the censoring times did let us avoid this complication but the censoring distribution is going to have a significant effect on the simulation. Any simulation should take care with its choice of censoring distribution and perhaps even test how that distribution influences the outcomes of the survival analysis.

We simulated the crossing hazards using the piecewise exponential distribution with one cutoff point. This creates a significant discontinuity in the hazard rate. An alternative to fix this would be to include more cutoff points throughout the time interval and making the change in the hazard rate more gradual. The hazard rates would still be discontinuous but to a much smaller extent.

An alternative method could be to model the hazards rate using continuous functions. For example choose one function that starts low and increases up to a certain point before planing out and another that starts high and ends low. Using that  $F(t) = 1 - \exp(-H(t))$  we could simulate from the time distribution for these hazard rates using the inverse distribution function method as described in Section 3.5.4.

A complication would be choosing the functions in question as there would be infinitely many ways to do so. One restriction would be that the cumulative hazard function  $H(t)$  would need to be invertible, specifically invertible using standard mathematical terms. For example, the hazard rate  $h(t) = \cos(t) + 1$  has cumulative hazard rate  $H(t) = \sin(t) + t$  which is not invertible with standard mathematical terms. In this case special methods would need to be used to invert the cumulative hazard function

### 5.3 Cox Regression variants

It is important to keep in mind that there is not one single Cox regression model or flexible parametric model (FPM). There are options within the Cox regression model that can make the model perform very differently. There are ways to make the Cox regression model more robust even in the presence of non-proportional hazards. With so many variants it can be difficult to talk about *the* Cox model and *the* FPM model. For example, if you fit the data to a Cox model and then independently fit a natural spline to an estimate you made of the hazard rate, does this still count as a Cox model? It is also possible to model interaction with time for your covariates with the Cox model as well. As such, saying that the Cox model only does this while the FPM does all of this is reductive.

That is why we have limited the Cox model to just the base version of the model. To define what counts as the base model of the Cox model, we use an argument from Royston and Parmar (2002). They state that while there are methods to model the base hazard rate and non-proportional effects within the Cox framework, there is no “‘natural’ widely accepted approach” for doing so. For FPM on the other hand, such modeling is all performed using splines.

If someone uses the Cox model while also estimating the base hazard rate function and model interaction with time, then there might not be a need to switch to using the FPM instead. However, if all they are doing is using the base Cox model then using FPM instead would be preferable due to its extra results.

One method of modeling non proportional hazards with the Cox model is to divide the time interval into pieces and fit a separate Cox regression model in each of the intervals. Optimally this would be done at points where the proportional hazards assumption is at its thinnest such as locations where the hazard rates cross each other. This would give us two sets of  $e^{\beta x}$  estimates where the estimate for a covariate with crossing hazards would be larger than 1 in one interval and smaller than 1 in the other. This would give us two models with approximately proportional hazards at the cost of needing to choose where the interval should be split and the extra problem of needing some way to interpret the results of these two models in relation to each other.

Another method to deal with non proportional hazards is stratification. If we knew that some covariate did not fulfill the proportional hazards assumption either through experience or from testing then we could stratify that covariate. The idea is that instead of having a single base hazard rate we instead have two depending on the value of the stratified covariate. Let  $x_s$  be the covariate we stratify. Then we would have the two base hazard rates  $h_0(t|x_s = 0) = h_{00}(t)$  and  $h_0(t|x_s = 1) = h_{01}(t)$ . This lets us account for some of the non-proportional hazards of the model while still avoiding actually modeling the shape of these base hazard rates.

### 5.4 Data sets with obviously non-proportional hazards.

In this thesis we have performed survival analysis on a data-set relating to dermatology with two treatments. While the methotrexate treatment could be described as a conservative treatment and the more general biological treatments as a more aggressive treatments, the difference in effects they have on the outcome of

VTE is not significantly high nor are they obviously non-proportional. Consequently, one could conclude that the Cox regression model is good enough for a model on this subject.

A better data-set for comparing outputs would be one where there is an obvious case of non-proportional hazards such as one where the estimated survival curves cross. Cancer data sets for example, are said to often have non-proportional hazards and a cancer data-set with surgery as an aggressive treatment would be likely to have non-proportional hazards. Using such a data-set for this kind of comparisons might have caused the Cox model to not be considered good enough anymore.

## 5.5 Model selection

When creating any type of model, a common problem is one of selecting which covariates should be included. The inclusion or exclusion of any one covariate can completely change the estimated coefficients of the model. When choosing which of the  $n$  total covariates should be included in the Cox model there are

$$\sum_{i=0}^n \binom{n}{i} = \sum_{i=0}^n \binom{n}{i} 1^n 1^{n-i} = (1+1)^n = 2^n$$

total models to choose from where the second equality is the binomial identity. For small  $n$ , one could test all possible models and choose the model that has the lowest AIC. If  $n$  is too large then stepwise regression can be used to reduce the amount of calculations necessary for the tests.

For FPM, selecting covariates becomes more complicated. Not only do we need to choose which covariates should be included, we also need to choose which of the chosen covariates should be modeled with interaction with time. That means that for any choice of  $i$  number of covariates to include in the model we then have  $2^i$  possible choices for interaction. This means that in total we have

$$\sum_{i=0}^n \left( \binom{n}{i} \sum_{j=0}^i \binom{i}{j} \right) = \sum_{i=0}^n \binom{n}{i} 2^i = \sum_{i=0}^n \binom{n}{i} 2^i 1^{n-i} = (2+1)^n = 3^n$$

possible choices of models. This means the number of models when modeling interaction with time grows exponentially faster than when modelling without. For example, if we have 10 covariates then without modelling interaction there are  $2^{10} = 1\,024$  choices for covariates while with interaction there are  $3^{10} = 59\,049$  choices. Our dermatology set has 28 covariates so the number of choices would be  $2^{28} = 268\,435\,458$  and  $3^{28} = 22\,876\,792\,454\,961$  without and with interaction respectively. So, when modelling interaction we have over 22 trillion possible choices.

The FPM takes a relatively long time to run, especially when a lot of variables are chosen to interact with time. In combination, these two facts mean that it is likely unreasonable to test every single model when the number of covariates is even slightly on the large side. Therefore, we would prefer to use some other method, for example, stepwise regression to choose our model. However, how we would apply stepwise regression to this choice is not entirely obvious. We could for example, first use stepwise regression to choose which covariates should be included in the model and then potentially test every single combination time dependent covariates for the final model. However, there is no guarantee that this will actually lead to a good final model as we are ignoring the effect of our choices of time dependent covariates on the model until we reach the end of the selection step.

An alternative might be to instead of only choosing which covariate should be included or excluded in each regression step, we also test all inclusions of adding time dependence to any covariate currently chosen. This would increase the number of possible choices to be tested in each step but we would be able to account for the effects of time dependent covariates on the model.

Another way to choose which variables should be time dependent would be to use some kind of test for non-proportional hazards. However, these tests are unreliable. log-minus-log survival plots and Schoenfeld



residuals must be judged visually and Schoenfeld tests have low power. The article “Why test for proportional hazards?” by Stensrud and Hernán (2020) argues that tests for proportional hazards should be avoided. For that reason, some other method of method selection would be preferred when using FPM.

## **5.6 Complications of survival analysis**

### **5.6.1 The frailty effect**

While hazard ratios are popular due to their ease of interpretation they tend to be biased. One reason for this is the “frailty effect” or the more descriptive name “unobserved individual heterogeneity”. The idea is that certain individuals are predisposed to more often suffer certain outcomes due to their genetics. This could mean that the majority of the people who do suffer from the outcomes in questions are just those who are genetically predisposed to them. These predisposed people would then likely suffer outcomes early on when exposed to aggressive treatments. The predisposed people in the conservative group would lay latent and only suffer their outcomes later on in the trial.

This means that the results of any survival analysis might just be modelling the effects of predisposed people rather than the general population.

### **5.6.2 Immortal time bias**

A common bias in medical trials and other types of survival analysis is the concept of immortal time bias or survivorship bias. The main gist of this bias is that there is a risk that any sample collected is not completely representative of the population at risk since the sample is likely to only contain survivors. This means that anyone that is not in the sample due to being eliminated either by the design of the selection or by being impossible to include because they are deceased do not have their information added to the model.

One example of the immortal time bias is that people who are awarded the Nobel price live longer than people who do not. This is because you have to live long to be able to receive a Nobel price and any person who dies young has no chance to win one. As such, the average age of Nobel prize winners are biased upwards.

### **5.6.3 Causality**

While hazard ratios can identify correlation between covariates and outcomes, they contain no information about the causal relationship between them. Consequently, a hazard ratio that is not equal to 1 does not necessarily mean that the covariate in question causes the change in rate of the outcome. Therefore you would need to study the causal relationship between the covariate and outcome if that is what you are interested in.

## 6 References

- Akaike, Hirotugu. 1974. “A New Look at the Statistical Model Identification.” *IEEE Transactions on Automatic Control* 19 (6): 716–23.
- Bickel, Peter J, and Anat Sakov. 2008. “On the Choice of  $m$  in the  $m$  Out of  $n$  Bootstrap and Confidence Bounds for Extrema.” *Statistica Sinica*, 967–85.
- Cox, David R. 1972. “Regression Models and Life-Tables.” *Journal of the Royal Statistical Society: Series B (Methodological)* 34 (2): 187–202.
- Grambsch, Patricia M, and Terry M Therneau. 1994. “Proportional Hazards Tests and Diagnostics Based on Wrighted Residuals.” *Biometrika* 81 (3): 515–26.
- Kaplan, Edward L, and Paul Meier. 1958. “Nonparametric Estimation from Incomplete Observations.” *Journal of the American Statistical Association* 53 (282): 457–81.
- Lambert, Paul C, and Patrick Royston. 2009. “Further Development of Flexible Parametric Models for Survival Analysis.” *The Stata Journal* 9 (2): 265–90.
- Lin, Haiqun, and Daniel Zelterman. 2002. “Modeling Survival Data: Extending the Cox Model.” Taylor & Francis.
- Royston, Patrick, and Mahesh KB Parmar. 2002. “Flexible Parametric Proportional-Hazards and Proportional-Odds Models for Censored Survival Data, with Application to Prognostic Modelling and Estimation of Treatment Effects.” *Statistics in Medicine* 21 (15): 2175–97.
- RStudio Team. 2020. *RStudio: Integrated Development Environment for r*. Boston, MA: RStudio, PBC. <http://www.rstudio.com/>.
- Schreck, Nicholas, Alla Slynko, Maral Saadati, and Axel Benner. 2024. “Statistical Plasmode Simulations—Potentials, Challenges and Recommendations.” *Statistics in Medicine*.
- Sjölander, Arvid, and Paul Dickman. 2024. “Why Test for Proportional Hazards—or Any Other Model Assumptions?” *American Journal of Epidemiology*, kwae002.
- Stensrud, Mats J, and Miguel A Hernán. 2020. “Why Test for Proportional Hazards?” *Jama* 323 (14): 1401–2.
- Therneau, Terry, and Elizabeth Atkinson. 1AD. “1 the Concordance Statistic.” CRAN.
- Zhou, Mai. 2001. “Understanding the Cox Regression Models with Time-Change Covariates.” *The American Statistician* 55 (2): 153–55.

## 7 Appendix

The appendix contains the graphs that were not presented and discussed in the main thesis.

### 7.1 Kaplan Meier graphs

These are the Kaplan-Meier graphs for covariates that were not shown in the results. Kaplan-Meier graphs are estimates of the survival function for the presence or absence of a covariate and are described in Section 3.1.1.

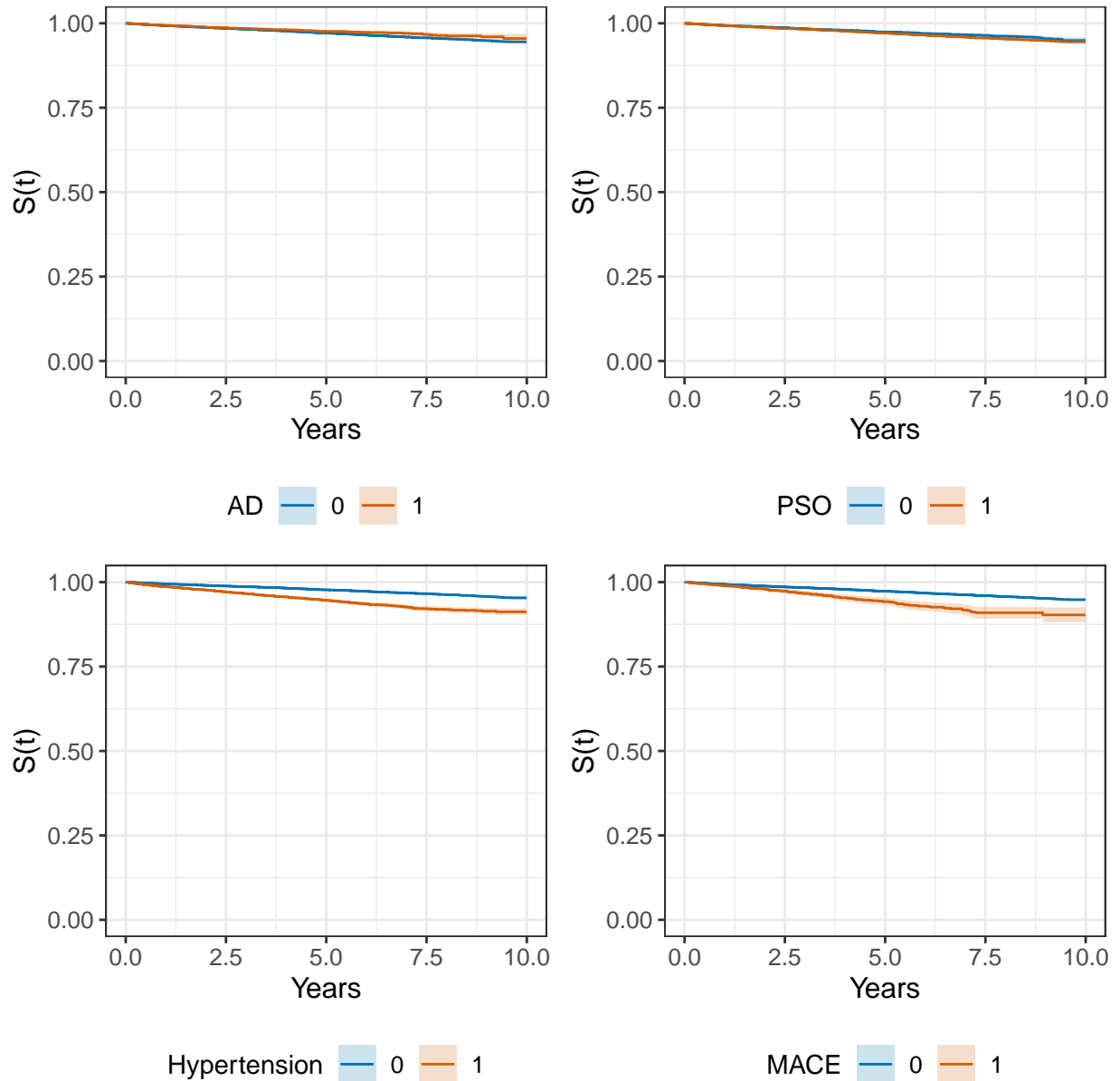


Figure 12: Kaplan-Meier estimate of the survival function with and without the presence of covariates history of AD, PSO, hypertension and MACE.

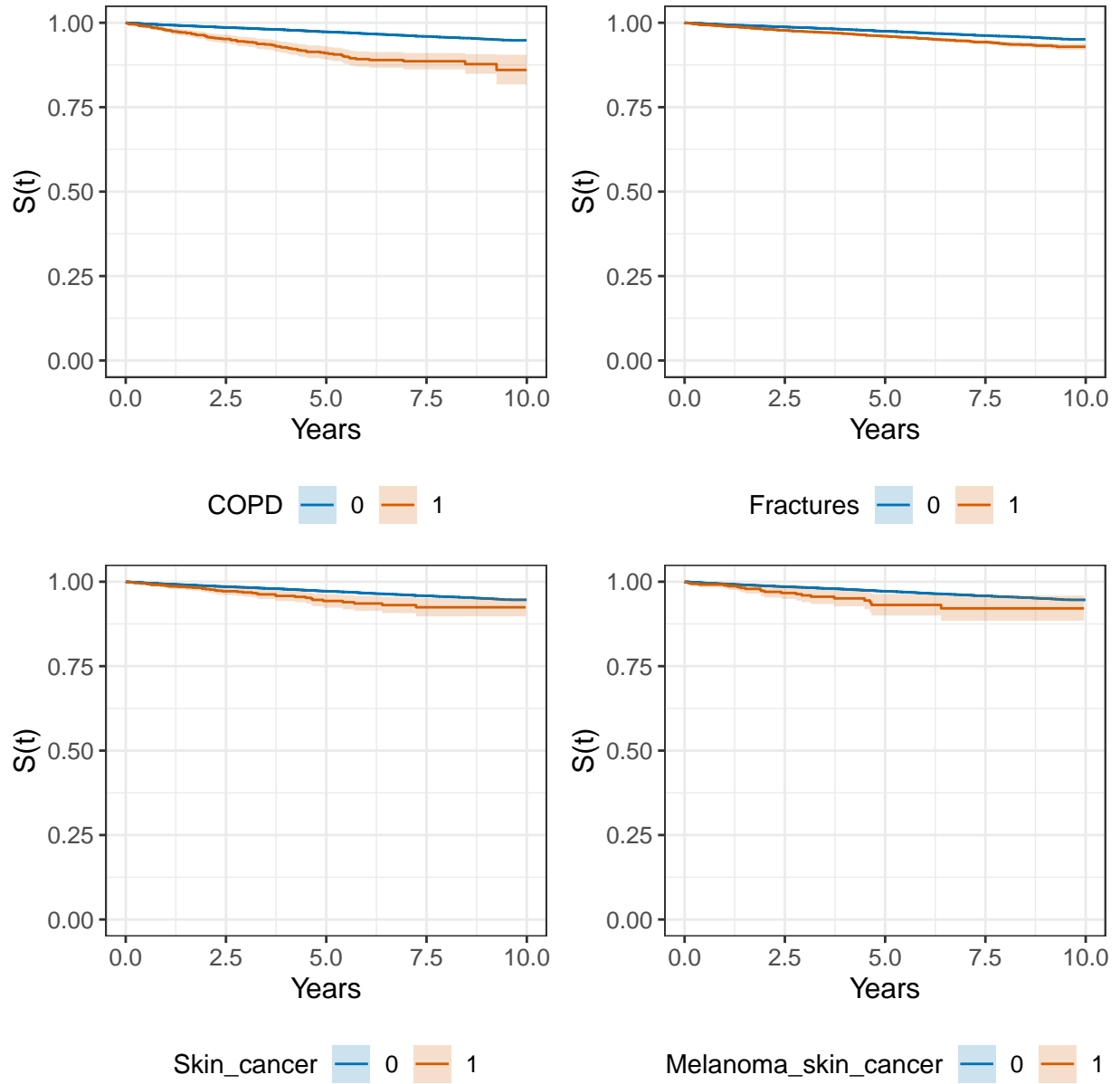


Figure 13: Kaplan-Meier estimate of the survival function with and without the presence of covariates history of COPD, fractures, skin cancer and melanoma skin cancer.

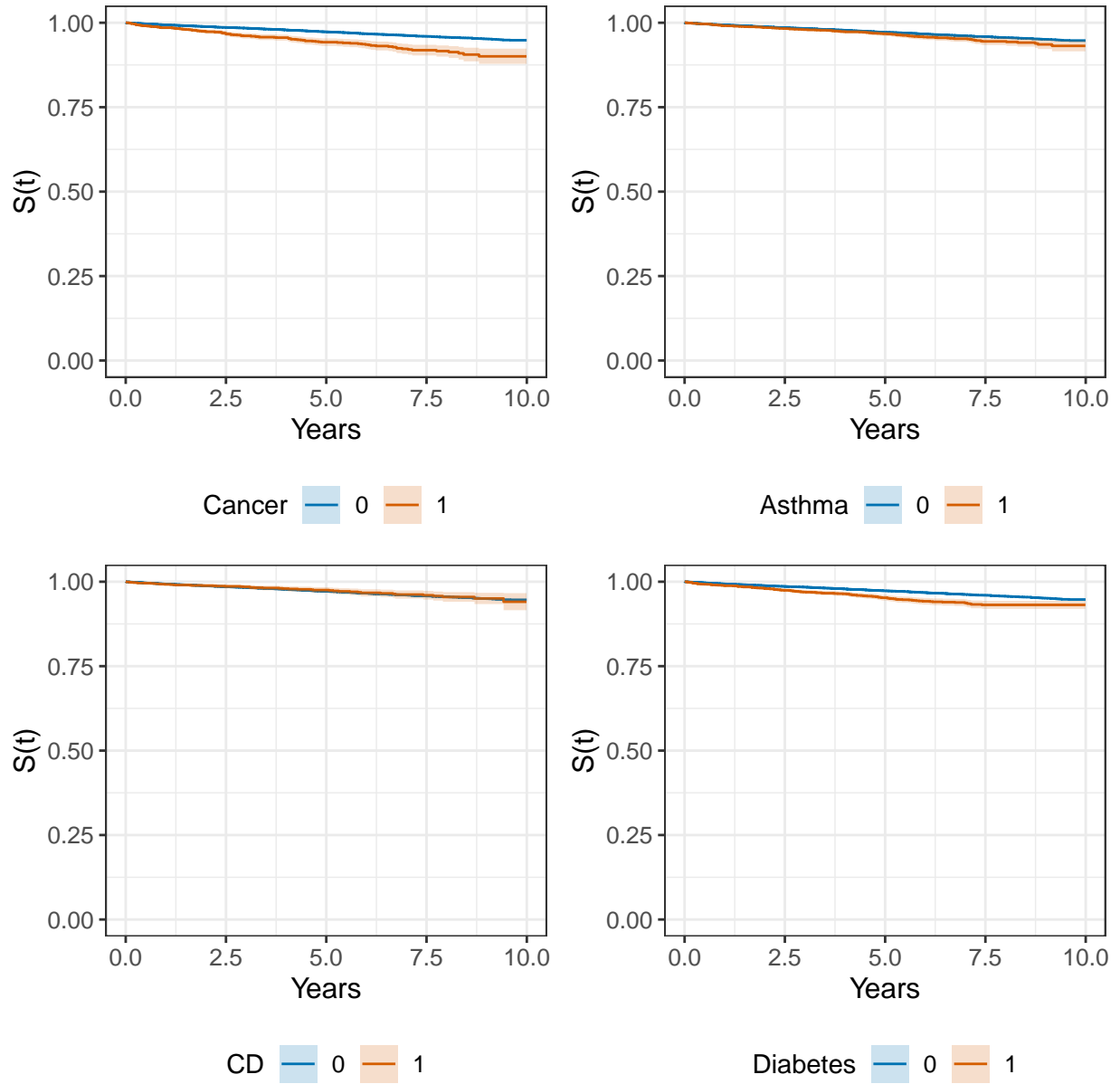


Figure 14: Kaplan-Meier estimate of the survival function with and without the presence of covariates history of cancer, asthma, CD and diabetes.

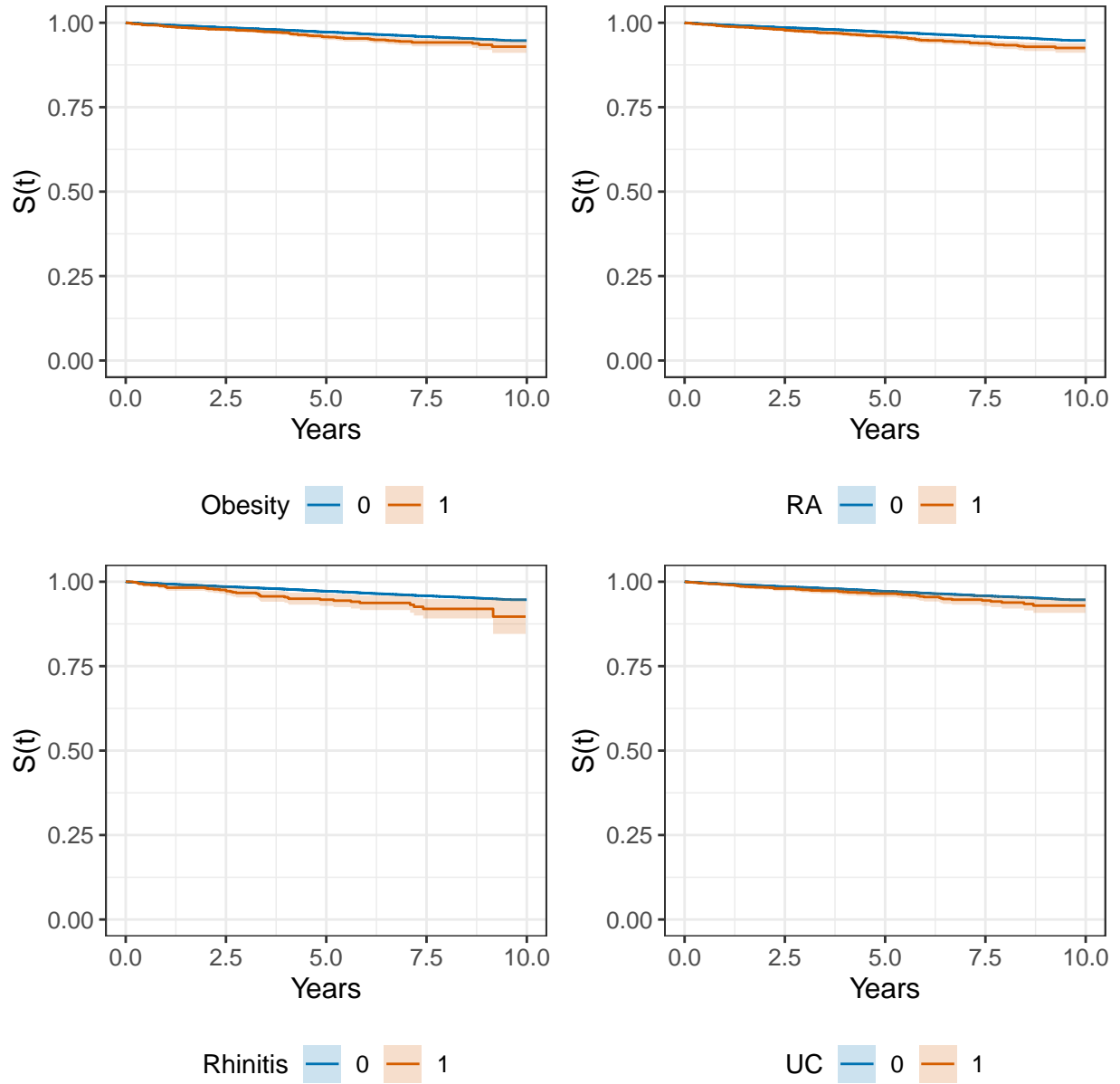


Figure 15: Kaplan-Meier estimate of the survival function with and without the presence of covariates history of obesity, RA, rhinitis and UC.

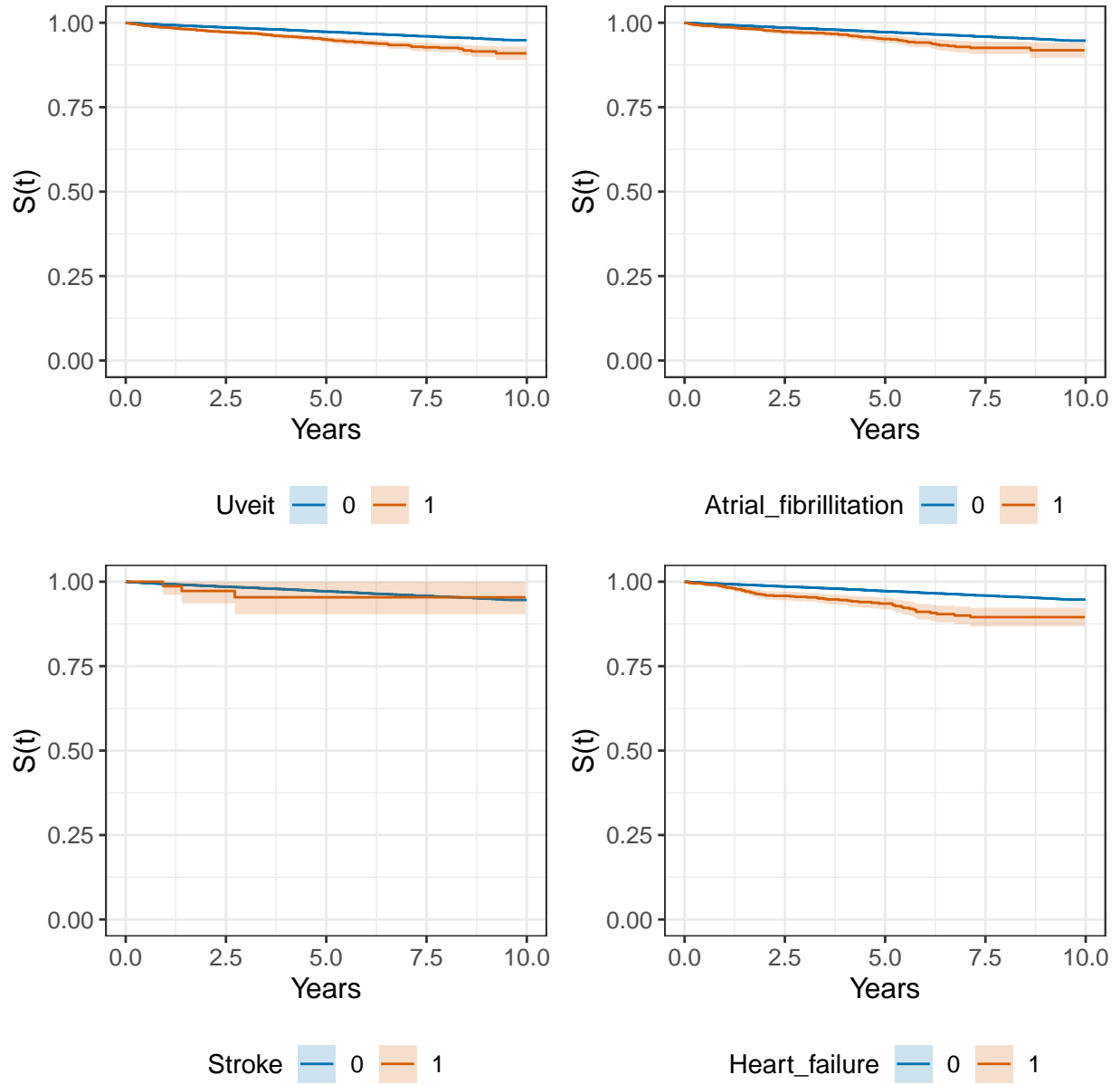


Figure 16: Kaplan-Meier estimate of the survival function with and without the presence of covariates history of uveit, atrial fibrillation, stroke and heart failure.

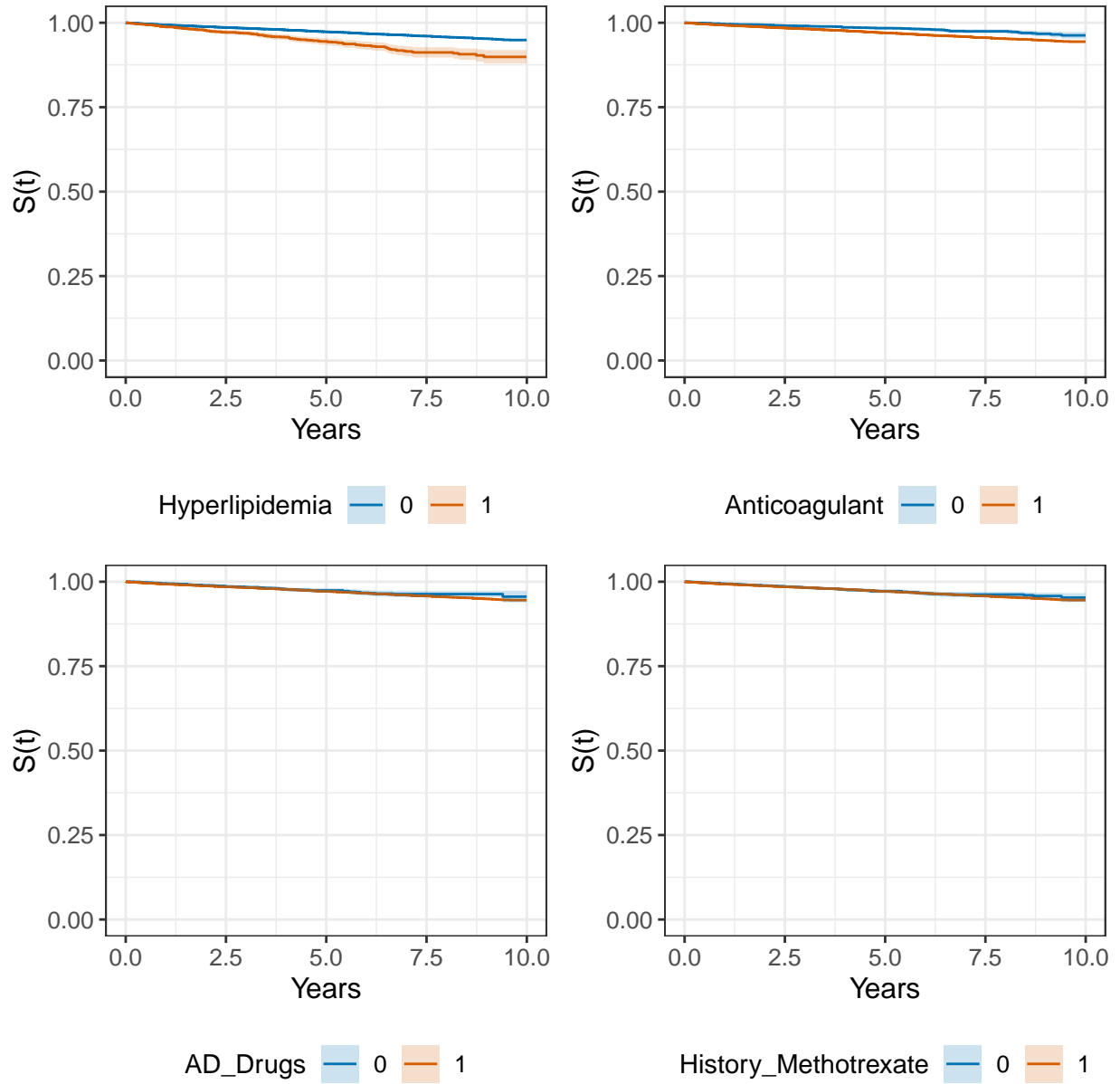


Figure 17: Kaplan-Meier estimate of the survival function with and without the presence of covariates history of hyperlipidemia, anticoagulant, AD drugs and methotrexate.



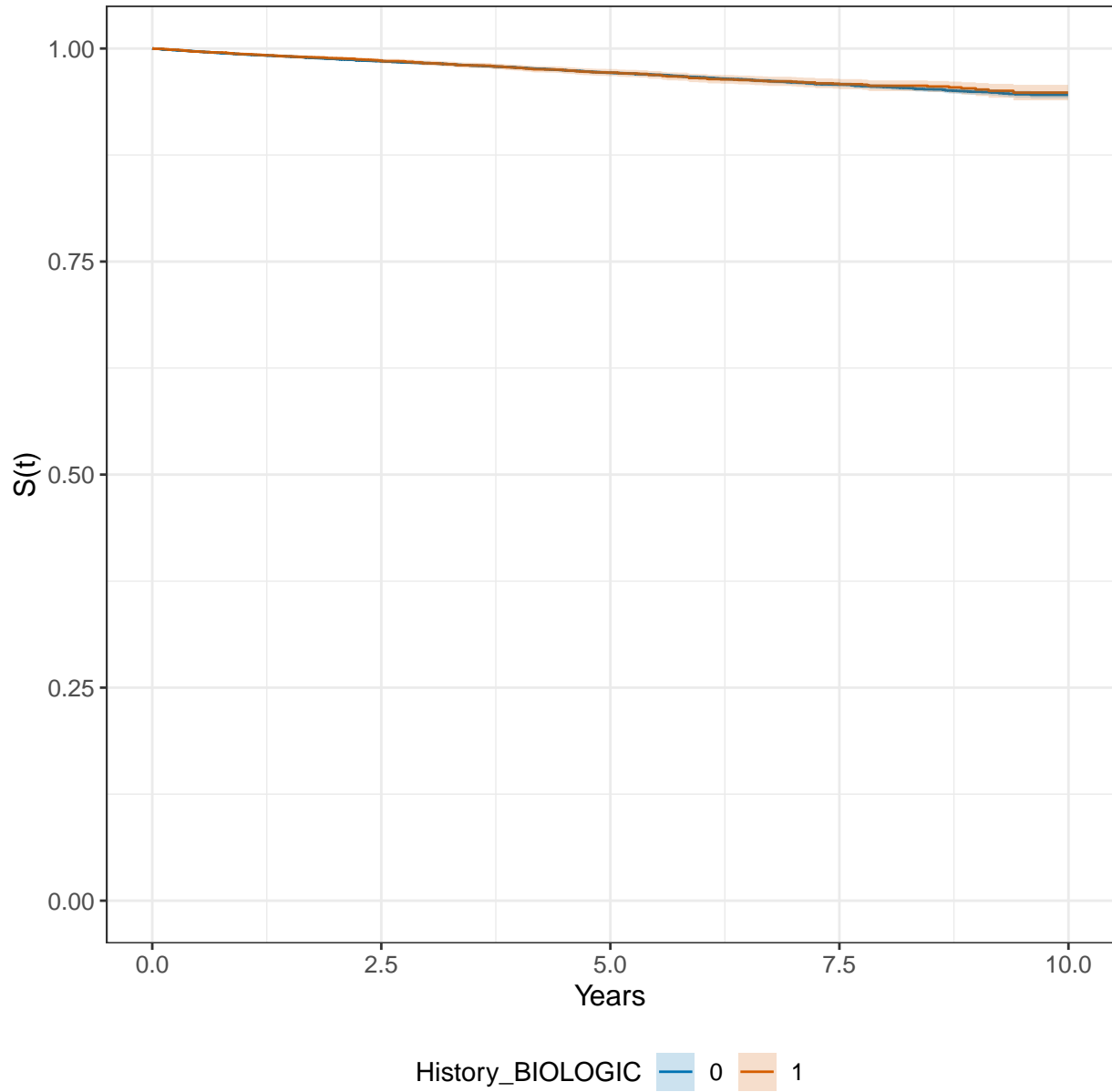


Figure 18: Kaplan-Meier estimate of the survival function with and without the presence of covariate history of biological medicine.

## 7.2 Log-minus-log survival plots

These are the log-minus-log survival plots for covariates that were not shown in Section 4.3. The theory for them are described in Section 3.4.1.

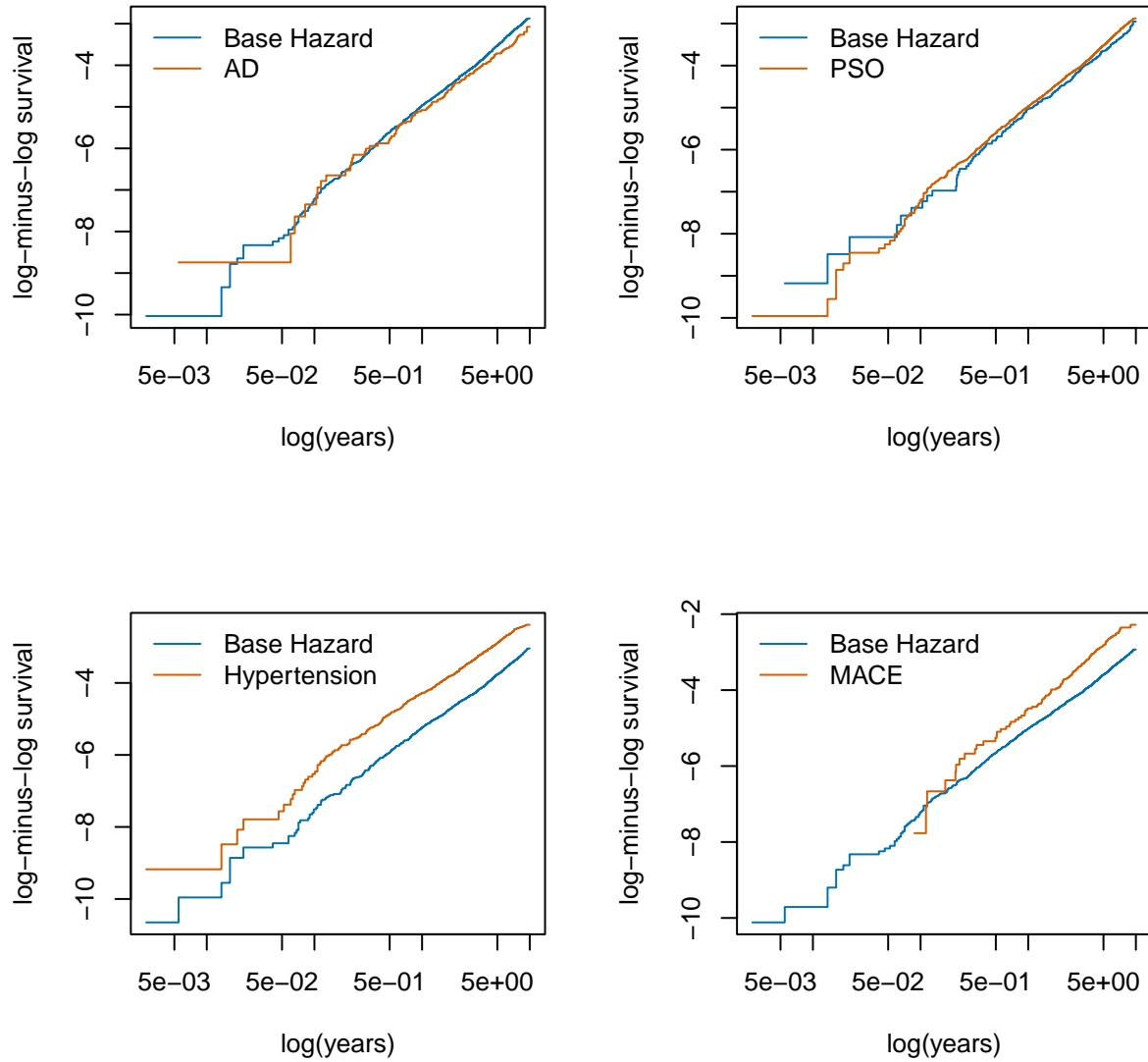


Figure 19: Log-minus-log transformation of the Kaplan-Meier estimate of the survival function with and without the presence of covariates history of AD, PSO, hypertension and MACE.

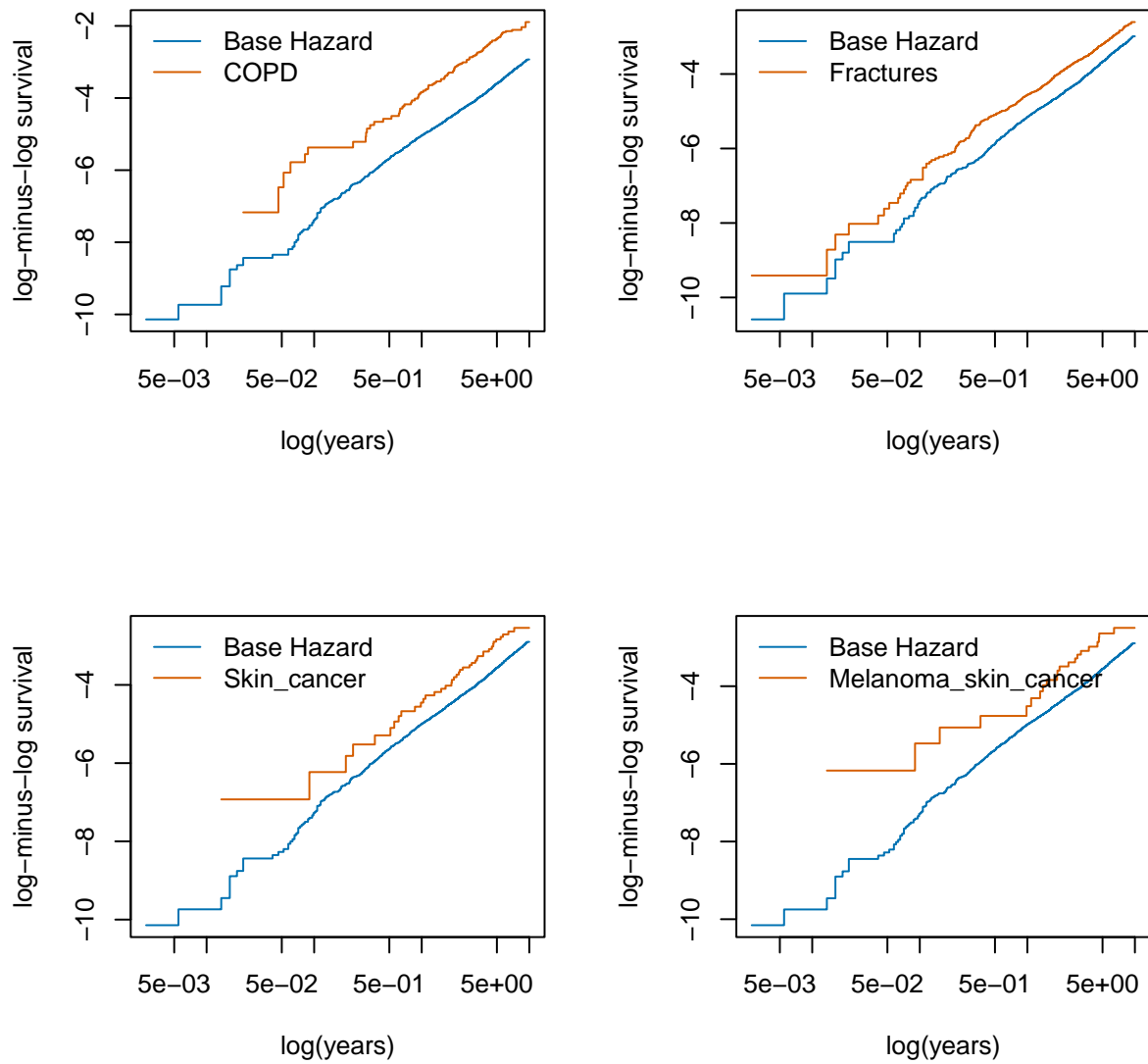


Figure 20: Log-minus-log transformation of the Kaplan-Meier estimate of the survival function with and without the presence of covariates history of COPD, fractures, skin cancer and melanoma skin cancer.

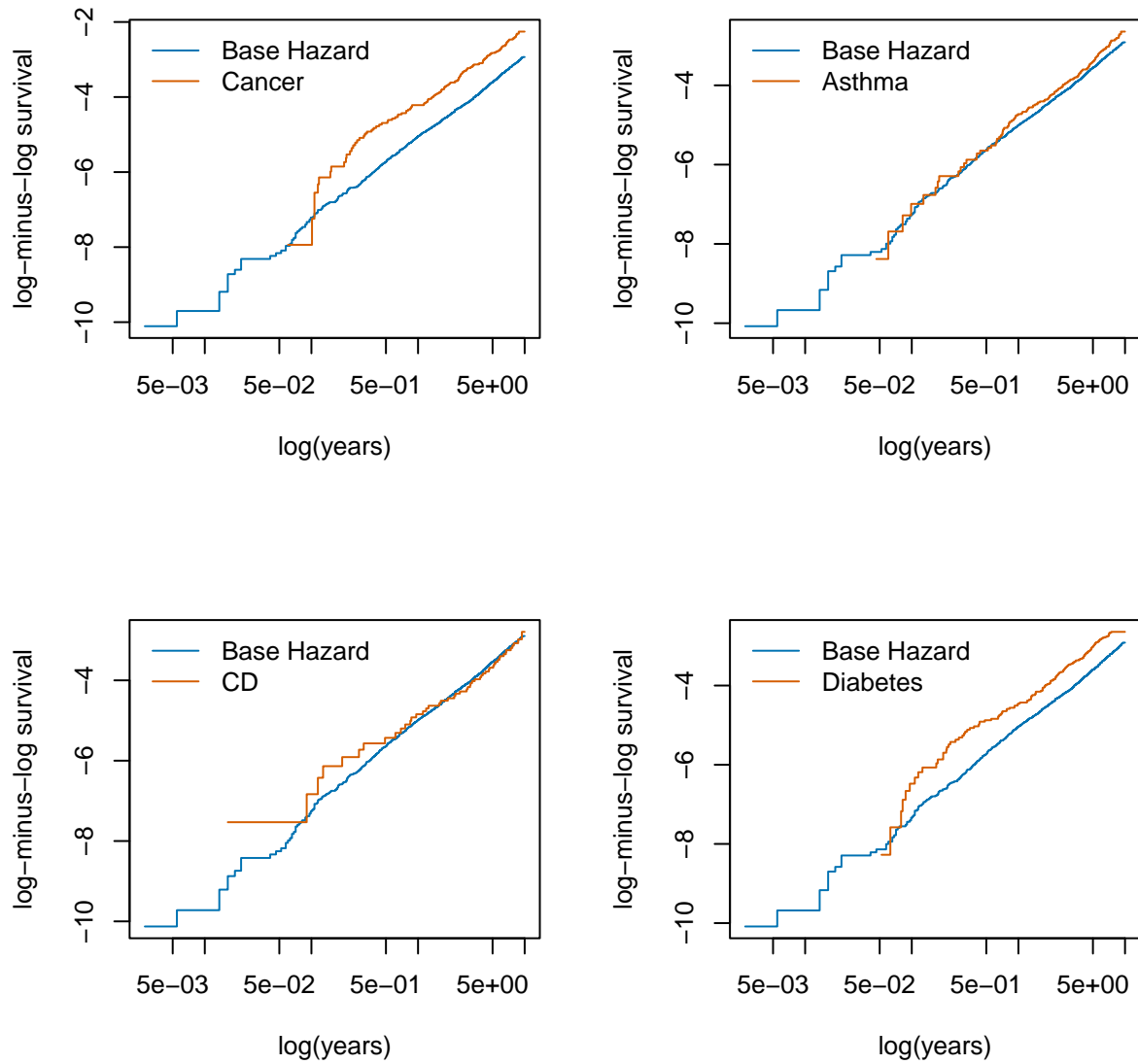


Figure 21: Log-minus-log transformation of the Kaplan-Meier estimate of the survival function with and without the presence of covariates history of cancer, asthma, CD and diabetes.

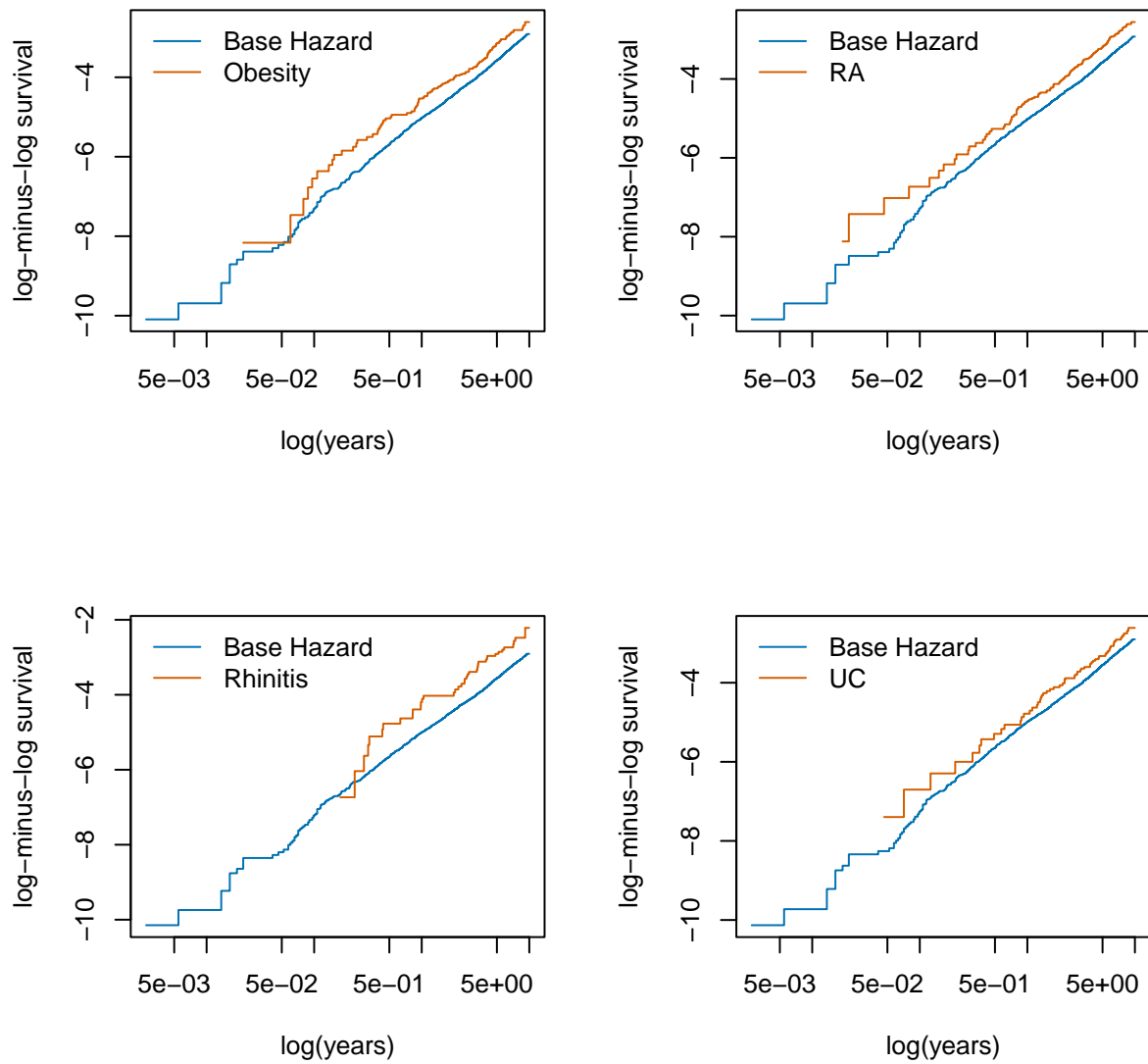


Figure 22: Log-minus-log transformation of the Kaplan-Meier estimate of the survival function with and without the presence of covariates history of obesity, RA, rhinitis and UC

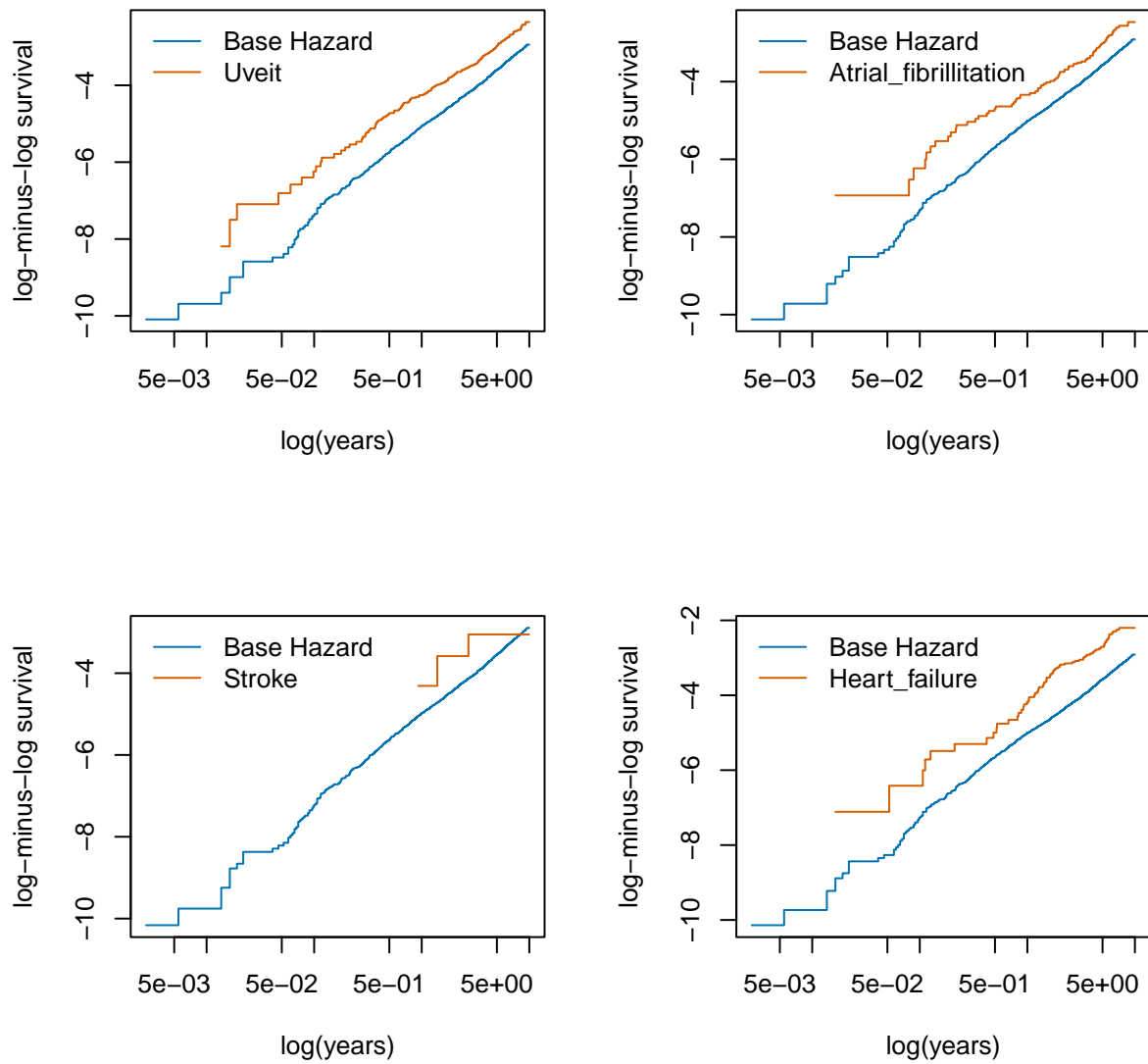


Figure 23: Log-minus-log transformation of the Kaplan-Meier estimate of the survival function with and without the presence of covariates history of uveit, atrial fibrillation, stroke and heart failure.

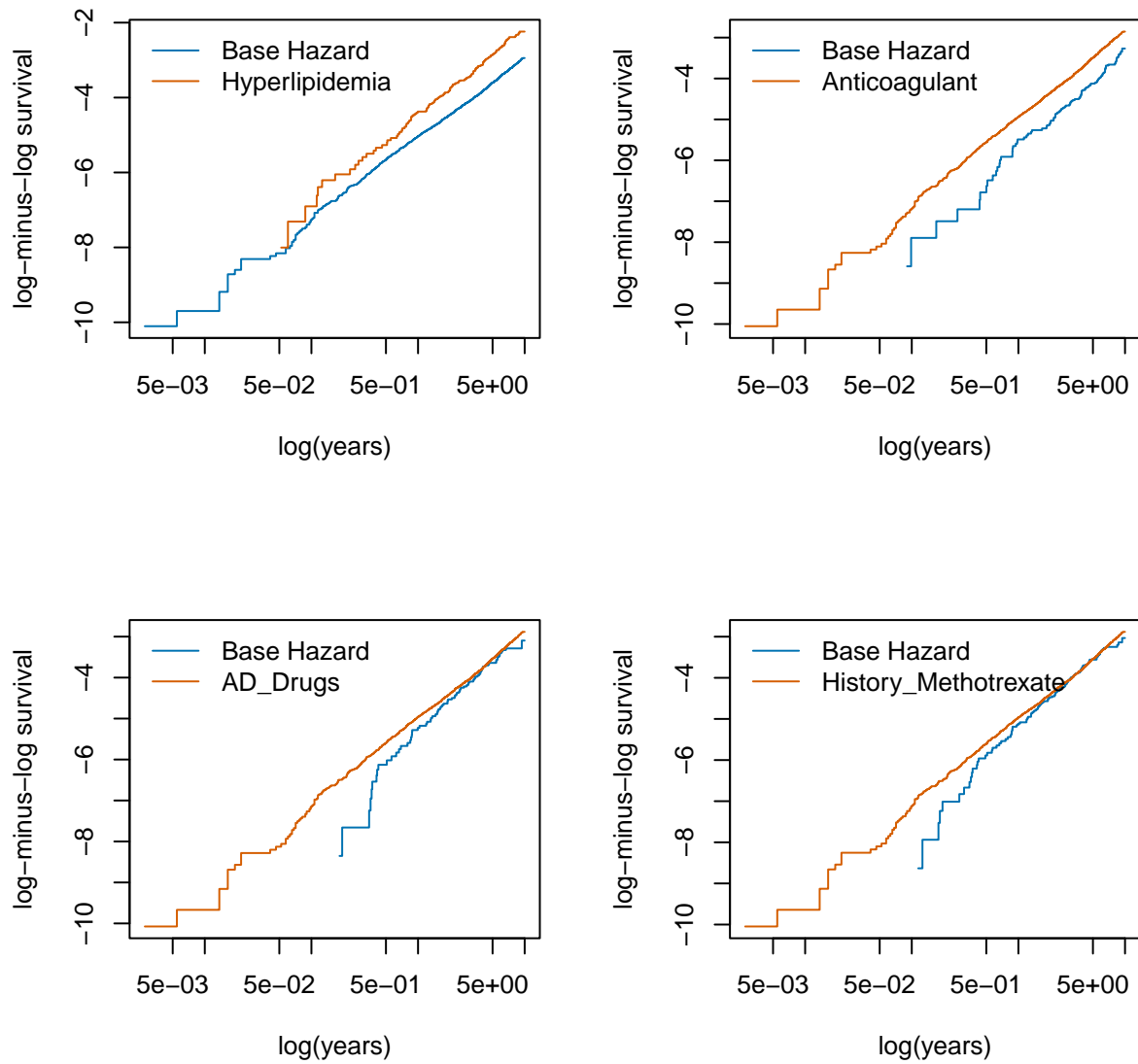


Figure 24: Log-minus-log transformation of the Kaplan-Meier estimate of the survival function with and without the presence of covariates history of hyperlipidemia, anticoagulants, AD drugs and methotrexate

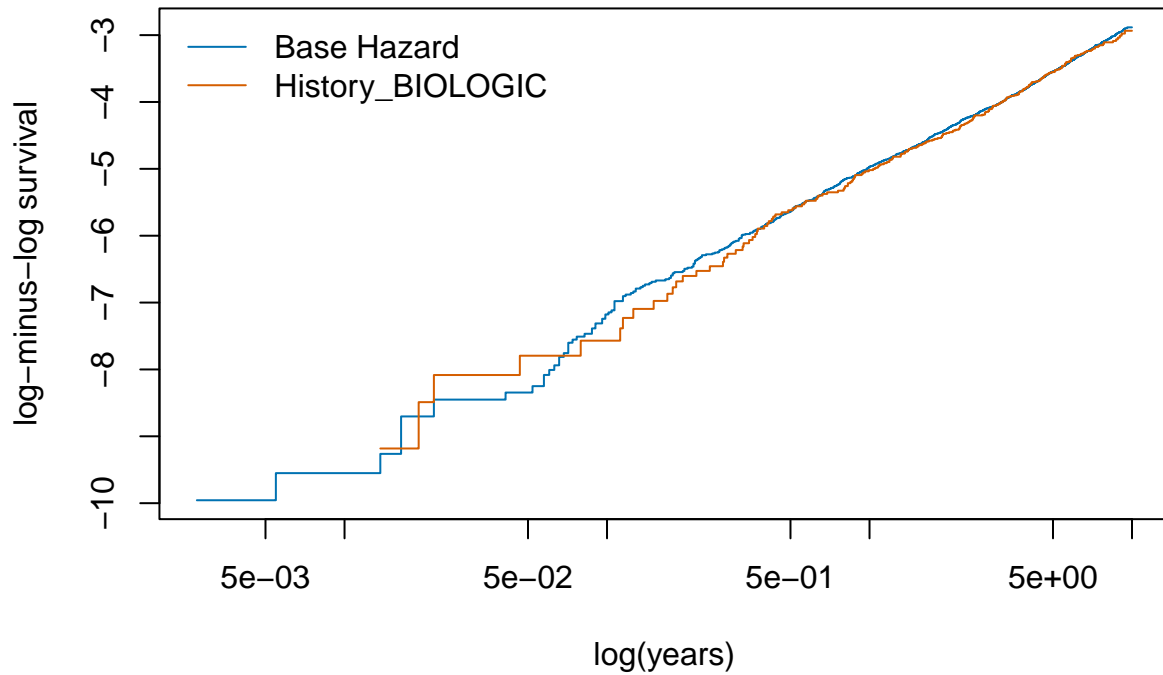


Figure 25: Log-minus-log transformation of the Kaplan-Meier estimate of the survival function with and without the presence of covariate history of biological medicine.

### 7.3 Schoenfeld residual graphs

These are the Schoenfeld residuals for the covariates that were not shown in Section 4.4.2. The Schoenfeld residuals are a graphical test for non-proportional hazards and are described in Section 3.4.2.



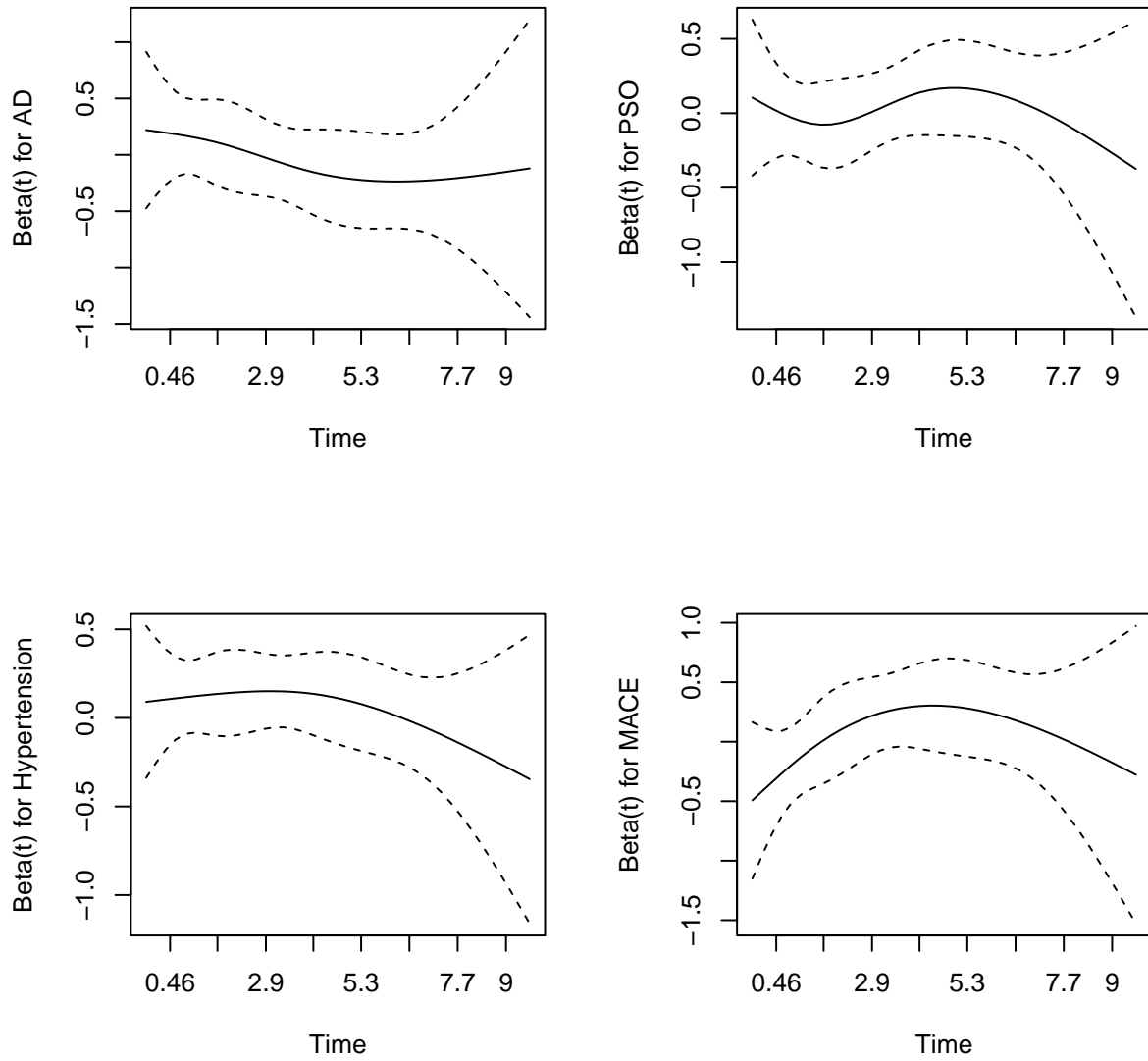


Figure 26: Schoenfeld residuals for the Cox regression estimates for the covariates history of AD, PSO, Hypertension and MACE.

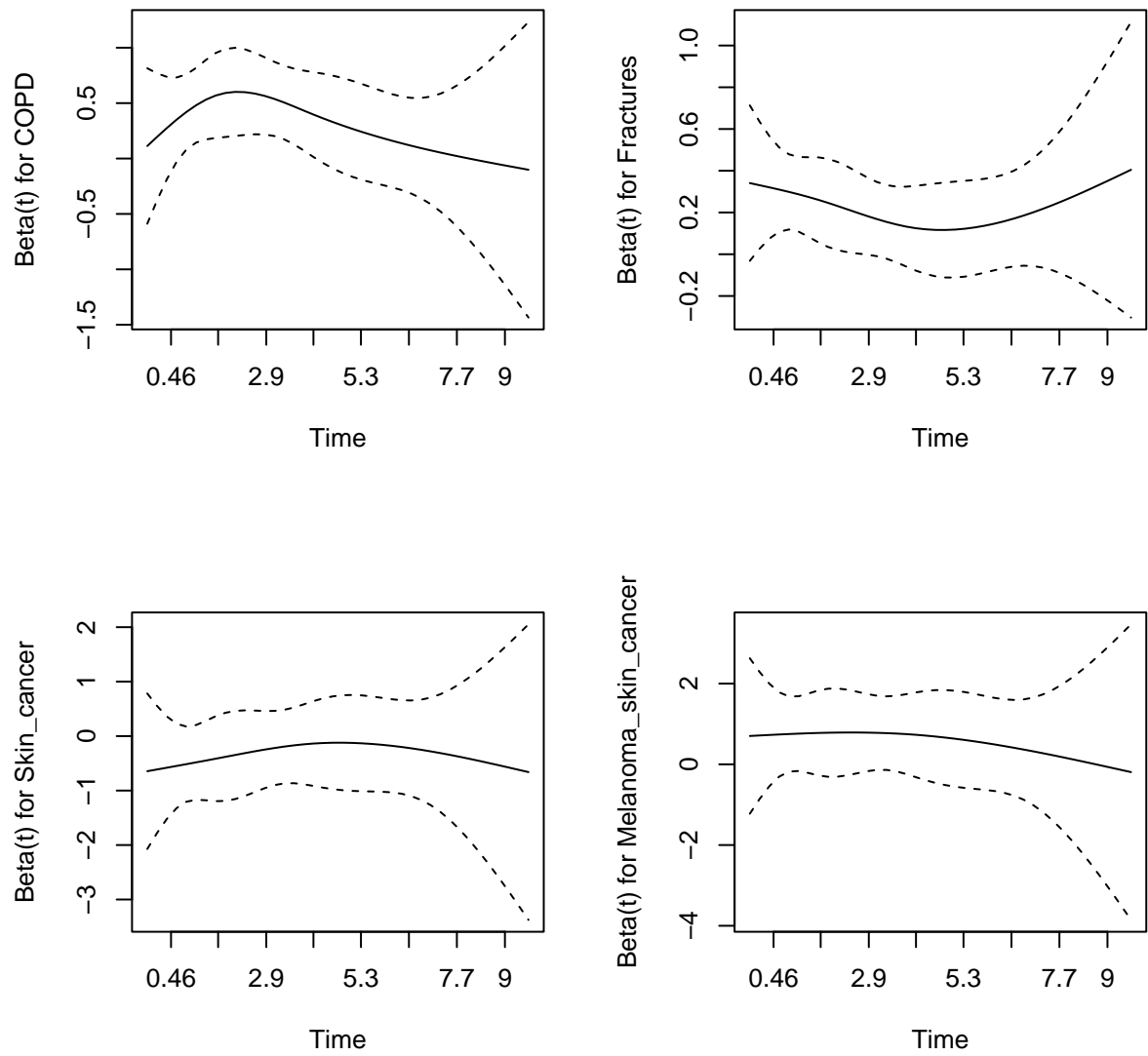


Figure 27: Schoenfeld residuals for the Cox regression estimates for the covariates history of COPD , fractures, skin cancer and melanoma skin cancer.

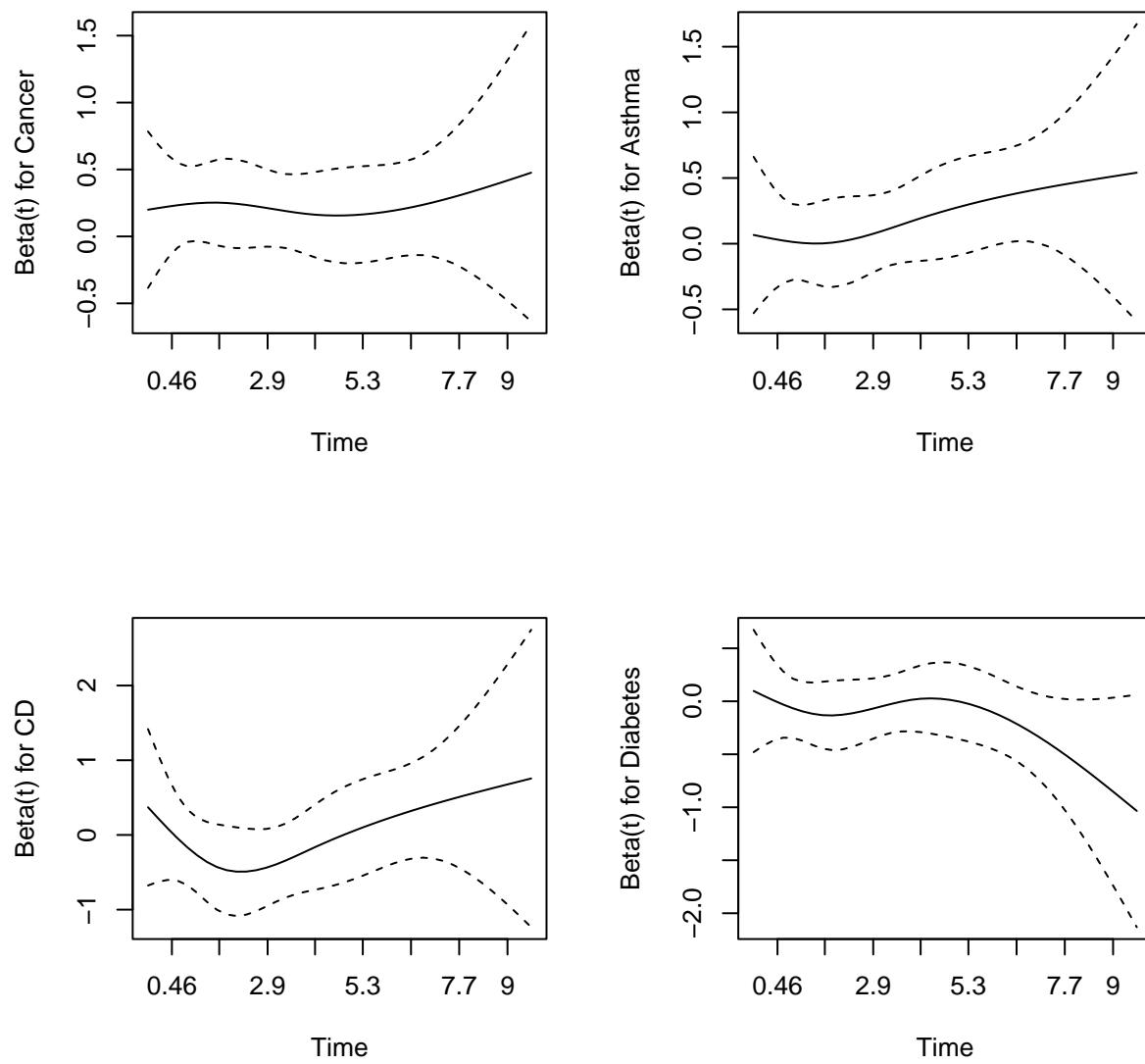


Figure 28: Schoenfeld residuals for the Cox regression estimates for the covariates history of cancer, asthma, CD and diabetes.

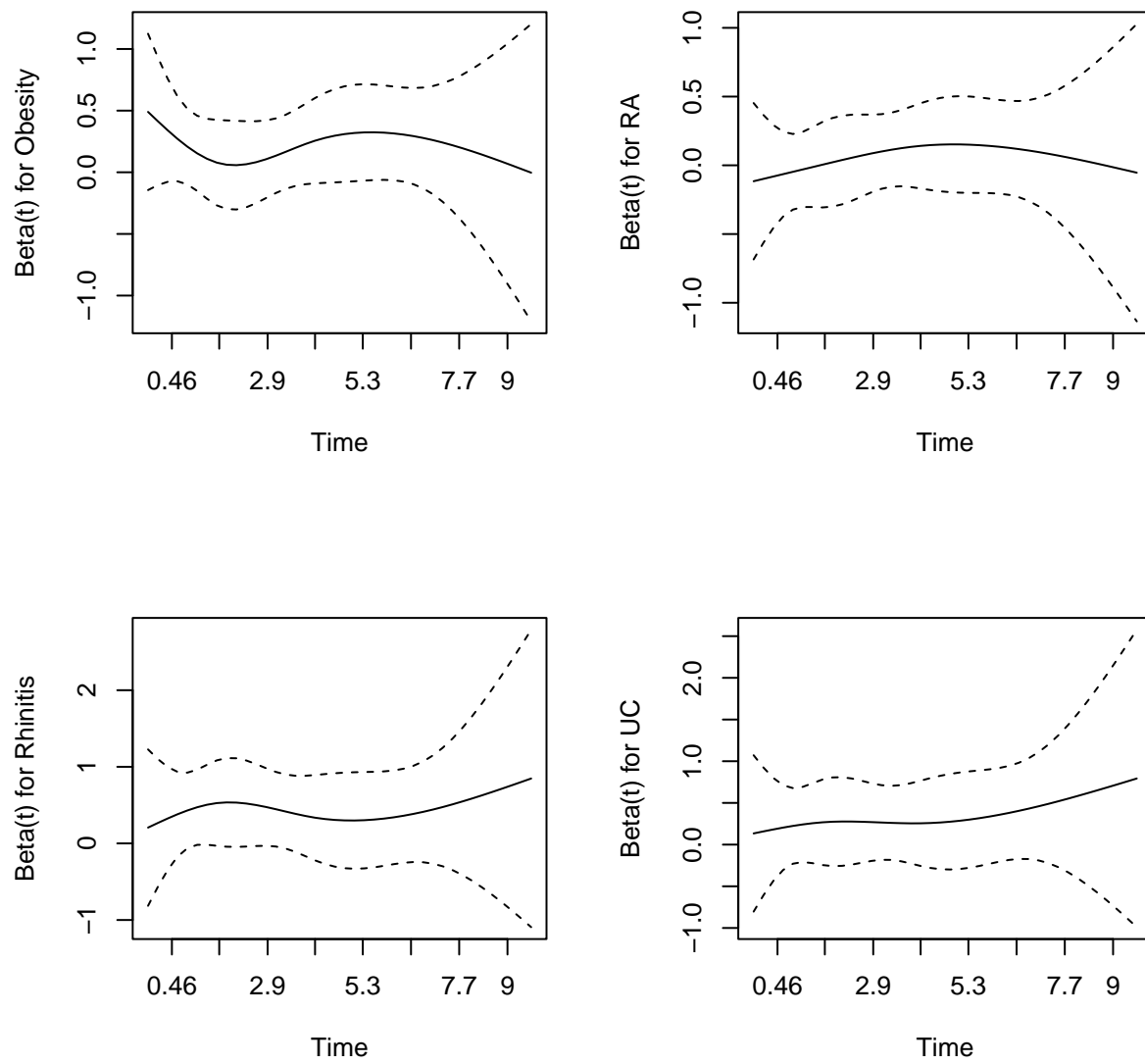


Figure 29: Schoenfeld residuals for the Cox regression estimates for the covariates history of obesity, RA, rhinitis and UC.

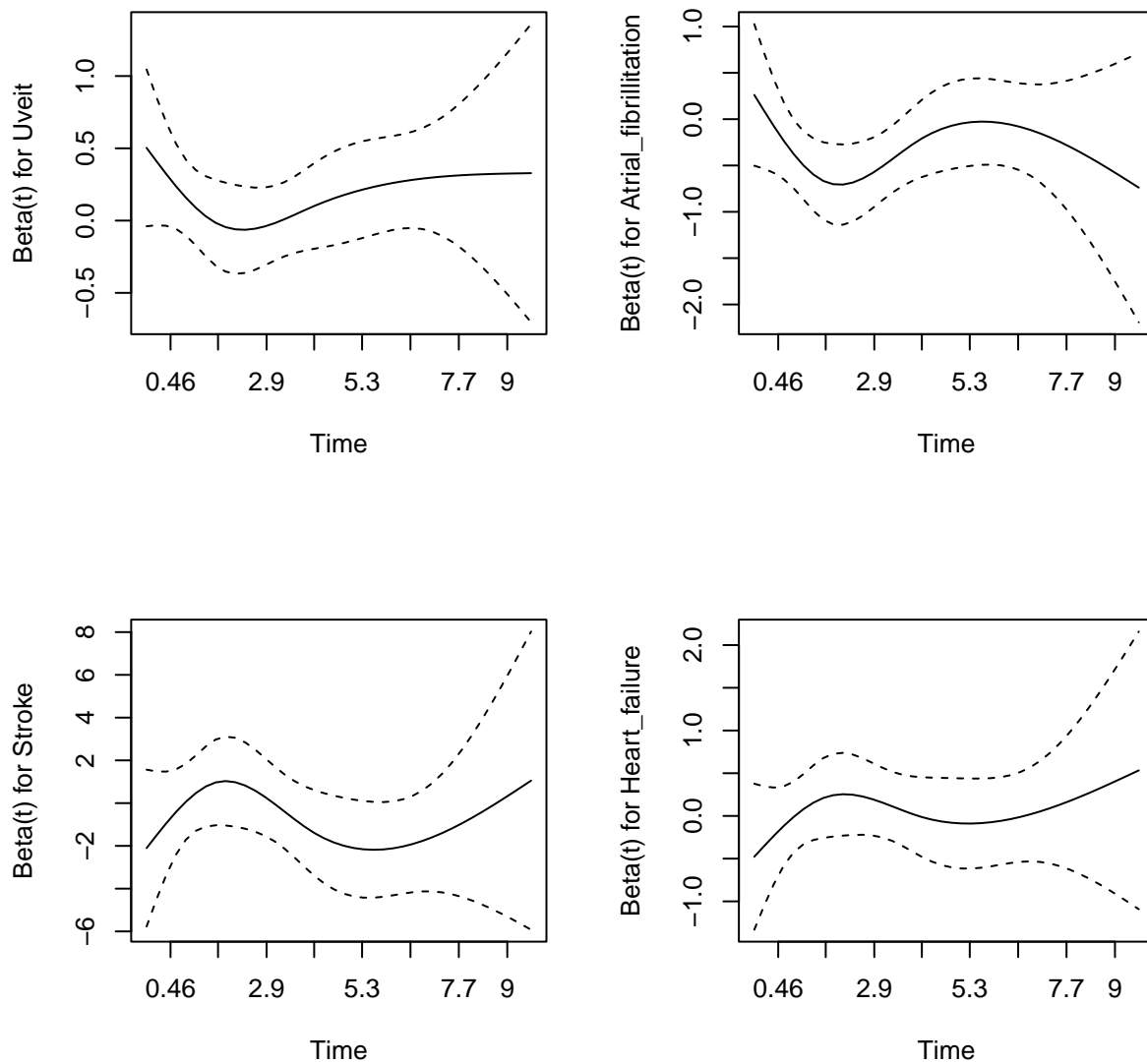


Figure 30: Schoenfeld residuals for the Cox regression estimates for the covariates history of Uveit, atrial fibrillation, stroke and heart failure

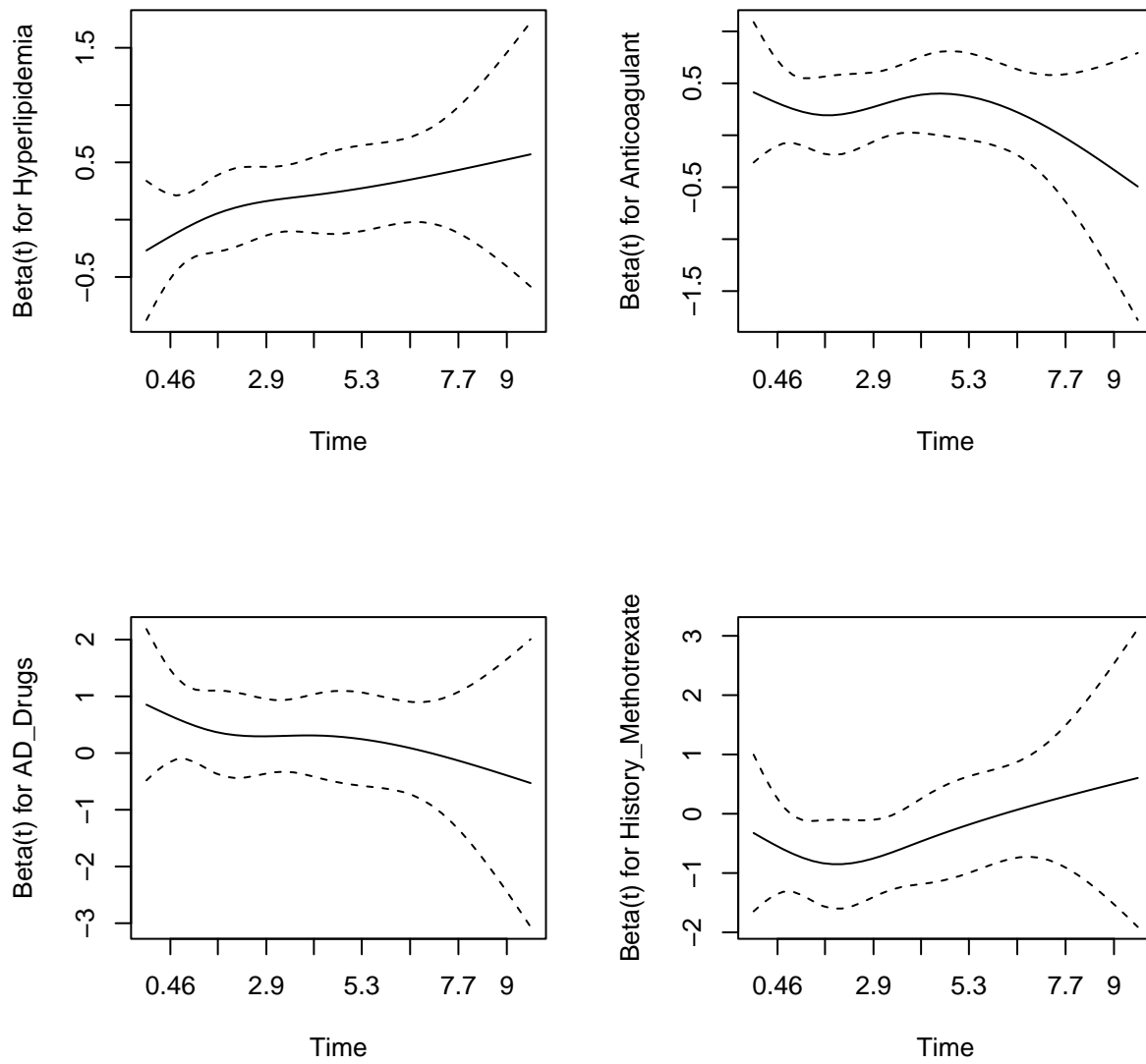


Figure 31: Schoenfeld residuals for the Cox regression estimates for the covariates history of hyperlipidemia, anticoagulants, AD drugs and methotrexate.

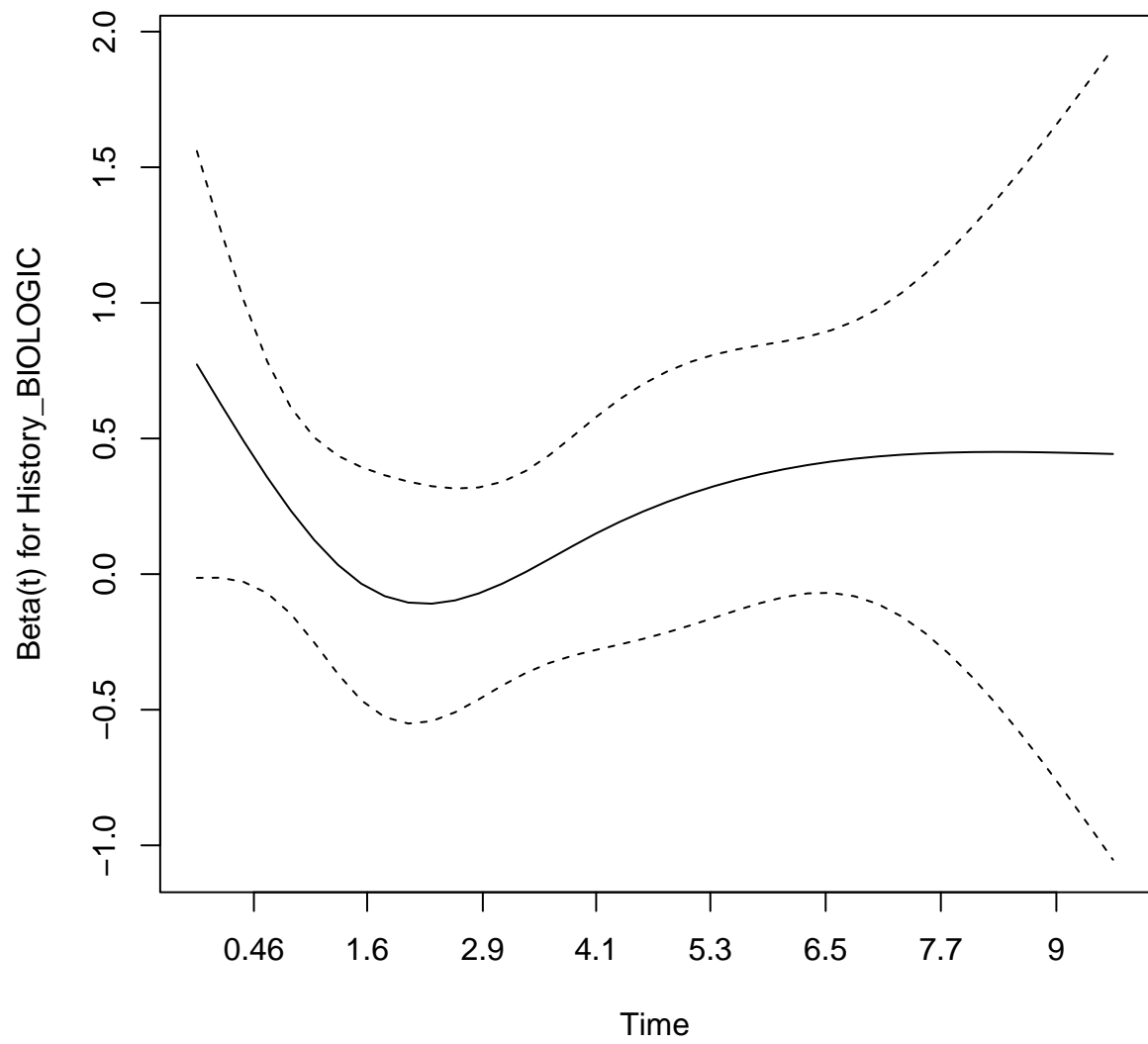


Figure 32: Schoenfeld residuals for the Cox regression estimates for the covariate history of biologic medicine.

## 7.4 Hazard rate graphs

These are the FPM estimated hazard rates for all the variables that were not shown in the results. Figures 33 through 39 shows the hazard rates. They are all proportional by assumption of the model.

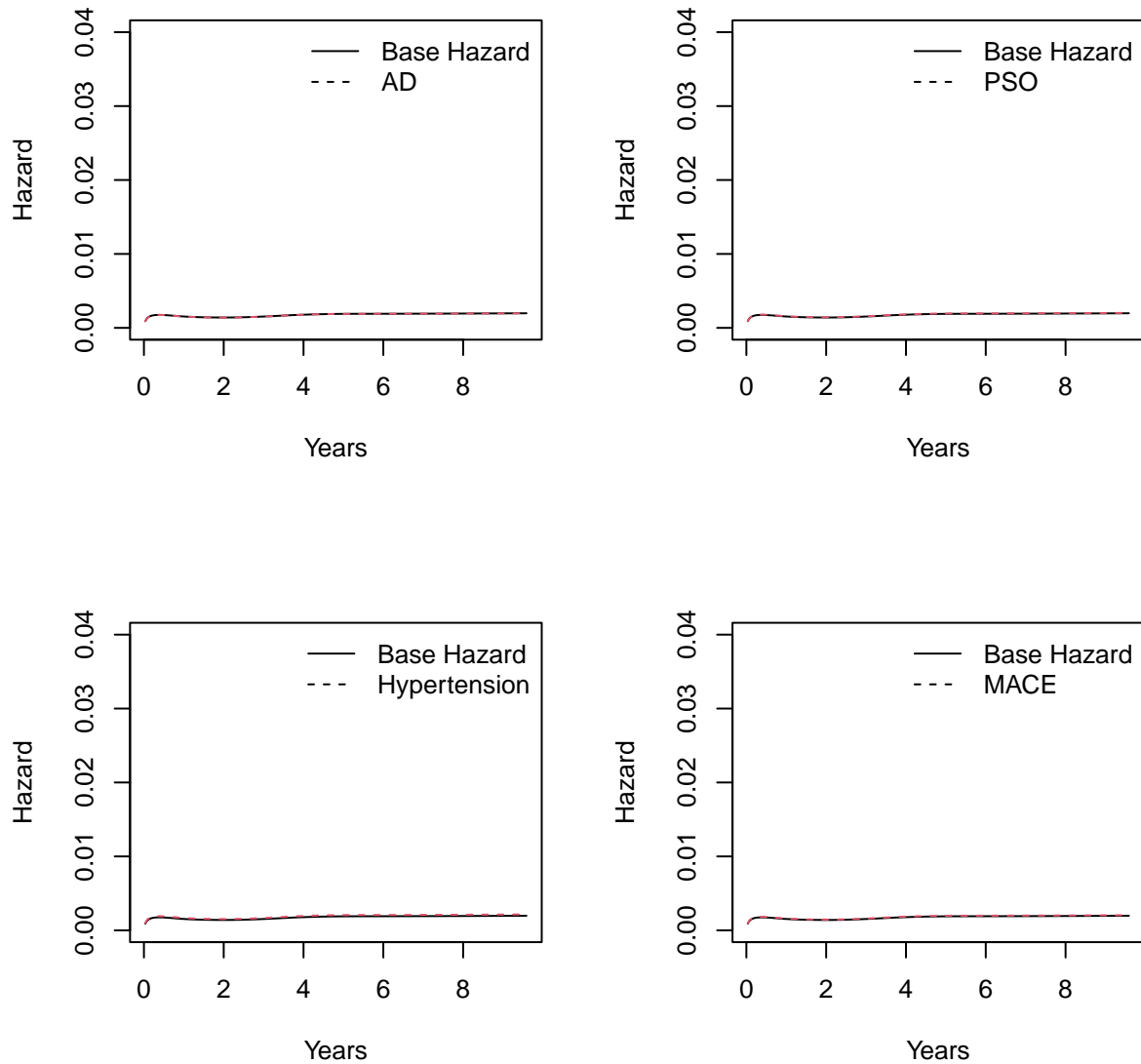


Figure 33: FPM estimated hazard curve for the covariates history of AD, PSO, hypertension and MACE



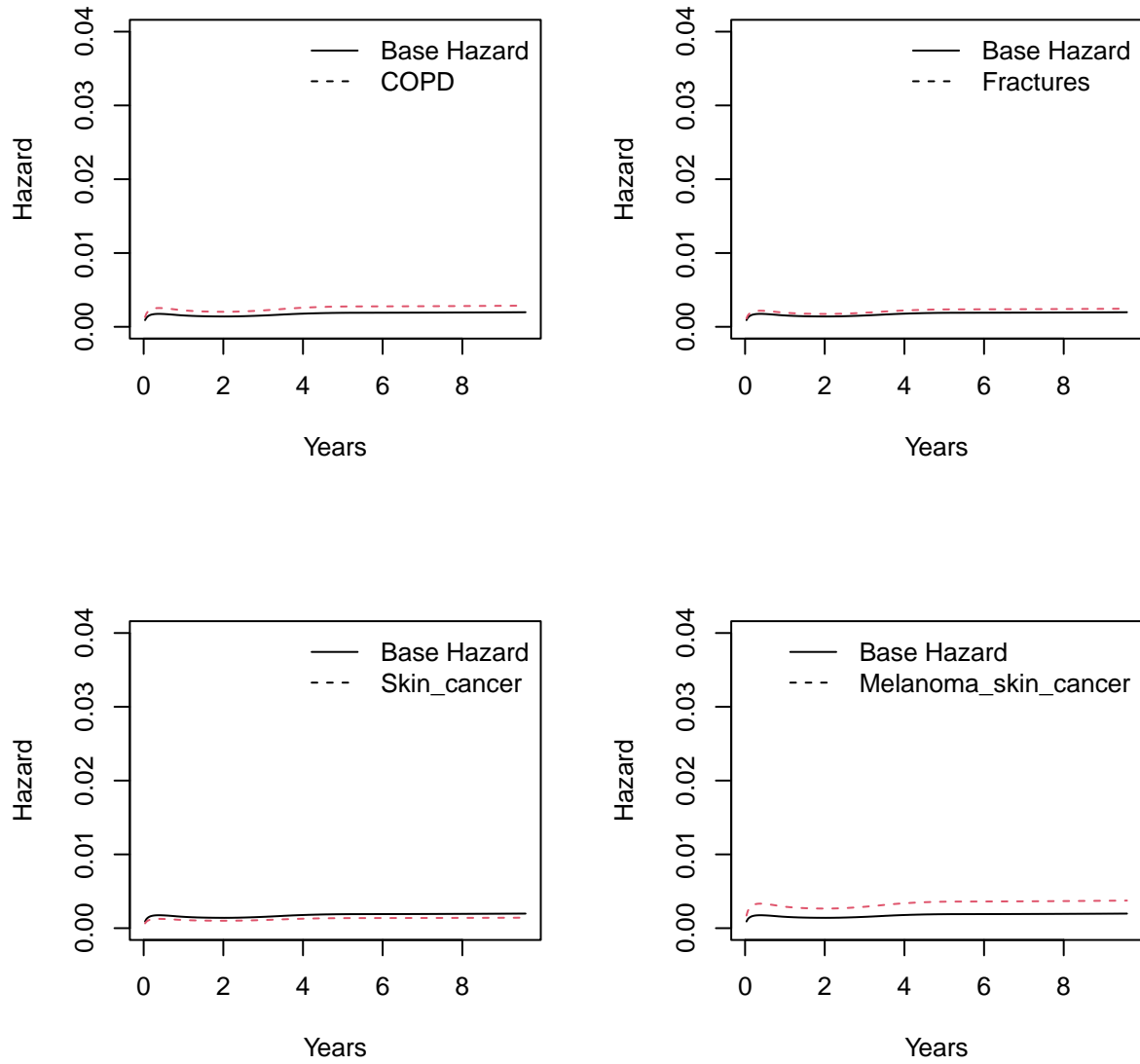


Figure 34: FPM estimated hazard curve for the covariates history of COPD, fractures, skin cancer and melanoma skin cancer.

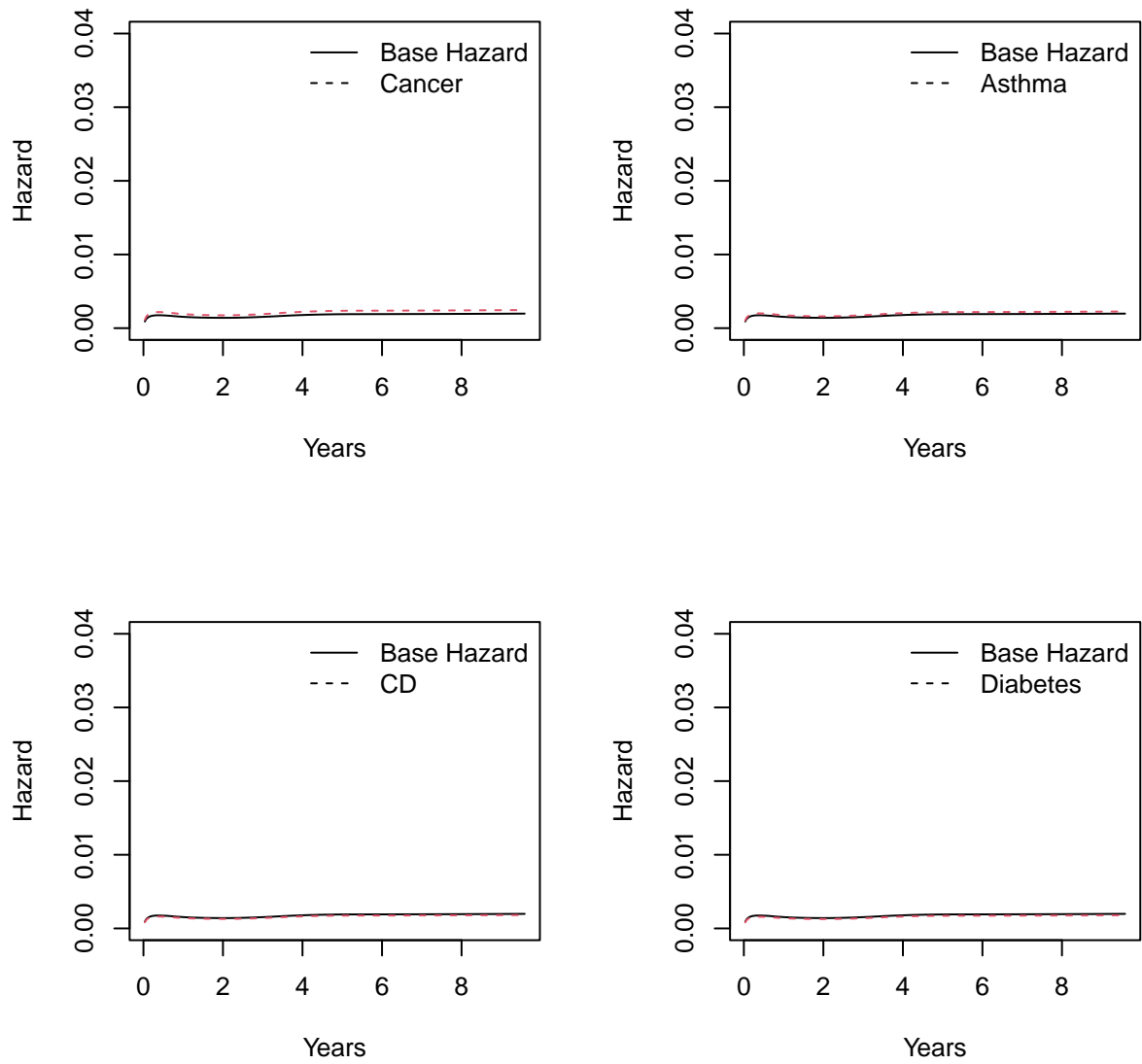


Figure 35: FPM estimated hazard curve for the covariates history of cancer, asthma, CD and diabetes

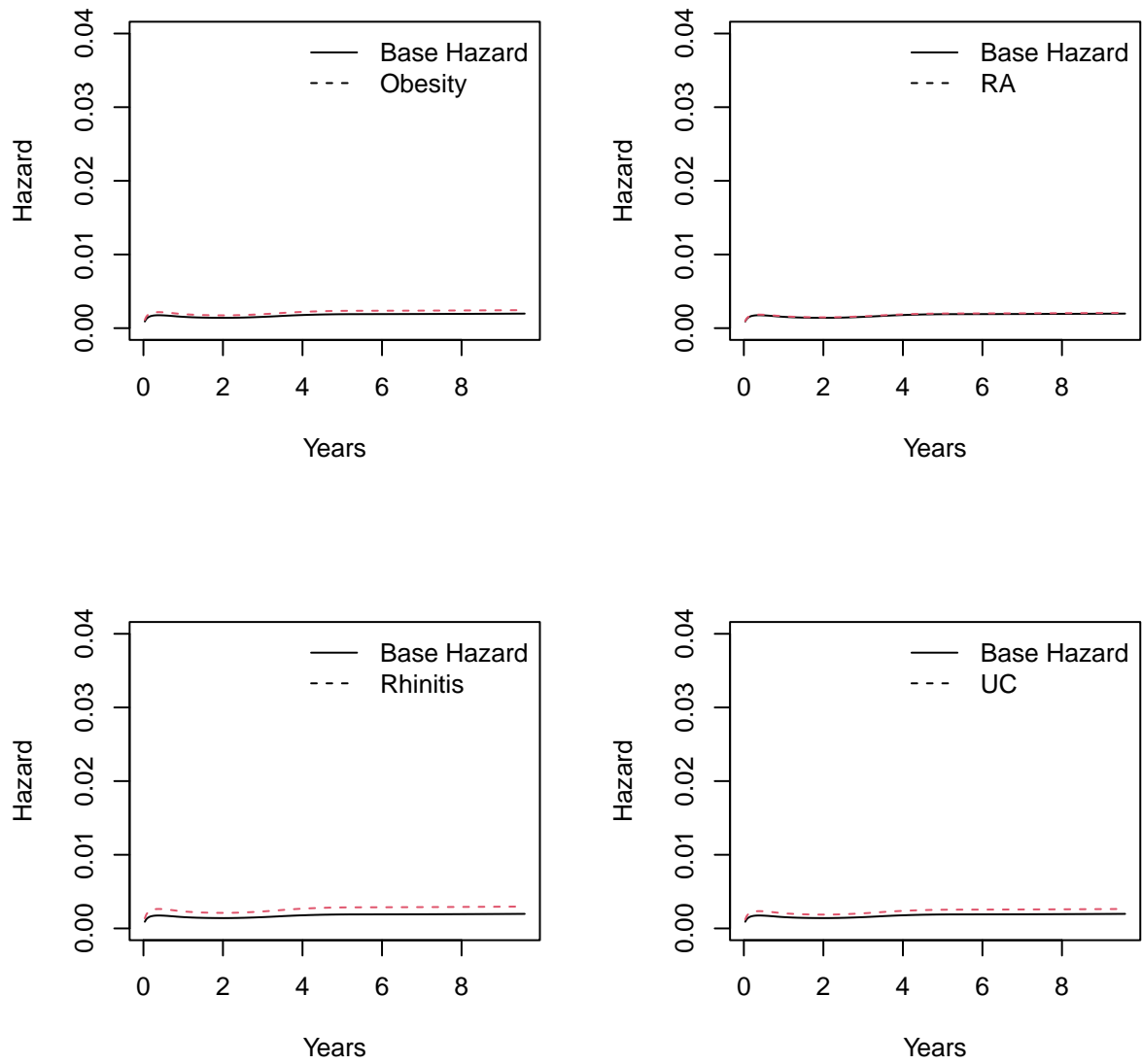


Figure 36: FPM estimated hazard curve for the covariates history of obesity, RA, Rhinitis and UC.

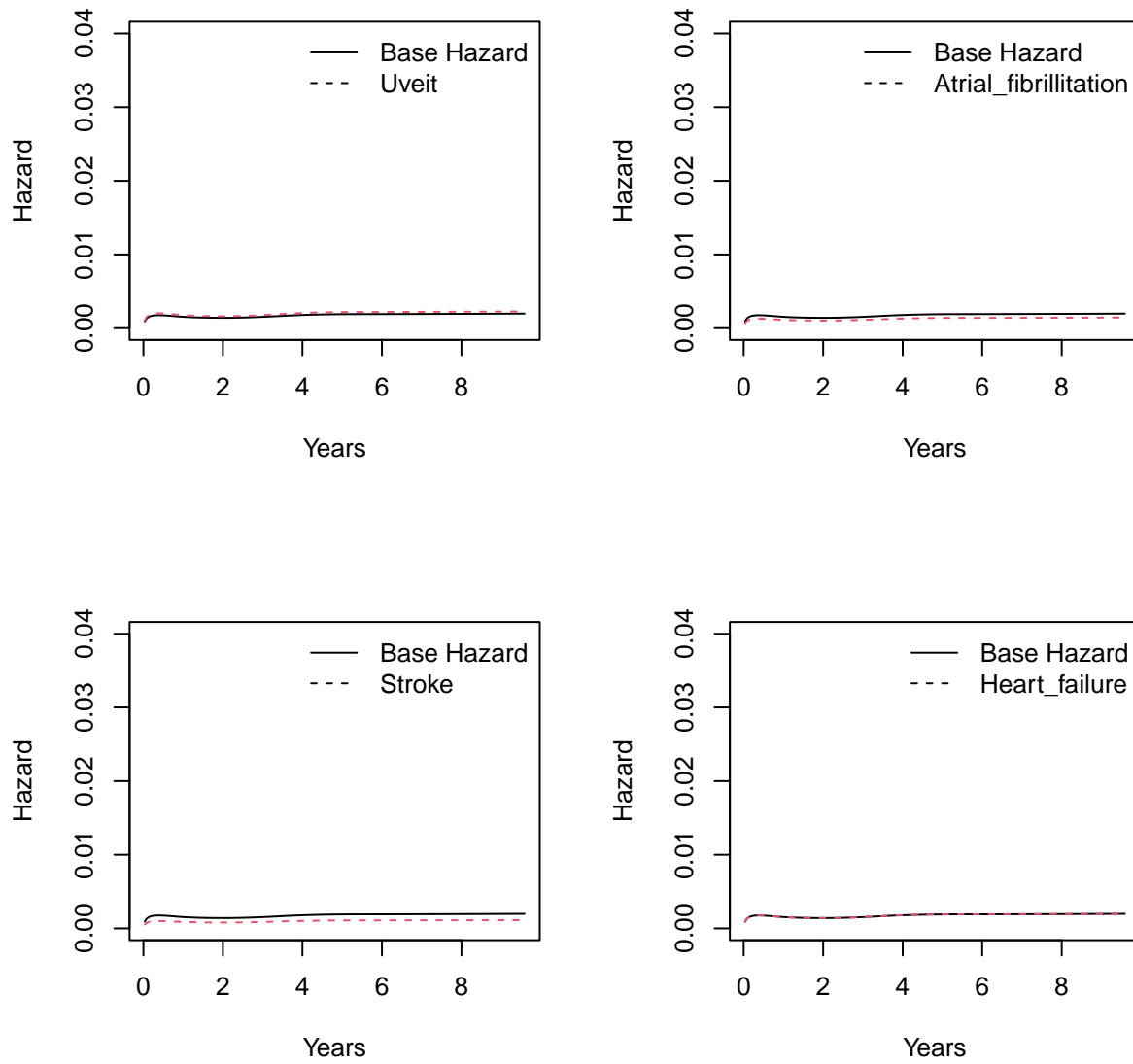


Figure 37: FPM estimated hazard curve for the covariates history of uveit, atrial fibrillation, stroke and heart failure.

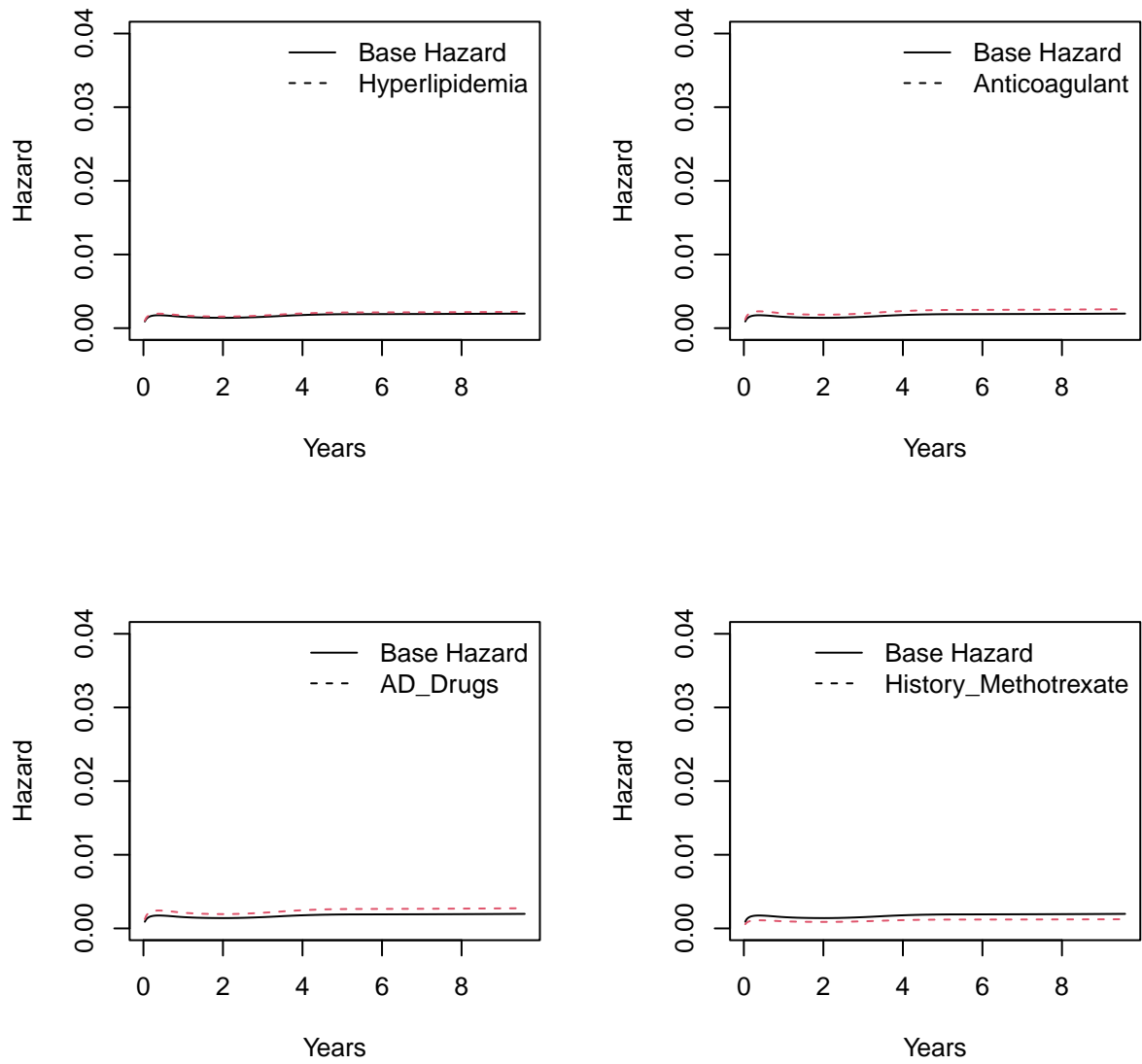


Figure 38: FPM estimated hazard curve for the covariates history of hyperlipidemia, anticoagulants, AD Drugs and Methotrexate.

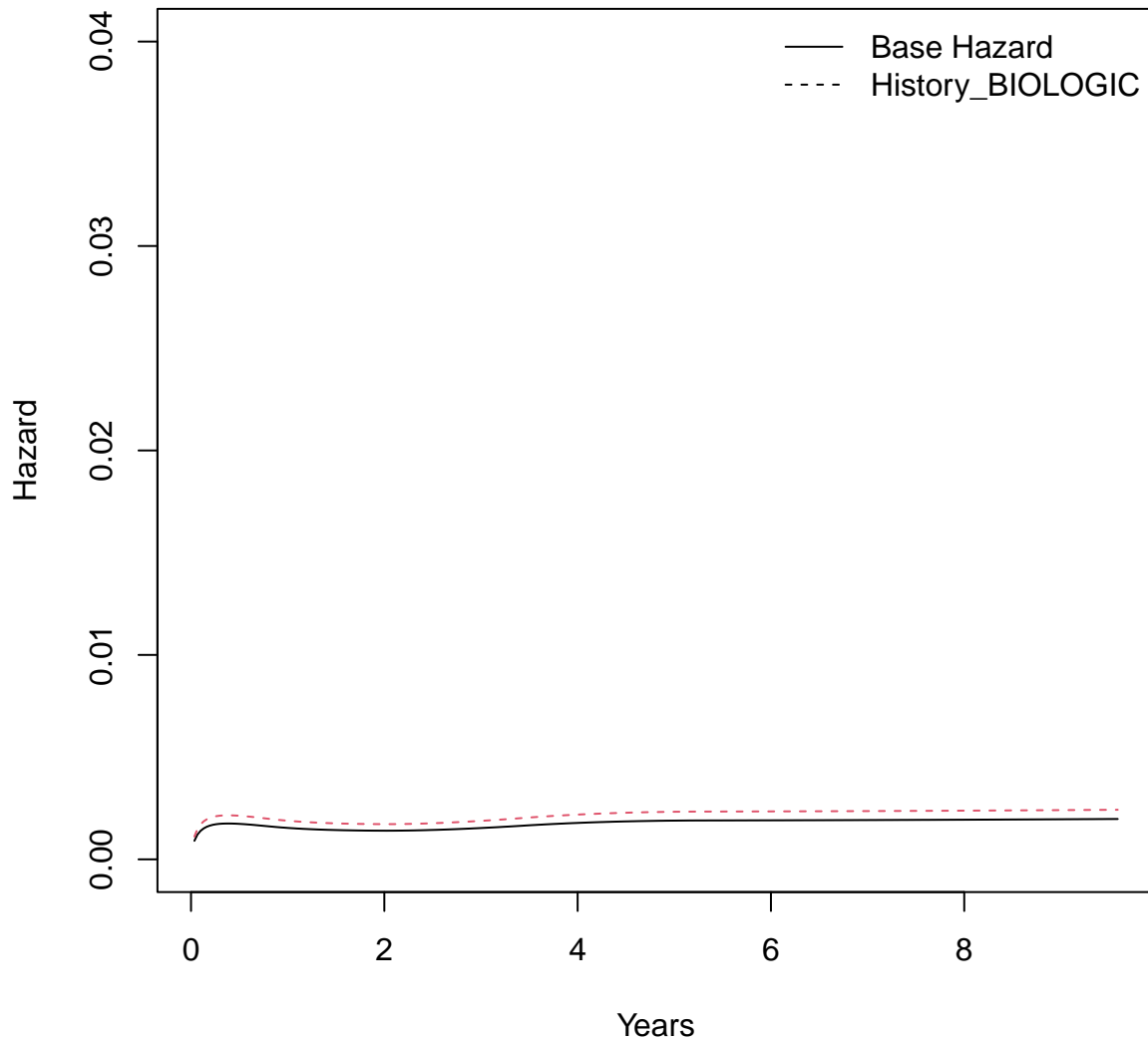


Figure 39: FPM estimated hazard curve for the covariate history of use of biological medicine.

## 7.5 Code

All the code for the modeling and simulation can be found at this github repository.

<https://github.com/Martin-Hans-Erik-Andersson/Surviving-non-proportional-hazards-with-the-flexible-parametric-model.git>

The data used is not public and as such is not included.