Stockholms universitet

Extrapolation of Survival Curves with an Application to Multiple Myeloma

Felix Wahl

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2016:2
Matematisk statistik
Juni 2016

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematiska institutionen

Mathematical Statistics
Stockholm University
Master Thesis **2016:2**
http://www.math.su.se

Extrapolation of Survival Curves with an Application to Multiple Myeloma

Felix Wahl*

June 2016

Abstract

Modeling of censored survival data has its difficulties. Especially when the interest lies in finding the mean survival time, making extrapolation necessary. In this master thesis the performance of some standard parametric models and the flexible parametric survival models of Royston Parmar (2002) are examined in the context of survival analysis. The performance is here defined as the ability to capture the true mean survival time. Furthermore, a variety of survival analysis methods is presented, involving non-parametric methods, modeling with covariates, etc. At our disposal we have a Swedish dataset containing registry data of 1606 multiple myeloma patients between January, 2000 and November, 2011. This data will first of all be used as a basis for a simulation study, where the performance of the standard parametric models and the flexible parametric survival models will be investigated. Secondly, part of this data will be modeled and conclusions from the simulation study regarding the amount of follow-up time needed will be tested. The simulation study yields some evidence that the flexible parametric survival models can help diminish the large biases one gets when modeling the data with a misspecified model. This holds even when they themselves are misspecified. Since the major problem with limited follow-up is the difficulty in fitting correctly specified models, this is a helpful result.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: felixwahl91@gmail.com. Supervisor: Mathias Lindholm.

Acknowledgements

I want to thank Amy Leval and Johan Liwing for giving me the opportunity to write my thesis at Janssen and for discussions and support throughout this thesis. Thank you also everyone at Janssen for giving me a great few months. I also want to thank my supervisor Mathias Lindholm at Stockholm University for the many hours of fun discussions and for valuable feedback and input. Finally I want to thank my wife Hanna Fues Wahl for reading this thesis in all of its versions and giving me helpful comments.

Contents

Abbreviations

AIC ..... Akaike information criterion

BIC ..... Bayesian information criterion

CDF ..... Cumulative distribution function

CI ..... Confidence interval

hdt ..... High-dose chemotherapy

MLE ..... Maximum likelihood estimator

MM ..... Multiple myeloma

OS ..... Overall survival

PDF ..... Probability density function

RP ..... Royston-Parmar

TTNT ..... Time to next treatment

Chapter 1

Introduction

In health economic modelling one is often concerned with modeling cost effectiveness to be able to make rational decisions in health policy. Cost effectiveness is studied by comparing, for instance, a treatments cost against its efficacy. To be able to say something about cost effectiveness one needs to have information about the indefinite future. In clinical trials there are limited follow-up times making it necessary to try to extrapolate our resulting knowledge into the future. This is not an easy task, how can we say something about what we have never observed.

In the above situation we use survival analysis, where interest lies in modeling time-to-events. An event could for example be death or the progression of a disease. The most commonly used method is to impose a distributional assumption on the data and model the observed part as well as possible. In most cases, when patient-level data from clinical trials is modeled, an exponential or Weibull distribution is used (Latimer, 2011). When treatment effects are of interest, mostly the Cox proportional hazards model is used. This is a model that has no distributional assumption but relies on the quite restrictive assumption that the rate at which events happen is proportional, by the same constant, between different values of a covariate over time. If this assumption does not hold we could stratify our data, but that would lead to a relatively large loss of power. There are other methods that might be promising but are not commonly used in health economic modelling of time-to-events. One of these are the flexible parametric survival models, also known as the Royston-Parmar models, developed by Royston & Parmar (2002). These are models that extend the standard parametric models by the use of splines, making them more flexible. Furthermore, with Royston-Parmar models explanatory variables can be incorporated in a quite simple way.

The aim of this thesis is to examine the performance of some of the standard parametric models and the Royston-Parmar models in terms of capturing the true mean time-to-event of the data generating distribution. We will also delve into some of the theory of survival analysis, which beyond the two methods above involves, among other, non-parametric estimation of survival curves and ways of modeling the effect of explanatory variables on the time-to-event.

At our disposal we have a Swedish dataset containing registry data of 1606 multiple myeloma

patients between January, 2000 and November, 2011. This data will first of all be used as a basis for a simulation study, where the performance of the standard parametric models and the Royston-Parmar models will be investigated. Secondly, part of this data will be modeled and conclusions from the simulation study will be tested. In the modelling of the data we will be concerned with modeling overall survival (OS) and time to next treatment (TTNT). OS is defined as the time from diagnosis until death and TTNT is defined as the time from one treatment start until the start of the next treatment or death. TTNT works as a proxy for disease progression and drug toxicity, which are more relevant outcome measures but not always available in registry data.

1.1 Outline

In Chapter 2 we go through the theory of survival analysis used in this thesis. In sections 2.1-2.5 we go through the basics of survival analysis. In Section 2.6 two nonparametric estimators are defined and in Section 2.7 parametric survival analysis is introduced. The Cox proportional hazards model is described in Section 2.8. The Royston-Parmar models are presented in Section 2.9 and in Section 2.10 a brief presentation of how one can connect counting processes and survival analysis is given. In Section 2.11 some other methods, not used in the computational part of this thesis, are mentioned. Finally, in Section 2.12 we discuss methods for model selection.

Chapter 3 describes the Swedish multiple myeloma dataset used in this thesis and some descriptive statistics are presented.

In Chapter 4 the computational methods of the thesis are presented. Section 4.1 describes the most important `R`-packages and functions used in the data handling and modeling. Then Section 4.2 goes on to present the simulation study and its procedure plus a simplified example is shown to illustrate how extrapolation performs in the best case scenario. Finally in Section 4.3 the results of the simulation study are examined.

The data analysis of the Swedish multiple myeloma dataset is found in Chapter 5. There the conclusions of the simulation study are tested on the data followed by some modeling using explanatory variables using both the Cox proportional hazards model and the Royston-Parmar models. The conclusions of the simulation study are also briefly tested on a survival dataset for patients with advanced lung cancer.

A summary and discussion of the thesis is found in Chapter 6. Furthermore, Appendix A provides some supplementary theory and Appendix B contains summary tables from the simulation study.

Chapter 2

Survival Analysis

In this chapter we introduce concepts around survival analysis used in this thesis. In sections 2.1-2.5 we go through the basics of survival analysis. In Section 2.6 two nonparametric estimators are defined and in Section 2.7 parametric survival analysis is introduced. The Cox proportional hazards model is described in Section 2.8, as a useful model to take covariates into account. The Royston-Parmar models are presented in Section 2.9 and in Section 2.10 we briefly present how one can connect counting processes and survival analysis. Finally in Section 2.11 some other useful methods, not used in the computational part of this thesis, are mentioned.

2.1 Introduction

In survival analysis we are interested in the time between two events. This can for instance be the time between diagnosis of a disease and disease progression or death from said disease. This time is referred to as the survival time or the time-to-event. A quantity we are interested in is the probability of the event happening, as described by the survival function. Furthermore, we are interested in the rate at which the event takes place, described by the hazard function.

The time-to-event is often not observed. In medical trials for instance, there is often a set follow-up time after which we do not observe the subjects any more. These observations are called censored, although they do still provide information about survival.

2.2 Censoring

The most common form of censoring considered in survival analysis is right censoring. This is when we know that the time-to-event is greater than some time quantity. There is also left censoring which is when we know that the time-to-event is less than some time quantity. This is not as common since we often know a point in time when the event has not happened. More natural is interval censoring. This is when we know that the time-to-event is greater than some time quantity and less than some other time quantity. For instance, if our event of interest is that some disease has progressed, then we can only know this once it is tested for and then we

only know that it has happened at or before the time of testing and after the time of a previous time of testing. In this thesis we will mainly be considering right censoring.

Often when we consider censoring we must think of what constitutes censoring. It might be the case that a death occurred by something else than the disease or illness the study is concerned with investigating. For instance a subject in a medical trial might die in a car accident and should perhaps thus be considered censored. This however is tricky since the car accident might have come from dizziness caused by the disease or some drug which is part of the clinical trial, and one might thus argue that this is not a censored observation. A common solution to this problem, which is not considered in this thesis, is to look at relative survival. See the book by Aalen, Borgan & Gjessing (2008) or the article by Andersson, Dickman, Eloranta & Lambert (2011) for an introduction to and modeling of relative survival.

If the time-to-event and the censoring time are not independent we have informative censoring. For most statistical procedures involving survival data we assume noninformative censoring. In this thesis we will be concerned with data that is mostly complete and our interest lies in examining fixed follow-up times, which is a form of non-informative censoring. Thus we will not worry about problems of informative censoring for the purposes of this thesis.

2.3 Data Setup

We will now take a closer look at the mathematical foundation of survival analysis, starting with the data setup.

We denote the time-to-event by T and the censoring time by C, where both T and C are continuous, non-negative random variables. We observe the outcome variable $Y = \min(T, C)$ together with a censoring indicator δ, where

$$\delta = \begin{cases} 1, & \text{if } T < C, \\ 0, & \text{else.} \end{cases}$$

Thus our data setup is one column for Y, one for δ and also optional additional columns with explanatory variables $\boldsymbol{X}$.

2.4 Important Functions in Survival Analysis

In survival analysis there are a number of functions that are of particular importance. These are the survival function, $S(t)$, the hazard function $\lambda(t)$, and the cumulative hazard function, $\Lambda(t)$. Each of these three functions, together with the cumulative distribution function (CDF), can be written in terms of all the other. The following definitions can be found in Therneau & Grambsch (2000).

The survival function of a random time to event, $T \geq 0$, is defined as

$$S(t) = \Pr(T > t) = 1 - F(t),$$

where $F(t)$ is the CDF of T. It is the probability of a subject surviving at least until time t. The hazard function is defined as

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)},$$

where $f(t)$ is the probability density function (PDF) of T. It is the conditional density of T given that $T > t$. It is also known as the hazard rate or the mortality rate (Nelson, 2005). It can be interpreted as the rate at which the event occurs at time t. The cumulative hazard function is

$$\begin{aligned}\Lambda(t) &= \int_0^t \lambda(t)dt \\ &= -\int_0^t \frac{S'(t)}{S(t)}dt \\ &= -\log S(t).\end{aligned}$$

Of these three functions, the hazard function is the function that is most accessible for interpretation in terms of the underlying data generating process.

To understand this we think of a discrete time process so that we can think in terms of probabilities and not densities. Say that we measure time in years. Then the hazard function evaluated at 75 years would give us the probability of an individual, alive at 75 years, dying before they turn 76 years old. For instance, the hazard of an exponential distribution is constant in time. Thus, the underlying process is time homogeneous and the rate at which events take place is independent of the current survival time. On the other hand, for a Gompertz distribution the hazard is monotonically increasing in time, which would correspond to an individual experiencing greater risk of an event the longer that individual has been alive. The Gompertz distribution, is often used in modeling the survival of the general population for this reason, especially since it's hazard function can have a shape that is rather flat for the early and mid years while then quite sharply increasing for older individuals. One could also use the Gompertz-Makeham distribution. This distribution differs from the Gompertz in that the hazard has a constant added to it (Melnikov & Romaniuk, 2006).

2.5 Likelihood Function for Censored Data

In this section we will derive the likelihood function for censored data (Aalen, Borgan & Gjessing, 2008). We assume that the censoring is non-informative. For subject i, who dies at time t_i and is not censored, i.e. $\delta_i = 0$, the likelihood contribution is

$$L(t_i, \delta_i) = f(t_i) = \lambda(t_i)S(t_i).$$

If the subject is instead censored at time t_i, thus $\delta_i = 1$, then the likelihood contribution is

$$L(t_i, \delta_i) = S(t_i).$$

Putting these two contributions together we have that the likelihood contribution for subject i is

$$\begin{aligned} L(t_i, \delta_i) &= f(t_i)^{1-\delta_i} \cdot S(t_i)^{\delta_i} \\ &= (\lambda(t_i)S(t_i))^{1-\delta_i} \cdot S(t_i)^{\delta_i} \\ &= \lambda(t_i)^{1-\delta_i} S(t_i). \end{aligned}$$

Thus, in a study of N independent subjects, the likelihood is

$$L(t_i, \delta_i) = \prod_{i=1}^{N} \lambda(t_i)^{1-\delta_i} S(t_i),$$

and the log-likelihood is

$$\begin{aligned} l(t_i, \delta_i) &= \sum_{i=1}^{N} (1-\delta_i) \log \lambda(t_i) + \log S(t_i) \\ &= \sum_{i=1}^{N} (1-\delta_i) \log \frac{f(t_i)}{S(t_i)} + \log S(t_i) \\ &= \sum_{i=1}^{N} (1-\delta_i) \log f(t_i) + \delta_i \log S(t_i). \end{aligned}$$

2.6 Nonparametric Estimators

In this section we introduce two nonparametric estimators, the Kaplan-Meier estimator of the survival function and the Nelson-Aalen estimator of the cumulative hazard function. These are appropriate when we want to visualize the data or we do not want to impose some assumed parametric model for the data.

2.6.1 Kaplan-Meier Estimator of the Survival Function

To estimate the survival function non-parametrically we can use the product limit method of Kaplan & Meier (1958), often called the Kaplan-Meier estimator. It is defined as

$$\hat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right) = \prod_{t_i \le t} \frac{n_i - d_i}{n_i}$$

where t_i is the time from the start at index i, d_i is the number of deaths between t_{i-1} and t_i and n_i is the number of subjects at risk at t_{i-1}. Also, note that $n_i = n_{i-1} - d_{i-1} - c_{i-1}$, where c_{i-1} are the number of subjects censored between t_{i-2} and t_{i-1}.

To estimate the variance of the Kaplan-Meier survival curve estimator we can use Greenwood's

formula of Greenwood (1926) which states that

$$\widehat{\text{Var}}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}. \tag{2.1}$$

Using the Kaplan-Meier estimator we can also find an estimator for the cumulative hazard function by the method of Peterson (1977) through

$$\hat{\Lambda}(t) = -\log \hat{S}(t). \tag{2.2}$$

2.6.2 Nelson-Aalen Estimator of the Cumulative Hazard Function

The Nelson-Aalen estimator, derived by Nelson (1972) and Aalen (1978), estimates the cumulative hazard function. It is defined as

$$\hat{\Lambda}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i}.$$

In Section 2.10 we outline how this estimator is derived using counting processes. The Nelson-Aalen estimator can be related to the Kaplan-Meier estimator through

$$\begin{aligned} -\log \hat{S}_{\text{KM}}(t) &= -\sum_{t_i \leq t} \log\left(1 - \frac{d_i}{n_i}\right) \\ &\approx \sum_{t_i \leq t} \frac{d_i}{n_i} \\ &= \hat{\Lambda}_{\text{NA}}(t). \end{aligned}$$

An estimator of the variance of this estimator is given in Hosmer & Lemeshow (1999) as

$$\widehat{\text{Var}}(\hat{\Lambda}(t)) = \sum_{t_i \leq t} \frac{d_i}{n_i^2}.$$

In the same way as equation (2.2), Breslow (1972) suggested that the survival function can be estimated by

$$\hat{S}(t) = \exp(-\hat{\Lambda}(t)),$$

where $\hat{\Lambda}(t)$ is the Nelson-Aalen estimator of the cumulative hazard function.

2.7 Parametric Survival Analysis

There are a number of different parametric distributions that can be used to model survival data. In this section we describe some of them. The advantage of using parametric distributions, instead of the mentioned nonparametric methods, is that we generally get smooth estimates of the hazard and survival functions, as well as the ability to extrapolate beyond the range of the

data (Royston & Lambert, 2011). In Section 2.12 we discuss methods that help us in selecting models.

2.7.1 Distributions

In Table 2.1 some of the different distributions considered in this thesis are shown. It is assumed that the reader is somewhat familiar with these distributions. For a short description of each, see Appendix A.1. Beyond the distributions shown in the table, we also describe the generalized gamma and the generalized F distributions for the sake of completeness. These are umbrella distributions that contain most of the distributions considered in this thesis as special cases.

Table 2.1: Parameters, probability density function, survival function, hazard function and mean of the distributions considered.

Distribution	Parameters	$f(t)$	$S(t)$	$\lambda(t)$	$\mathrm{E}[T]$
Exponential	$\beta > 0$	$\beta e^{-\beta t}$	$e^{-\beta t}$	β	$\frac{1}{\beta}$
Weibull	$a > 0, b > 0$	$\frac{a}{b}\left(\frac{t}{b}\right)^{a-1} e^{-\left(\frac{t}{b}\right)^{a}}$	$e^{-\left(\frac{t}{b}\right)^{a}}$	$\frac{a}{b}\left(\frac{t}{b}\right)^{a-1}$	$b\Gamma\left(1+\frac{1}{a}\right)$
Log-normal	$\mu, \sigma > 0$	$\frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\log t-\mu)^2}{2\sigma^2}}$	$\Phi\left(-\frac{\log t-\mu}{\sigma}\right)$	$\frac{\frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\log t-\mu)^2}{2\sigma^2}}}{\Phi\left(-\frac{\log t-\mu}{\sigma}\right)}$	$e^{\mu+\frac{1}{2}\sigma^2}$
Fréchet	$a > 0, b > 0$	$\frac{a}{b}\left(\frac{t}{b}\right)^{-(1+a)} e^{-\left(\frac{t}{b}\right)^{-a}}$	$1 - e^{-\left(\frac{t}{b}\right)^{-a}}$	$\frac{a\left(\frac{t}{b}\right)^{-(1+a)}}{b\left(e^{\left(\frac{t}{b}\right)^{-a}}-1\right)}$	$b\Gamma\left(1-\frac{1}{a}\right), \quad a > 1$
Gompertz	$a > 0, b > 0$	$ae^{\frac{t}{b}} e^{ab\left(1-e^{\frac{t}{b}}\right)}$	$e^{ab\left(1-e^{\frac{t}{b}}\right)}$	$ae^{\frac{t}{b}}$	$be^{ab}\int_{ab}^{\infty} \frac{e^{-t}}{t} dt$
Log-logistic	$a > 0, b > 0$	$\frac{\frac{a}{b}\left(\frac{t}{b}\right)^{a-1}}{\left(1+\left(\frac{t}{b}\right)^{a}\right)^2}$	$\frac{1}{1+\left(\frac{t}{b}\right)^{a}}$	$\frac{\frac{a}{b}\left(\frac{t}{b}\right)^{a-1}}{1+\left(\frac{t}{b}\right)^{a}}$	$\frac{b\frac{\pi}{a}}{\sin\left(\frac{\pi}{a}\right)}$, if $a > 1$
Gamma	$a > 0, s > 0$	$\frac{x^{a-1}e^{-\frac{x}{s}}}{s^a\Gamma(a)}$	$1 - \frac{\gamma\left(a,\frac{t}{s}\right)}{\Gamma(a)}$	$\frac{t^{a-1}e^{\frac{t}{s}}}{s^a\left(\Gamma(a)-\gamma\left(a,\frac{t}{s}\right)\right)}$	as

2.7.2 Examples of Hazard Functions

In Figure 2.1 we show some example hazard functions which show the possible shapes, using the standard parametric distributions described in subsection 2.7.1. The parameter values are not of importance since we just want to demonstrate possible shapes. The parameter values are therefore not printed. Note that each distribution only provides a small number, i.e. at most, three different principal shapes.

2.7.3 Truncated Distributions

One problem with the standard parametric distributions is that they have support on the whole positive real line. Thus unrealistically large survival times will have a non-zero probability of occurring. A simple solution to this is to truncate our distribution to only have support on realistic values.

Let T be a random time-to-event with PDF $f(t)$ and CDF $F(t)$. The random time-to-event we are interested in is then $T|T < \tau$, for some given constant τ. This random variable has the

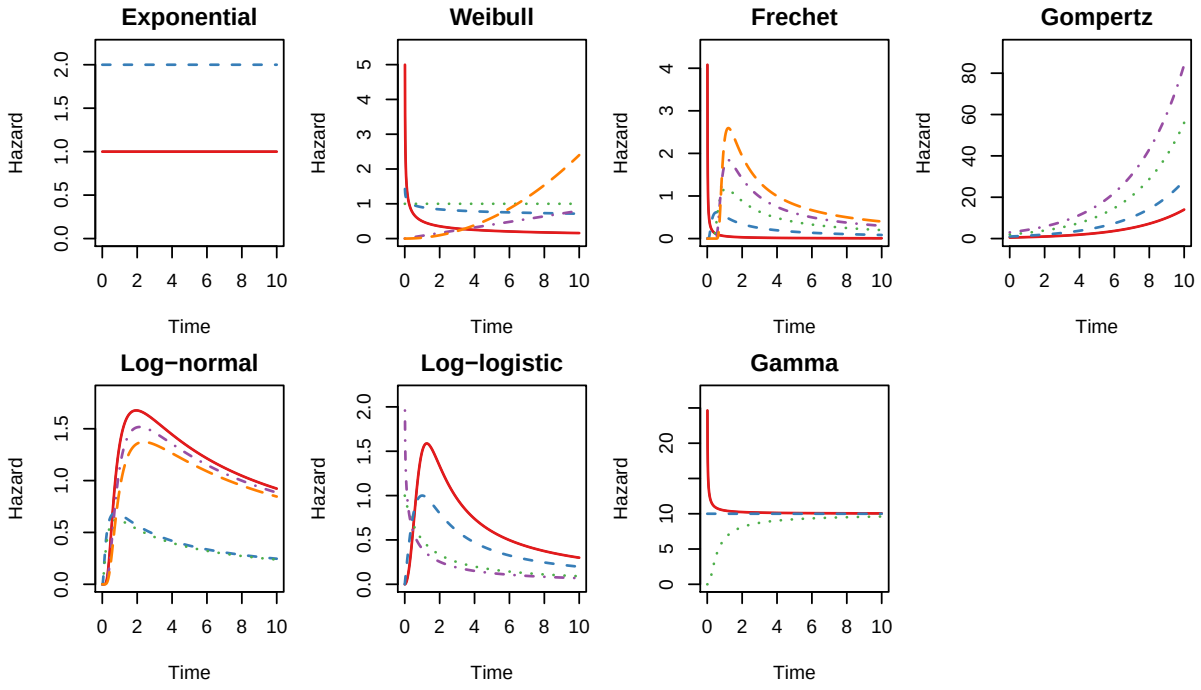

Figure 2.1: Examples of hazard functions for the different parametric distributions described in subsection 2.7.1. The parameter values are not of importance since we just want to demonstrate possible shapes. The parameter values are therefore not printed.

PDF

$$f(t|T < \tau) = \frac{f(t)I(t \leq \tau)}{F(\tau)},$$

where $I(t \leq \tau)$ is the indicator function which is 1 if $t \leq \tau$. The CDF of this distribution is

$$\begin{aligned} F(t|T < \tau) &= \frac{\int_0^t f(s)I(s \leq \tau)ds}{F(\tau)} \\ &= \left(\frac{F(t)}{F(\tau)}\right)^{I(t \leq \tau)}. \end{aligned}$$

This can be solved analytically but for our needs it will be sufficient to calculate this in `R` using the standard functions for the distributions.

To simulate observations from a truncated distribution we can use inverse transform sampling, assuming calculability of the quantile function for the distribution. We sample uniform random numbers between 0 and $F(\tau)$ and insert into the quantile function of the non-truncated distribution to get a sample from the truncated distribution.

If the quantile function is not available analytically and is slow to compute numerically, we can sample from the non-truncated distribution, provided it can be sampled from. Then we throw away values above τ to get a sample from the truncated distribution. This should be quite fast and lead to few rejections since the total probability mass above τ should be quite small.

In Figure 2.2 the hazard function of a truncated log-normal distribution with parameters $\mu = 0$, $\sigma = 1$ and $\tau = 12$ can be seen together with the non-truncated distribution's hazard with

the same parameters. This seems promising since we get the initial peak which would be biologically plausible for many diseases and then the tail has increasing hazard reflecting ageing of subjects.

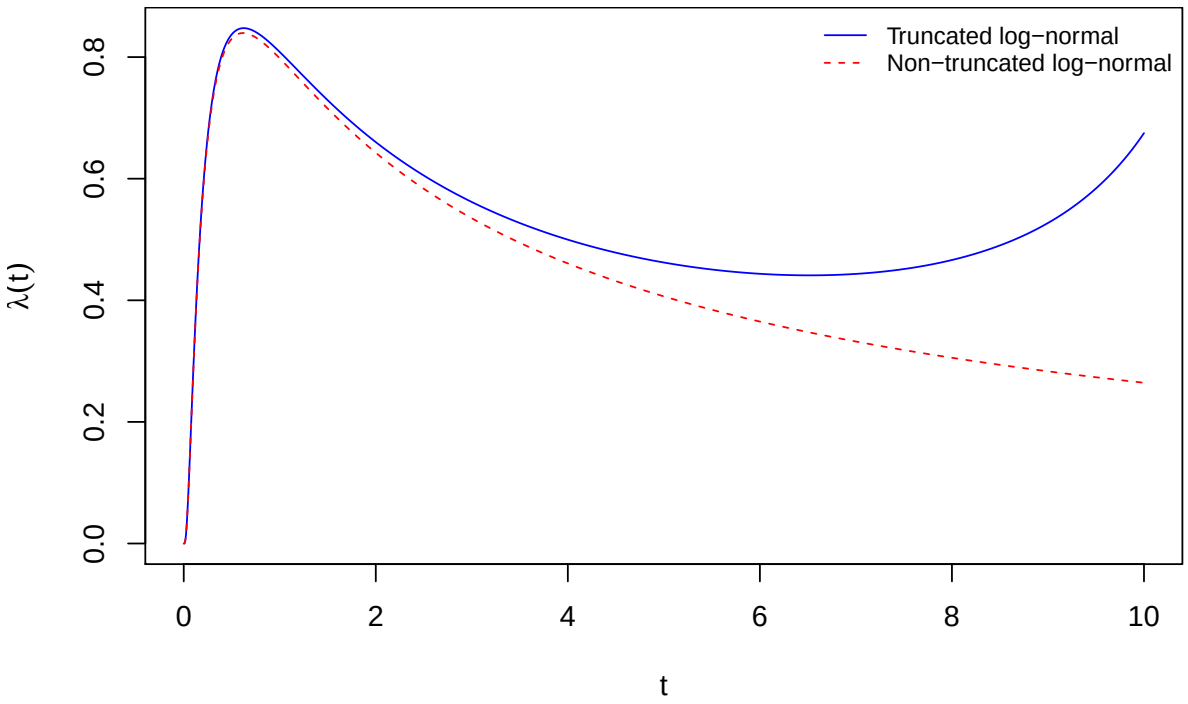

Figure 2.2: Hazard functions of a truncated log-normal distribution with parameters $\mu = 0$, $\sigma = 1$ and $\tau = 12$ and a non-truncated distribution's hazard with the same parameters.

2.8 Cox Proportional Hazards Model

This section follows Hosmer & Lemeshow (2008) and Kalbfleisch & Prentice (2011). As with any data, subjects are not necessarily all the same and we want to introduce explanatory variables which might affect the outcome. The most common way to introduce explanatory variables is to use the Cox proportional hazards model (Cox, 1972) which is a semi-parametric regression model. Semi-parametric here meaning that we have a parametric and a non-parametric part. The parametric part is the regression coefficients and the non-parametric part is the baseline hazard, which we do not estimate.

The proportional hazards model is defined through the hazard function in the following way

$$\lambda(t; \boldsymbol{X}) = \lambda_0(t) r(t, \boldsymbol{X}), \quad t > 0. \tag{2.3}$$

That is, the hazard function for a subject with explanatory variables $\boldsymbol{X}$, is proportional to some baseline hazard function $\lambda_0(\cdot)$ through the relative risk function $r(t, \boldsymbol{X})$. The most common setup for the relative risk function is

$$r(t, \boldsymbol{X}) = \exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right).$$

With this relative risk function the model is usually called the Cox proportional hazards model. One of the neat things about the model in equation (2.3) is that the hazard ratio is

$$\mathrm{HR}(t, \boldsymbol{X}_1, \boldsymbol{X}_0) = \frac{r(t, \boldsymbol{X}_1)}{r(t, \boldsymbol{X}_0)},$$

and thus does not depend on the baseline hazard function. For the Cox proportional hazards model this means that

$$\mathrm{HR}(t, \boldsymbol{X}_1, \boldsymbol{X}_0) = e^{(\boldsymbol{X}_1 - \boldsymbol{X}_0)'\boldsymbol{\beta}}.$$

This yields the interpretation that a one unit increase in explanatory variable x_i, holding all others fixed, corresponds to subjects having the event of interest at $\exp(\beta_i)$ times the rate.

When the relative risk function is independent of time, i.e. $r(t, \boldsymbol{X}) = r(\boldsymbol{X})$, we can get the survival function by first calculating the cumulative hazard function

$$\begin{aligned}
\Lambda(t; \boldsymbol{X}) &= \int_0^t \lambda(s) ds \\
&= r(\boldsymbol{X}) \int_0^t \lambda_0(s) ds \\
&= r(\boldsymbol{X}) \Lambda_0(t).
\end{aligned}$$

Then we use the relationship between this and the survival function to get

$$\begin{aligned}
S(t; \boldsymbol{X}) &= e^{-\Lambda(t; \boldsymbol{X})} \\
&= e^{-r(\boldsymbol{X})\Lambda_0(t)} \\
&= \left(e^{-\Lambda_0(t)}\right)^{r(\boldsymbol{X})} \\
&= S_0(t)^{r(\boldsymbol{X})}.
\end{aligned}$$

Although one of the nice things about the Cox proportional hazards model is that we do not need to make a distributional assumption for the survival curve, this will not apply in this thesis since we are interested in the extrapolation of survival curves and thus need more than just hazard ratios. Of course we can always specify a parametric distribution and still use the Cox proportional hazards model setup and find the parameter values through regular maximum likelihood estimation. Some models inherently have proportional hazards, such as the exponential distribution, where each covariate setup gets their own parameter value.

2.8.1 Maximum Likelihood Estimation

To estimate the parameters $\boldsymbol{\beta}$ we use the partial likelihood function as introduced by Cox (1972). We will not go through its derivation, see Kalbfleisch & Prentice (2011) or Hosmer & Lemeshow

(2008) for a more detailed approach. The partial likelihood function for $\boldsymbol{\beta}$ is

$$L_p(\boldsymbol{\beta}; \boldsymbol{X}) = \prod_{i:\delta_i=0} \left(\frac{r(\boldsymbol{\beta}, \boldsymbol{X}_i)}{\sum_{j:Y_j \geq Y_i} r(\boldsymbol{\beta}, \boldsymbol{X}_i)} \right)$$
$$= \prod_{i:\delta_i=0} \left(\frac{e^{\boldsymbol{X}_i'\boldsymbol{\beta}}}{\sum_{j:Y_j \geq Y_i} e^{\boldsymbol{X}_j'\boldsymbol{\beta}}} \right).$$

Thus, the log partial likelihood function is

$$l_p(\boldsymbol{\beta}; \boldsymbol{X}) = \sum_{i:\delta_i=0} \left[\boldsymbol{X}_i'\boldsymbol{\beta} - \log \left(\sum_{j:Y_j \geq Y_i} e^{\boldsymbol{X}_j'\boldsymbol{\beta}} \right) \right]. \tag{2.4}$$

This equation can be solved by using, for instance, the Newton-Raphson algorithm (Kalbfleisch & Prentice, 2011).

2.8.2 Schoenfeld Residuals

Here we follow along Hosmer & Lemeshow (1999). Schoenfeld residuals, proposed by Schoenfeld (1982), are residuals for Cox proportional hazards models. They are derived from the derivative of the log partial likelihood, see Equation (2.4). That is, we have

$$\frac{\partial l_p(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{i:\delta_i=0} \left[x_{ik} - \frac{\sum_{j:Y_j \geq Y_i} x_{jk} e^{\boldsymbol{X}_j'\boldsymbol{\beta}}}{\sum_{j:Y_j \geq Y_i} e^{\boldsymbol{X}_j'\boldsymbol{\beta}}} \right].$$

The individual Schoenfeld residual for the ith subject for the kth covariate is then estimated by

$$\hat{r}_{ik} = \delta_i \left(x_{ik} - \frac{\sum_{j:Y_j \geq Y_i} x_{jk} e^{\boldsymbol{X}_j'\hat{\boldsymbol{\beta}}}}{\sum_{j:Y_j \geq Y_i} e^{\boldsymbol{X}_j'\hat{\boldsymbol{\beta}}}} \right),$$

where $\hat{\boldsymbol{\beta}}$ is the partial likelihood estimator of $\boldsymbol{\beta}$. Since these residuals are based on the score function, they should be distributed around zero. For greater diagnostic power Grambsch & Therneau (1994) scaled the Schoenfeld residuals by an estimator of the variance of the residuals. That is, the scaled Schoenfeld residuals are

$$\hat{\boldsymbol{r}}_i^* = \left[\widehat{\mathrm{Var}}(\hat{\boldsymbol{r}}_i) \right]^{-1} \hat{\boldsymbol{r}}_i.$$

We leave out the computation of the variance and instead show an approximation for the variance that is used in the function `cox.zph` in the package `survival`. It was suggested by Grambsch & Therneau (1994) based on their experience that the variance matrix often is close to constant over time. If we assume that the distribution of a covariate is the same for different risk sets, then an approximation of the variance is

$$\left[\widehat{\mathrm{Var}}(\hat{\boldsymbol{r}}_i) \right]^{-1} = m \widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}),$$

where m is the number of observed survival times. The variance of $\hat{\boldsymbol{\beta}}$ is straight forward to obtain through the Fisher information matrix of the partial log likelihood.

2.8.3 Testing the Proportional Hazards Assumption

Continuing on the previous subsection we can construct a test of the proportional hazards assumption. We follow Klein & Kleinbaum (2005) who propose that one ranks time and test whether the ranks and the scaled Schoenfeld residuals are correlated. The thought process being that if the proportional hazards assumption holds, then there is independence between time, or the ranks of time, and the (scaled) Schoenfeld residuals. Independence implies that they are uncorrelated. Although the converse does not hold, this will be close enough for our purposes. The test of the correlation coefficient being 0, that is $H_0 : \rho = 0$, is done via a chi-square test.

We can also plot the scaled Schoenfeld residuals against time to visually asses the proportional hazards assumption. A random pattern independent of time would support the assumption.

2.9 Royston-Parmar Models

Having introduced the standard parametric models and some non-parametric models, we now go on to the flexible parametric survival models, also called the Royston-Parmar models.

This section follows Royston & Lambert (2011). Royston-Parmar (RP) models are generalizations of standard parametric survival models. They are based on the use of a restricted cubic spline function to allow for greater flexibility in modeling. The basic concept is to linearize the survival function in log time and then adding basis functions of restricted cubic splines. This also yields a natural way of adding explanatory variables to the models.

We start the section by going through what restricted cubic splines are, then we move on to show how we go about generalizing the standard parametric survival models. After that we look at the likelihood and parameter estimation for the RP models and finally we have a short discussion about the RP models sensitivity to the number and location of the knots of the splines.

2.9.1 Restricted Cubic Splines

A cubic spline with K knots can be written as

$$s(x;\boldsymbol{\gamma}) = \sum_{j=0}^{3} \gamma_{0j} x^j + \sum_{i=1}^{K} \gamma_{i3} (x - k_i)_+^3,$$

where

$$x_+ = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases}$$

(Royston & Lambert, 2011). Restricted cubic splines are cubic splines that are forced to be linear before the first knot and after the last knot. In our modeling we will let the first knot be the minimum of the log uncensored survival times and the last knot be the maximum of the log uncensored survival times. With this boundary condition we can write a restricted cubic spline with m interior knots $k_2, ..., k_{m+1}$ and two boundary knots $k_{\min}$ and $k_{\max}$ as

$$s(x; \boldsymbol{\gamma}) = \gamma_0 + \sum_{i=1}^{m+1} \gamma_i z_i(x),$$

where

$$z_1(x) = x,$$

and

$$z_j(x) = (x - k_j)_+^3 - \lambda_j (x - k_{\min})_+^3 - (1 - \lambda_j)(x - k_{\max})_+^3,$$

for $j = 2, ..., m + 1$. Here we have that

$$\lambda_j = \frac{k_{\max} - k_j}{k_{\max} - k_{\min}}.$$

In the following subsections we will use restricted cubic splines to generalize some survival functions. These models are the Royston-Parmar models.

2.9.2 Generalizing the Weibull Distribution (Proportional Hazards Model)

In this and the following two subsection we generalize some survival functions by using restricted cubic splines. We also look at how we can add explanatory variables in modeling the survival function.

The Weibull distribution has the survival function

$$S(t) = e^{-\left(\frac{t}{b}\right)^a}.$$

By taking the logarithm of the cumulative hazard function, $\Lambda(t) = -\log S(t)$, we get

$$\log \Lambda(t) = \log(-\log(S(t))) = a \log t - a \log b = \gamma_0 + \gamma_1 \log t.$$

This sum of two components can be generalized by the spline function $s(\log t; \boldsymbol{\gamma})$ by writing

$$\log \Lambda(t) = s(\log t; \boldsymbol{\gamma}) = \gamma_0 + \gamma_1 \log t + \sum_{i=1}^{k} \gamma_{i+1} z_i(\log t),$$

where z_i, $i = 1, ..., k$, are basis functions of the restricted cubic splines. To model with explanatory variables we can extend the equation to be

$$\log \Lambda(t; \boldsymbol{X}) = s(\log t; \boldsymbol{\gamma}) + \boldsymbol{X}'\boldsymbol{\beta}.$$

By letting $\log \Lambda_0(t) = s(\log t; \boldsymbol{\gamma})$ we see that

$$\Lambda(t; \boldsymbol{X}) = \Lambda_0(t) e^{\boldsymbol{X}'\boldsymbol{\beta}},$$

which means that this is a proportional hazards model.

2.9.3 Generalizing the Log-Logistic Model (Proportional Odds Model)

For the log-logistic model we have

$$S(t) = \frac{1}{1 + \left(\frac{t}{b}\right)^a}.$$

Thus we have

$$\text{logit}\,(1 - S(t)) = a \log t - a \log b = \gamma_0 + \gamma_1 \log t.$$

By adding the spline function $s(\log t; \boldsymbol{\gamma})$ and potential explanatory variables $\boldsymbol{X}$ we can write the generalization as

$$\text{logit}\,(1 - S(t; \boldsymbol{X})) = s(\log t; \boldsymbol{\gamma}) + \boldsymbol{X}'\boldsymbol{\beta}.$$

Following the same thought as for the Weibull generalization we let $\text{logit}\,(1 - S_0(t)) = s(\log t; \boldsymbol{\gamma})$ and exponentiate to get

$$\frac{1 - S(t; \boldsymbol{X})}{S(t; \boldsymbol{X})} = \frac{1 - S_0(t)}{S_0(t)} e^{\boldsymbol{X}'\boldsymbol{\beta}}.$$

Thus we see that this is a proportional odds model.

2.9.4 Generalizing the Log-Normal Model (Probit Model)

For the log-normal model we have

$$S(t) = \Phi\left(-\frac{\log t - \mu}{\sigma}\right).$$

Thus we have

$$-\Phi^{-1}(S(t)) = \frac{1}{\sigma}(\log t - \mu) = \gamma_0 + \gamma_1 \log t.$$

In the same way as before we can thus generalize this by writing

$$-\Phi^{-1}(S(t; \boldsymbol{X})) = s(\log t; \boldsymbol{\gamma}) + \boldsymbol{X}'\boldsymbol{\beta}.$$

If t was fixed this would be a regular binary probit model.

2.9.5 Likelihood Function and Parameter Estimation

In this subsection we look at the likelihood contributions for the three different RP models introduced in this section. We derive explicitly the one for the Weibull RP model.

Let $\eta_i(t_i) = s(\log t_i; \boldsymbol{\gamma}) + \boldsymbol{X}_i\boldsymbol{\beta}$. Now, from Section 2.5 we know that the log-likelihood contribution

for subject i is

$$l_i(t_i, \delta_i) = (1 - \delta_i) \log \lambda(t_i) + \log S(t_i).$$

For the Weibull RP model we have

$$\log \Lambda(t_i) = \eta_i(t_i).$$

Thus the likelihood contribution for a censored subject is

$$\begin{aligned} l_i(t_i, \delta_i = 1) &= \log S(t_i) \\ &= \log e^{-\Lambda(t_i)} \\ &= -\Lambda(t_i) \\ &= -e^{\eta_i(t_i)}, \end{aligned}$$

and for an uncensored subject it is

$$\begin{aligned} l_i(t_i, \delta_i = 0) &= \log \lambda(t_i) + \log S(t_i) \\ &= \log \left(\frac{d}{dt} \Lambda(t) \mid_{t=t_i} \right) - e^{\eta_i(t_i)} \\ &= \log \left(\frac{d}{dt} e^{\eta_i(t)} \mid_{t=t_i} \right) - e^{\eta_i(t_i)} \\ &= \log \left(\eta_i'(t_i) e^{\eta_i(t_i)} \right) - e^{\eta_i(t_i)} \\ &= \log \eta_i'(t_i) + \eta_i(t_i) - e^{\eta_i(t_i)}. \end{aligned}$$

Here we have that

$$\begin{aligned} \eta_i'(t_i) &= \frac{d\eta_i(t_i)}{dt_i} \\ &= \frac{ds(\log t; \boldsymbol{\gamma})}{dt_i} \\ &= t_i^{-1} \frac{ds(\log t; \boldsymbol{\gamma})}{d \log t_i}, \end{aligned}$$

where

$$\begin{aligned} \frac{ds(\log t; \boldsymbol{\gamma})}{d \log t_i} &= \gamma_1 + \sum_{i=2}^{m+1} \gamma_i \frac{dz_i(\log t_i)}{d \log t_i} \\ &= \gamma_1 + \sum_{j=2}^{m+1} \gamma_i \left(3(\log t_i - k_j)_+^2 - 3\lambda_j (\log t_i - k_{\min})_+^2 - 3(1 - \lambda_j)(\log t_i - k_{\max})_+^2 \right). \end{aligned}$$

The above expression can now be optimized using standard software routines. We leave out the computations for the log-logistic and log-normal models, but for completeness the log-likelihood contributions can be seen below. For the log-logistic RP model we have

$$l_i(t_i, \delta_i = 1) = -\log \left(1 + e^{\eta_i(t_i)} \right),$$

and

$$l_i(t_i, \delta_i = 0) = \eta_i(t_i) + \log \eta_i'(t_i) - 2 \log \left(1 + e^{\eta_i(t_i)} \right).$$

For the log-normal RP model we have

$$l_i(t_i, \delta_i = 1) = \log\left(\Phi\left(-\eta_i(t_i)\right)\right),$$

and

$$l_i(t_i, \delta_i = 0) = \log \eta_i'(t_i) + \log \phi\left(\eta_i(t_i)\right).$$

2.9.6 Sensitivity to Number of Knots and Location of Knots

Royston & Lambert (2011) perform an investigation of the sensitivity of the number of knots and the location of the knots on the survival function. They do this by fitting log-logistic RP models to a breast cancer data set and then compare the survival functions visually. They first fit four models with 0, 1, 2 and 3 knots, respectively. The model with 0 knots is quite different from the other three, but the three models with knots are difficult to separate, indicating, at least for this data set, that one knot gives the needed extra flexibility. After that they fitted 10 models with 2 interior knots, whose positions were selected randomly. This again yielded curves that were difficult to distinguish from one another.

The above example does however not prove generally that the number of knots and knot positions do not influence, to a great extent, the survival function. In this thesis we will choose the number of knots using AIC and BIC, see Section 2.12.2, unless there is some reason to not choose the model with the lowest AIC or BIC, such as a biologically implausible hazard. The knot location will be chosen as equally-spaced percentiles of the empirical log uncensored survival times, as suggested by Royston & Parmar (2002).

2.10 Counting Process Formulation of Survival Analysis

There is a natural connection between counting processes and survival analysis that we will demonstrate in this section. Specifically we will derive the Nelson-Aalen estimator of the cumulative hazard function, which was presented in subsection 2.6.2. In subsection 2.10.1 we describe the Martingale residual. It is nice to see, but for the uninterested reader this section is not necessary for continued understanding of this thesis.

The following part of this section follows Crowder (2001). Let $N(t)$ be a counting process. This means that:

- $N(t) \geq 0$ takes integer values.
- $N(s) \leq N(t)$ for $s \leq t$.
- $dN(t) = N(t) - N(t-)$ is equal to either 0 or 1.
- $\mathrm{E}[N(t)] < \infty$.

Above we use the notation that $N(t-) = \lim_{\epsilon \downarrow 0} N(t - \epsilon)$. A filtration $\mathcal{F}_t$ of a counting process can be said to be all the information we know about the counting process up to and including

time t. We will use the notation $\mathcal{F}_{t-}$ to be the information up to but not including time t.

We can write the expectation or probability

$$\mathrm{E}[dN(t)|\mathcal{F}_{t-}] = P\left(dN(t) = 1 \mid \mathcal{F}_{t-}\right) = d\Lambda(t),$$

where

$$\Lambda(t) = \int_0^t \lambda(s)ds,$$

and $\lambda(t)$ is the intensity of $N(t)$ and $\Lambda(t)$ is the integrated intensity function.

Let $M(t) = N(t) - \Lambda(t)$, we then have

$$\begin{aligned}\mathrm{E}[dM(t)|\mathcal{F}_{t-}] &= \mathrm{E}[dN(t) - d\Lambda(t)|\mathcal{F}_{t-}] \\ &= d\Lambda(t) - d\Lambda(t) \\ &= 0,\end{aligned}$$

which is the Martingale property. This together with

$$\begin{aligned}\mathrm{E}[|M(t)|] &= \mathrm{E}[|N(t) - \Lambda(t)|] \\ &\leq \mathrm{E}[|N(t)|] + \mathrm{E}[|\Lambda(t)|] \\ &< \infty,\end{aligned}$$

shows that $M(t)$ is a Martingale with respect to the filtration $\mathcal{F}_{t-}$. In the above calculations we used the triangle inequality.

In survival analysis the time-to-event is denoted by the random variable T and the censoring indicator is denoted by C. We can connect this to a counting process by

$$N(t) = I(T \leq t, C = 1).$$

The intensity of this counting process is

$$\lambda(t) = Y(t)h(t),$$

where $h(t)$ is the hazard and $Y(t) = I(T \geq t)$ is an indicator for being at risk.

If we have a sample of n individuals with common hazard function then we can construct the summed counting process $N_+(t) = \sum_{i=1}^n N_i(t)$ and its corresponding integrated intensity function $\Lambda_+(t)$, Martingale process $M_+(t)$ and number at risk function $Y_+(t)$. To estimate the cumulative hazard function for this sample we write

$$N_+(t) = \Lambda_+(t) + M_+(t),$$

which yields

$$\begin{aligned}dN_+(t) &= d\Lambda_+(t) + dM_+(t) \\ &= Y_+(t)dH(t) + dM_+(t),\end{aligned}$$

where H is the cumulative hazard function and $N_+(t)$ and $Y_+(t)$ are our observed data. Taking conditional expectation with respect to $\mathcal{F}_{t-}$ yields

$$dN_+(t) = Y_+(t)dH(t),$$

since $dM_+(t)$ is a Martingale with respect to $\mathcal{F}_{t-}$ and thus has zero expectation. Rearranging yields an estimate for the cumulative hazard function

$$d\hat{H}(t) = \frac{dN_+(t)}{Y_+(t)}.$$

Integrating yields

$$\begin{aligned}\hat{H}(t) &= \int_0^t \frac{dN_+(t)}{Y_+(t)} \\ &= \sum_{t_j \leq t} \frac{dN_+(t_j)}{Y_+(t_j)},\end{aligned}$$

where t_j, $j = 1, ..., n$, are the event times, including censored times, of the subjects. The summands are taken to be zero if $Y_+(t_j)$ is zero. This is the Nelson-Aalen estimator of the cumulative hazard described in subsection 2.6.2. One can also derive the Kaplan-Meier estimator using the counting process formulation, but this is not done here.

2.10.1 Martingale Residuals

We can now construct a set of residuals using the counting process formulation of survival analysis. These are based on the Martingale process $M(t)$ described above. An estimator of the Martingale process for the ith individual at the event/censoring time t_i, called the Martingale residual, is (Hosmer & Lemeshow, 1999)

$$\begin{aligned}\hat{M}(t_i) &= \delta_i - \hat{\Lambda}(t_i) \\ &= \delta_i - Y_i(t_i)\hat{H}(t_i) \\ &= \delta_i - \hat{H}(t_i),\end{aligned}$$

where δ_i is 1 if the ith subject had an event at t_i and 0 if the time-to-event was censored at t_i. Since $Y(t) = I(T \geq t)$ and $T = t_i$, in this case we have $Y_i(t_i) = I(t_i \geq t_i) = 1$, and thus the above equation follows. We can use this estimator for the Martingale process as a residual for subject i, noting that since $M(t)$ is a Martingale we should have $\sum_i \hat{M}_i = 0$ and the residuals should be evenly distributed around 0. We can therefore plot the Martingale residuals for different covariates to see if there is some systematic deviation from this.

2.11 Other Methods

In this section we briefly mention some methods that can be useful for survival data, but which are not used in this thesis.

2.11.1 Restricted Mean Survival Time and Median Survival Time

Two useful methods for computing comparative effectiveness is to use restricted mean survival time or the median survival time. The restricted mean survival time of a random time-to-event T is $\mathrm{E}[\min(\tau, T)]$ for some constant time τ (Royston & Parmar, 2013). Thus we are not worried about the behavior of the tail. The median survival time is the survival time $t_{0.5}$ such that $S(t_{0.5}) = 0.5$.

The good thing about these are that they are easier to estimate and often do not require extrapolation. In short, they are more robust than estimators of the unrestricted mean time-to-event. The problem with them is that in health economic evaluations we need the unrestricted mean to be able to say something of use about the cost of, for instance, a treatment. Because of this they will not be considered in this thesis.

2.11.2 Mixture Distributions

Another method is to use mixture distributions. These models imply that different subjects belong to different distributions by some latent variable that is constant over time. An example being that some subjects get cured while others do not. This example is conceptually not useful for our application since it would imply that there are individuals that get cured, which is currently not the case for multiple myeloma. However, there might be other latent variables we could catch by using mixture distributions. One could also think of mixture distributions as just another way of gaining flexibility. In this thesis we will be satisfied with the extra flexibility of the RP models, and the investigation of these, and not consider mixture distributions.

2.11.3 Bayesian Model Averaging

We now take a look at Bayesian model averaging. This is a way to incorporate several models instead of having to decide on one specific model. In the context of Bayesian model averaging, the posterior distribution of the mean, μ, given the data $\boldsymbol{T}$ is

$$P(\mu|\boldsymbol{T}) = \sum_{i=1}^{K} P(\mu|M_i, \boldsymbol{T}) P(M_i|\boldsymbol{T}),$$

where M_i, $i = 1, ..., K$ are the models considered (Hoeting, Madigan, Raftery & Volinsky, 1999). The posterior distribution of M_i is

$$P(M_i|\boldsymbol{T}) = \frac{P(\boldsymbol{T}|M_i)P(M_i)}{\sum_{j=1}^{K} P(\boldsymbol{T}|M_j)P(M_j)},$$

where

$$P(\boldsymbol{T}|M_i) = \int P(\boldsymbol{T}|M_i, \boldsymbol{\theta}_i) P(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i.$$

Here $\boldsymbol{\theta}_i$ is the parameter vector of model M_i. If we let

$$P(M_j) = \frac{1}{K}, \quad j = 1, ..., K,$$

we get

$$\begin{aligned} P(M_i|\boldsymbol{T}) &= \frac{P(\boldsymbol{T}|M_i)}{\sum_{j=1}^{K} P(\boldsymbol{T}|M_j)} \\ &\propto P(\boldsymbol{T}|M_i) \\ &= \int P(\boldsymbol{T}|M_i, \boldsymbol{\theta}_i) P(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i \end{aligned}$$

This is just the expectation of the likelihood under the prior distribution of the likelihood parameters under model M_i. In practice the above can thus be found by inserting MCMC samples into the likelihood.

The reason for not incorporating this into the simulation study of Section 4.2 is not because these models are not of interest, but because of the computation time. Simulating enough MCMC samples for convergence takes quite some time, then one has to do that repeatedly a large number of times. Add to that the fact that we want to investigate several different data generating models and follow-up times and the computational time gets out of hand.

2.11.4 Discrete Time Modeling

Since time-to-event data often is measured in some discrete time scale, such as days, it would make sense to model survival times using a discrete distribution or process.

We could think of a simplified example where the survival probability each day is the same over time, which would closely resemble an exponential survival time in the continuous case. Thus we could, for each day, aggregate the data into the number of patients alive, n_t, and the number of deaths, d_t. Here t can, for instance, be the number of days from diagnosis. We could then say that

$$N_t \sim \text{Bin}(n_{t-1}, p),$$

where N_t is a random variable expressing the number of subjects alive at time t, before taking censoring into account. With censoring we would have $n_t = n_{t-1} - d_{t-1} - c_{t-1}$ where c_t are the number of censored patients at day t. If we do not believe in the assumption of a constant survival probability in time we could model the probability, p, in time, with covariates, or perhaps as an autoregressive process.

An equivalent approach in the case of a constant p is to model each patients survival time in days as a geometric random variable, that is, the number of days until the first event.

We could also go down on the individual level and for instance model the event for each day and individual as outcomes of Bernoulli random variables. That is

$$d_{it} \sim \text{Bernoulli}(\pi_{it}),$$

where d_{it} is an indicator for whether individual i had an event on day t. We then let

$$\text{logit}(\pi_{it}) = \boldsymbol{X}_{it}\boldsymbol{\beta} + \sum_{k=0}^{K} \phi_k(t)$$

where ϕ_k, $k = 1, ..., K$ are some basis functions and $\boldsymbol{X}_{it}$ is the covariate pattern for individual i at day t. This is a logistic regression model where we treat each day and patient as being independent. One problem with this is that there is no guarantee that the probability of an event will stay away from 0, or very small values, when t is greater than our largest observed survival time. Thus we might get a survival curve that flattens out and never goes to 0.

2.12 Model Selection

In this section we discuss methods for model inspection and selection. In subsection 2.12.1 we talk about visual inspection and in subsection 2.12.2 AIC and BIC are introduced as a method of relative model comparison. The visual tools are not used in this thesis since they are difficult to implement in a simulation study, despite this they deserve a mention.

2.12.1 Visual Inspection

A simple first step in model inspection for survival models is to visually compare a model to the Kaplan-Meier curve. This is an informal way of quickly assessing if a model gives a reasonable fit to the data at hand. One should be aware that in the presence of heavy censoring or interval censoring, it can be quite difficult to tell if the model fits the curves nicely. Also, in the tail of the Kaplan-Meier curve there are often quite few subjects, giving an unstable behavior. Still it can be very useful as a first step.

Another useful visual inspection is to plot the (log) cumulative hazard derived from the Kaplan-Meier estimator. This gives an indication of whether the hazard is constant, monotonically increasing or decreasing or if it is non-monotonic with respect to time. This can then inform the choice of model. For instance, a constant hazard, that is a linearly increasing cumulative hazard with respect to time, would lend credence to an exponential model. Also, plotting the cumulative hazard stratified on some explanatory variable would show if the proportional hazards assumption, underlying the Cox proportional hazard model, is reasonable.

Linearize for Visualization

A good visual tool for inspecting which distributions might be suitable to model the data, is to try to find a quantity, derivable from the survival function, that is linear in some transformation of time. Below is a list of such quantities.

- Exponential: $\Lambda(t) = \beta t$,
- Weibull: $\log \Lambda(t) = -a \log b + a \log t$,

- Fréchet: $\log\left(-\log\left(1-S(t)\right)\right) = a\log b - a\log t$,
- Log-logistic: $\text{logit}(S(t)) = a\log b - a\log t$,
- Log-normal: $\Phi^{-1}(S(t)) = -\frac{\mu}{\sigma} - \frac{1}{\sigma}\log t$.

2.12.2 AIC & BIC

The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are measures we can use to do relative comparison between models. They are based on the value of the maximized log likelihood for a given model but with an adjustment through a penalization term that is based on the number of parameters in the model. AIC is defined as (Agresti, 2013)

$$\text{AIC} = -2\hat{l} + 2k,$$

and BIC is defined as

$$\text{BIC} = -2\hat{l} + k\log(n),$$

where $\hat{l}$ is the maximized log-likelihood, n is the number of observations and k is the number of parameters in the model.

The model with minimal AIC, among a set of candidate models, is the model that minimizes the expected Kullback-Leibler divergence. The Kullback-Leibler divergence is the information lost when we use a certain model to approximate reality. The model with minimal BIC is the model that maximizes the posterior probability of the model under the prior $1/R$, where R is the number of candidate models. See Burnham (2004) for a detailed discussion and comparison of AIC and BIC.

AIC and BIC can and will be used in this thesis to determine if one or some models provide a much better fit, relatively, compared to other models.

Chapter 3

Data Description

In this chapter we present and describe the Swedish multiple myeloma dataset used in this thesis. We want the reader to note that the dataset is observational registry data and thus there will be a selection bias for different sets of covariates. The dataset is not a randomized clinical trial and it is thus difficult to talk about causality in our analyses.

Before we discuss the Swedish dataset, a Finnish multiple myeloma dataset deserves a mention. Quite some time was spent investigating this dataset, but it turned out to be contaminated by a survivorship bias. The data was most likely collected backwards in time. Data should be collected by looking at all patients diagnosed at a particular date and then go forward in time. In this case it seems as if one date was picked and all patients alive were collected, then they went backwards in time collecting all alive patients at each time point. Thus the further back in time one goes, patients will seem to live longer. In short, a survivorship bias was present. After removing the time period were the survivorship bias was present, not enough follow-up remained to be of use in studying the accuracy of fitting the tail. Thus this dataset was excluded from the data analysis in this thesis.

The Swedish data consists of registry data of 1606 patients with multiple myeloma between January, 2000 and November, 2011. Of these patients, approximately 47% are censored.

Out of all the patients, approximately 32% received high-dose chemotherapy (hdt) followed by a stem cell transplant. These will be excluded from the analyses done in this thesis since they are very different from those that do not receive hdt and would thus need stratification rather than modeling as an explanatory variable in a joint model. Furthermore, the outcome for these patients is much better and the follow-up time we have is not adequate to capture the whole distribution. Thus the data that will be analysed consists of 1098 multiple myeloma patients. Out of these, approximately 36% are censored. 49% of the patients are men. Hereafter we will refer to this subset as our data.

In Figure 3.1 a histogram of the age at diagnosis for the subjects in our dataset can be seen. The mean age at diagnosis is 74 years.

In Figure 3.2 the number of newly diagnosed patients per year can be seen. The increasing

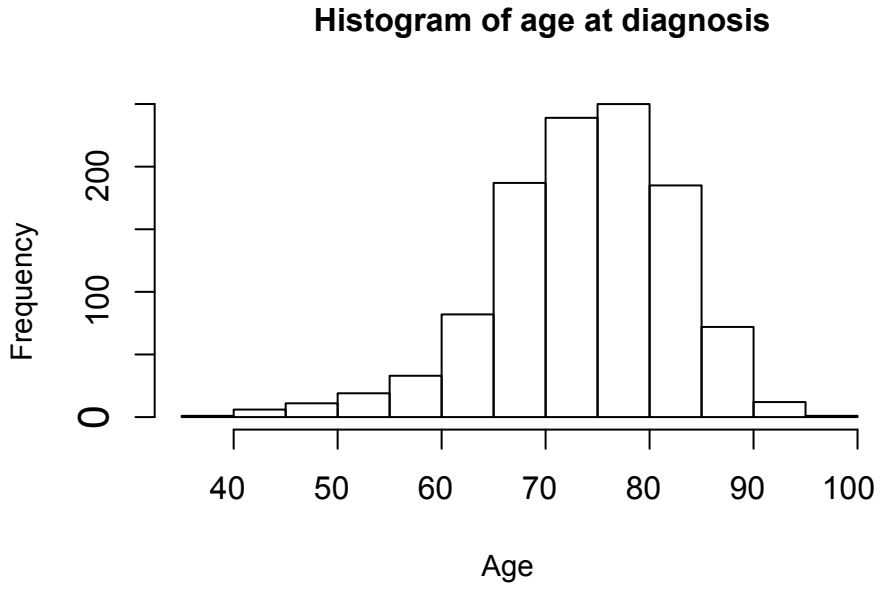

Figure 3.1: Histogram of the age at diagnosis for patients with multiple myeloma that do not receive hdt in their first line of treatment.

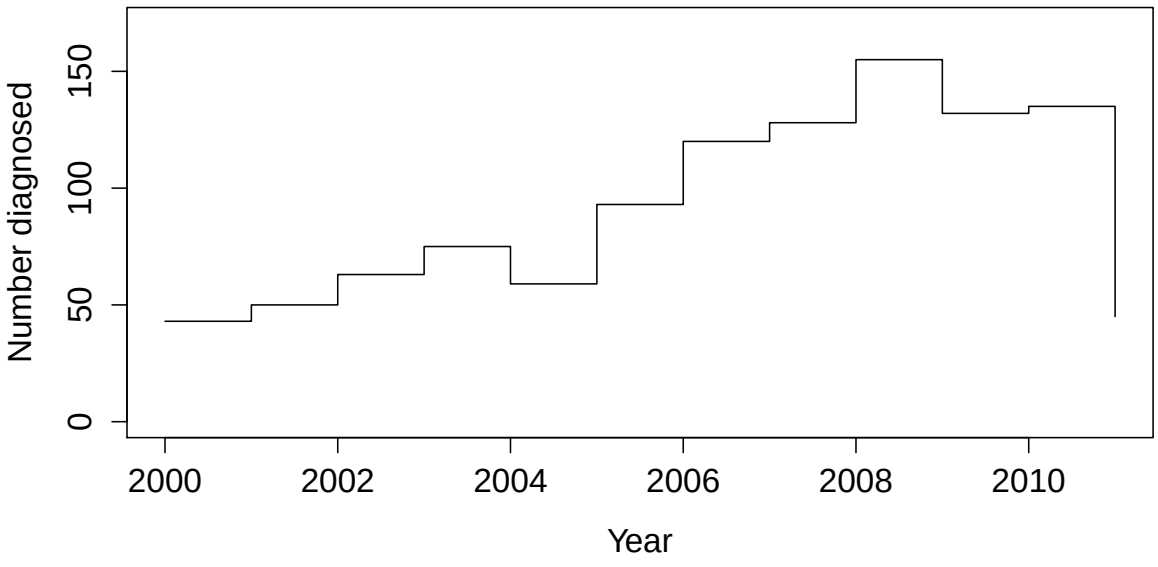

Figure 3.2: Number of newly diagnosed patients in the registry per year who do not receive hdt in their first line of treatment.

pattern does not reflect an increase of multiple myeloma, but instead the fact that as time passes more and more hospitals start entering their patients into the registry.

In the dataset there are a number of baseline variables for each patient. These include type of multiple myeloma, various biological lab values like hemoglobin and creatinine and stage of MM according to the international staging system. In this thesis we will only make use the stage of MM instead of the biological lab values. In the data there are 137 patients with stage 1 MM, 313 with stage 2 and 252 with stage 3. This leaves 396 patients with an unidentified stage.

The dataset has for each patient and line of treatment a start date for the treatment and the treatment type. The treatment type is classified into old and new therapies where approximately 29% of patients get new therapies in the first line of treatment. This changes over time, see Figure 3.3 for the proportion of new therapies over time. The distinction between new and old is one in time. Approximately around the year 2000, after a long period of no progress, a lot of new therapies were discovered. Thus treatments introduced before 2000 are classified as old and those introduced after as new.

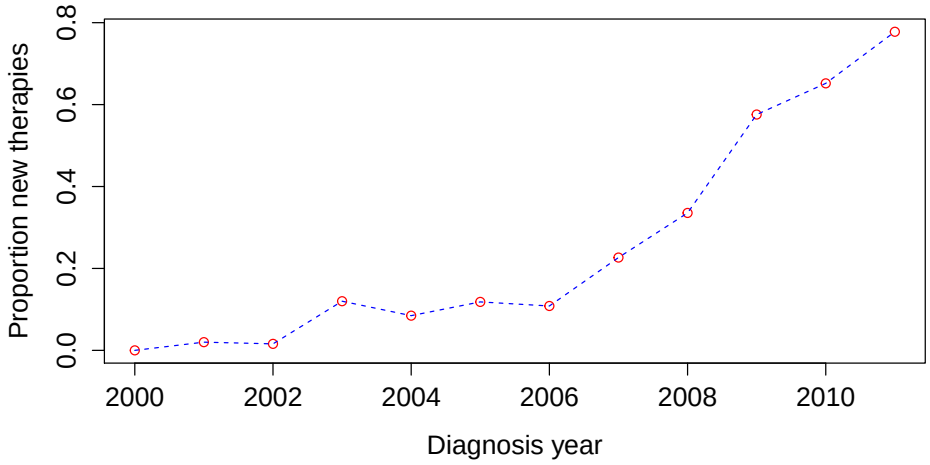

Figure 3.3: Proportion of new therapies in the first line of treatment over time for patients that do not receive hdt in their first line of treatment.

The time-to-events are specified as follows: overall survival (OS) is the time between the date of diagnosis and the date of death or the last date of follow-up in the case of censoring. Time-to-next-treatment (TTNT) is the time between the start of a treatment line and the start of the next treatment line, death or the last date of follow-up in the case of censoring. Figure 3.4 shows the Kaplan-Meier curves for OS and TTNT in line 1 for patients that do not receive hdt.

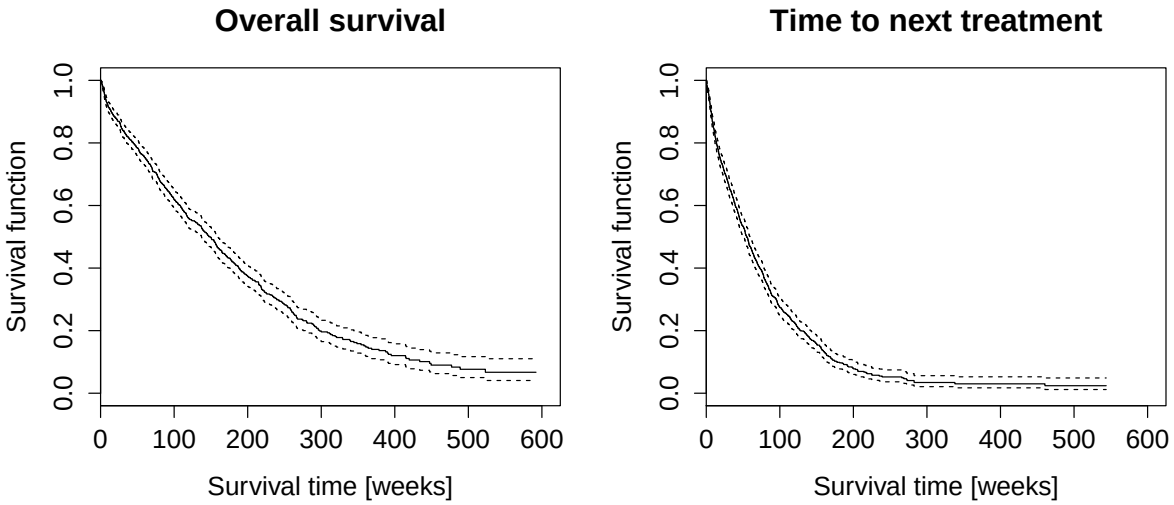

Figure 3.4: Kaplan Meier curves for overall survival and time to next treatment in line 1 for patients that do not receive hdt. The dashed lines are 95% pointwise confidence intervals calculated using Greenwood's formula, see Equation 2.1.

Chapter 4

Computational Methods

In this chapter the computational methods of the thesis are described. We begin in Section 4.1 by going through the R-packages that have been of particular use in the making of this thesis. Then we describe a simulation study in Section 4.2, where we compare the standard parametric models to the flexible parametric Royston-Parmar models.

4.1 R Programming

The computational part of this thesis, both data handling and analyses, have been performed in R. In this section the most important and time saving packages and functions that have been used are described in short.

Since this thesis has used real world data, there has been a need for data management. This has been helped by the use of the packages `dplyr` and `lubridate`. The former contains convenient functions for working with data frames and the latter contains functions for parsing and manipulating dates.

To speed up the simulations in Section 4.2 the package `parallel`, which is a part of base R, has been used. Specifically the function `mclapply` has been used for parallel computation.

Several functions from the `survival` package have been used, namely:

- `Surv` to create a survival object. These are used as the response variable in model formulas.
- `survfit` to create Kaplan-Meier survival curves.
- `coxph` to fit Cox proportional hazards models.
- `cox.zph` to test the proportional hazards assumption and to get Schoenfeld residuals. Note that to get the form we present in this thesis one has to use the argument `transform = ''rank''`.

To fit the standard parametric models we have used the function `flexsurvreg` in the package `flexsurv`. For the Royston-Parmar models we have used `flexsurvspline` in the same

package.

The function `pehaz` in the package `muhaz` is used to estimate piecewise exponential hazard functions from right-censored data, as done in Figure 5.7.

Finally, we have used the function `jacobian` in the package `numDeriv` to numerically find the gradient for the delta method in the cases when the gradient cannot be analytically found. See appendices A.2 and A.3.

All of the functions mentioned above are consistent with the theory presented in this thesis.

4.2 Simulation Study

To investigate how well the standard parametric models and the Royston-Parmar models perform, and also how they compare, we will perform a simulation study. We will simulate data from some known distributions with known parameter values and then, for different imposed follow-up times, fit the standard distributions and the Royston-Parmar models. These fitted distributions will then be compared, through some different measures, to the true model. In particular, we will pay attention to how well the models can capture the true mean time-to-event. This since often the reason for wanting the whole survival function, is to be able to calculate the mean. Thus the question is, even if we cannot capture the shape of the true distribution, can we capture the true mean?

The rough outline of our simulation study is as follows:

1. Choose patient entry and censoring distribution,
2. Choose true survival time distribution,
3. Simulate N samples from the true distribution,
4. Fit standard parametric and Royston-Parmar models to each of the samples,
5. Find estimate and confidence interval of the mean for the models in step 4,
6. Compare the result of step 5 to the population mean of the true distribution.

We will look at the performance of each model separately and also the performance of choosing models based on AIC or BIC.

4.2.1 Simulation Procedure

In this subsection we will use three subsets of the Swedish data described in Chapter 3. This is done since we want subsets that have three underlying true distributions that are at least somewhat different from each other. Furthermore, we use subsets since we want stable, reasonably homogeneous subpopulations for our purposes. These subsets are

1. TTNT for line 1 given new therapies and not given a stem-cell transplant. Ages at diagnosis range from 68 to 78 years (1st and 3rd quantiles).

2. TTNT for line 2 given old therapies and not given a stem-cell transplant in line 2. Ages at diagnosis range from 67 to 78 years (1st and 3rd quantiles).

3. TTNT for line 2 given new therapies and not given a stem-cell transplant in line 2. Ages range from 59 to 74 years (1st and 3rd quantiles).

Patient Entry and Censoring

In a clinical trial the subjects would not all enter the study at a fixed point. Instead researchers have to pick subjects during the study period as new subjects get the disease. To reflect this, we investigate the duration between arrival times in our data and then simulate from some fitting model how individuals enter the study.

In this application, it is not unreasonable to assume that arrivals follow a homogeneous Poisson process and thus the arrival times are exponentially distributed.

To find the rate we bin all subjects diagnosed between 2007-11-01 and 2009-12-31 on a weekly basis. We choose this timeline since this is a time period were all hospitals involved have collected data. Fitting a Poisson model to these weekly number of diagnoses yields $\lambda = 4.79$ new diagnosed patients per week. It is estimated that about 80% of the hospitals in Sweden are represented and thus we divide λ by 0.8 and round up to get our final number of diagnoses per week to be 6. Thus we assume that the arrivals follow a homogeneous Poisson process with parameter $\lambda = 6$. This means that the arrival times are exponentially distributed with a rate of $1/6$. In Figure 4.1 a quantile-quantile plot of the fitted Poisson distribution can be seen. It is not perfect, but close enough to be able to assume exponential arrival times in our simulation study.

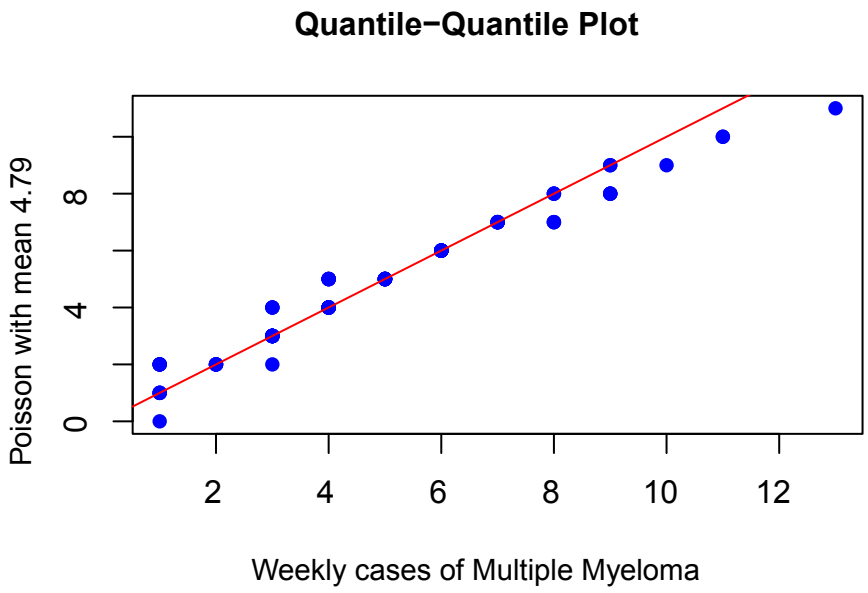

Figure 4.1: Quantile-quantile plot of the fitted Poisson distribution to the weekly incidence of multiple myeloma in the Swedish data.

Choice of True Distributions

We will investigate three different possible data generating distributions. These are chosen based on being the best, relatively by AIC, fitting models to the subsets described above. The following are the models, fitted to the data described above in the same order:

1. Exponential with rate $\beta = 0.0081$ and an expected value of 123.4.
2. Weibull Royston-Parmar model with 1 knot at $\log t = 3.42$ and an expected value of 55.4.
3. Truncated log-normal with $\mu = 3.82$, $\sigma = 1.39$ and $\tau = 2000$. The expected value is 109.0.

The probability density functions of these distributions can be seen in Figure 4.2.

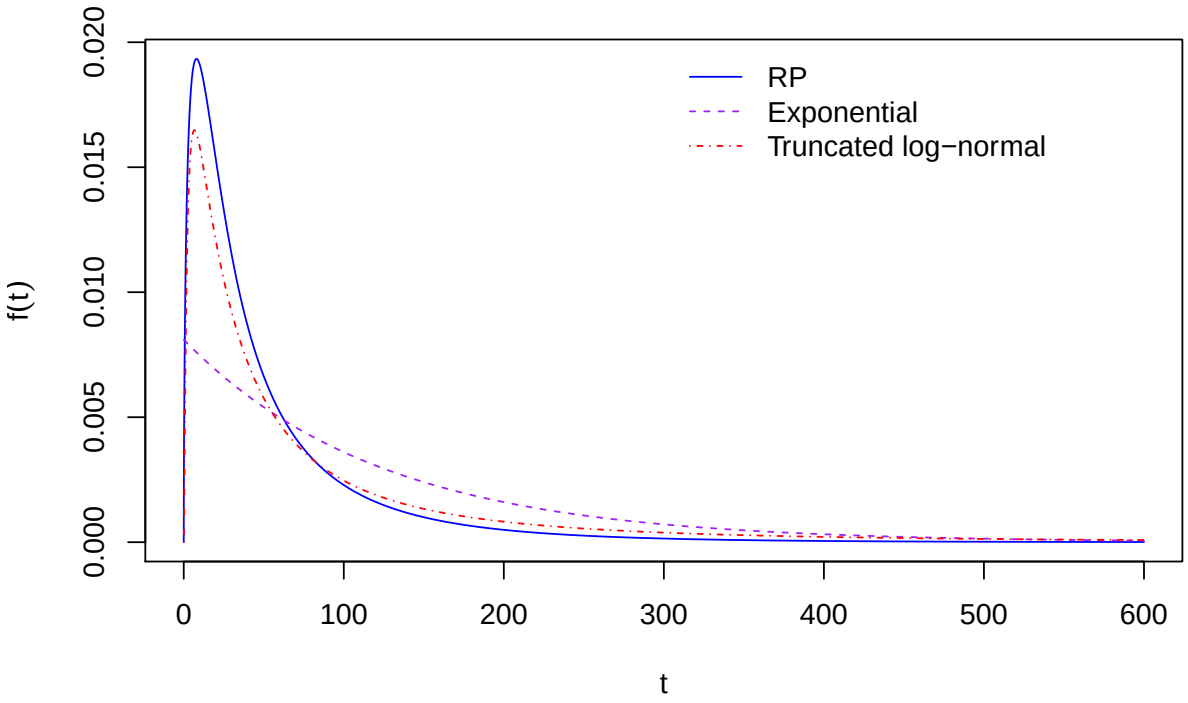

Figure 4.2: Probability density functions of the models that are used to simulate data from in the simulation study.

Confidence Intervals of the Mean

To find confidence intervals for the mean survival time we use the asymptotic normal distribution of the maximum likelihood estimator (MLE) and the delta method, see Appendix A.2. If there is no closed form equation for the gradient, a numerical approximation is carried out by Richardson's extrapolation (Richardson & Gaunt, 1927), see Appendix A.3.

4.2.2 Simple First Example

We will now go through a motivating and simple example to see how well estimation of the mean survival time is possible for different followup times and sample sizes in the best case scenario.

We will simulate data from an exponential distribution with mean 1 and then fit an exponential distribution to the data.

First, let us go through the theoretically derived relationship between the asymptotic distribution of the MLE and the sample size and censoring time. The likelihood is

$$L(\mu; \boldsymbol{X}, \boldsymbol{\delta}) = \prod_{i=1}^{n} f(x_i)^{\delta_i} S(x_i)^{1-\delta_i}.$$

Separating the sample into an observed sample $\boldsymbol{Y}$ of size n_o and a censored sample $\boldsymbol{Z}$ of size n_c yields the likelihood

$$L(\mu; \boldsymbol{Z}, \boldsymbol{Y}) = \prod_{i=1}^{n_o} f(y_i) \prod_{i=1}^{n_c} S(z_i).$$

If we have a fixed censoring time c, then all observations $z_i = c$, $i = 1, ..., n_c$ and all observations $y_i < c$, $i = 1, ..., n_o$. Thus the log-likelihood is

$$\begin{aligned} l(\mu; \boldsymbol{Z}, \boldsymbol{Y}) &= \sum_{i=1}^{n_o} \log f(y_i) + \sum_{i=1}^{n_c} \log S(c) \\ &= \sum_{i=1}^{n_o} \left(-\frac{y_i}{\mu} - \log \mu \right) + \sum_{i=1}^{n_c} -\frac{c}{\mu} \\ &= -\frac{n_o \bar{y}}{\mu} - n_o \log \mu - \frac{n_c c}{\mu} \\ &= -\frac{n_o \bar{y} + n_c c}{\mu} - n_o \log \mu. \end{aligned}$$

From this we can get the score function

$$S(\mu; \boldsymbol{Z}, \boldsymbol{Y}) = \frac{n_o \bar{y} + n_c c}{\mu^2} - \frac{n_o}{\mu}.$$

Setting equal to 0 and solving for μ yields

$$\hat{\mu}_{\text{MLE}} = \frac{n_c \bar{y} + n_c c}{n_o}.$$

The Fisher information is

$$I(\mu) = \mathrm{E}\left[2\frac{n_o \bar{y} + n_c c}{\mu^3} - \frac{n_o}{\mu^2} \middle| \mu \right].$$

To calculate this we first need the expectation

$$\begin{aligned} \mathrm{E}[Y] &= \mathrm{E}[X | X < c] \\ &= \int_0^c y \frac{f_X(y)}{F_X(c)} \mathrm{d}y \\ &= \mu - \frac{c e^{-\frac{c}{\mu}}}{1 - e^{-\frac{c}{\mu}}}. \end{aligned}$$

Secondly we need to calculate the expected number of censored observations,

$$\mathrm{E}[n_c] = nP(X \geq c) = n e^{-\frac{c}{\mu}}.$$

Using this we get the Fisher information

$$I(\mu) = \frac{n}{\mu^2}\left(1 - e^{-\frac{c}{\mu}}\right).$$

Thus the asymptotic distribution of the MLE is, see Lehmann & Casella (1998),

$$\hat{\theta}_{\text{MLE}} \sim N\left(\mu, \frac{\mu^2}{n\left(1 - e^{-\frac{c}{\mu}}\right)}\right).$$

If we censor at a specific quantile we can rewrite this quite nicely. The quantile function of the exponential distribution is

$$F^{-1}(p) = -\mu \log(1 - p).$$

Inserting $c = c(p) = F^{-1}(p)$ yields

$$\hat{\theta}_{\text{MLE}} \sim N\left(\mu, \frac{\mu^2}{np}\right).$$

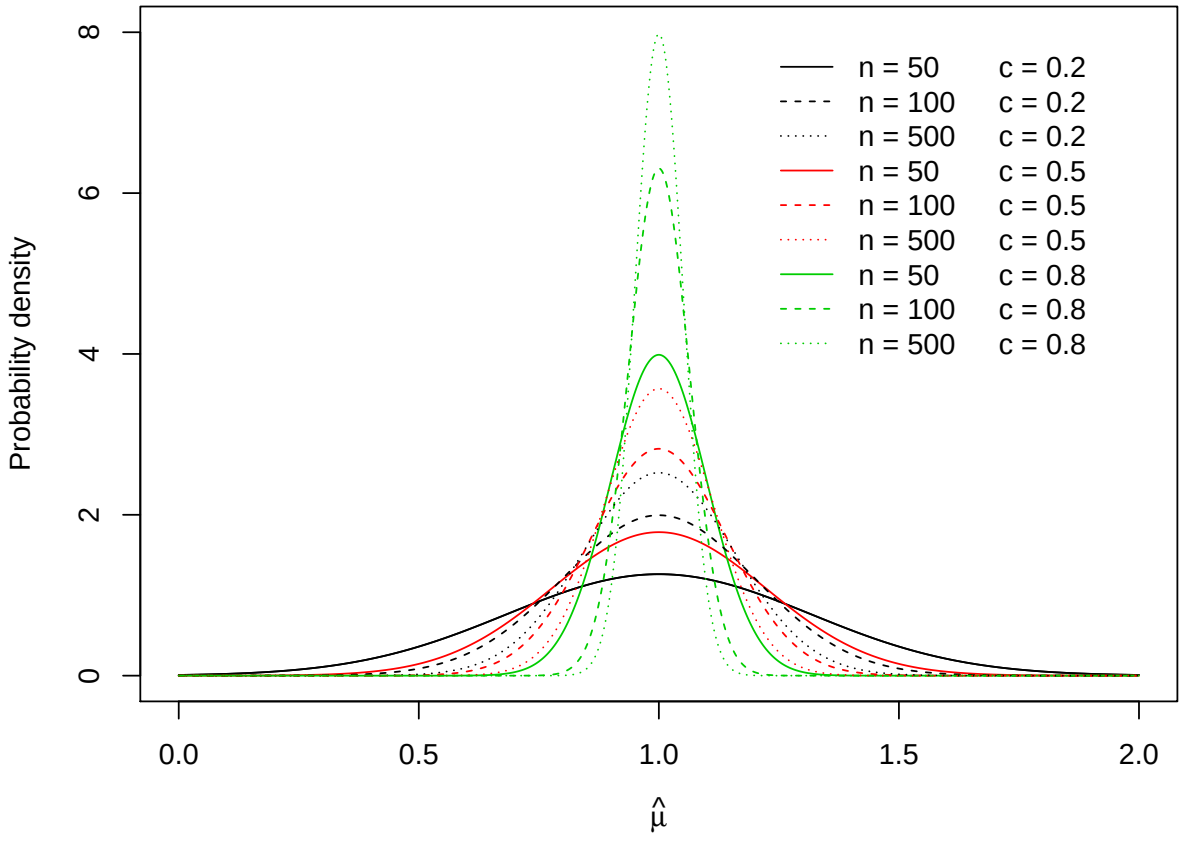

Figure 4.3: Density of the asymptotic normal distribution of the MLE of μ for different sample sizes n and different censoring quantiles c.

In Figure 4.3 we can see the asymptotic normal distribution of the MLE estimate of μ for different sample sizes n and different censoring quantiles c. It is no surprise that having a larger censoring quantile is more important than having a large sample size. 50 observations with

censoring at the 0.8 quantile yields a tighter density than 500 observations with censoring at the 0.5 quantile.

Now we simulate samples of size $n = 50, 100, 500, 1000$ and with censoring quantiles $c = 0.2, 0.5, 0.8$. For each combination we simulate 10000 samples and then compute 95% confidence intervals of the mean for each of these samples. In Figure 4.4 the coverage and average length of these confidence intervals can be seen. The coverages are consistent with the nominal value and the average lengths of the confidence intervals are consistent with the theoretically derived lengths, as indicated by the dashed lines in Figure 4.4. Here again we see that having a reasonably long follow-up is quite important. In Figure 4.4 we can also see quantile-quantile plots of the shortest and longest follow-up time. With a short follow-up and small sample size the asymptotic result is far from fulfilled, but except for that case, the asymptotic result seems to approximately hold.

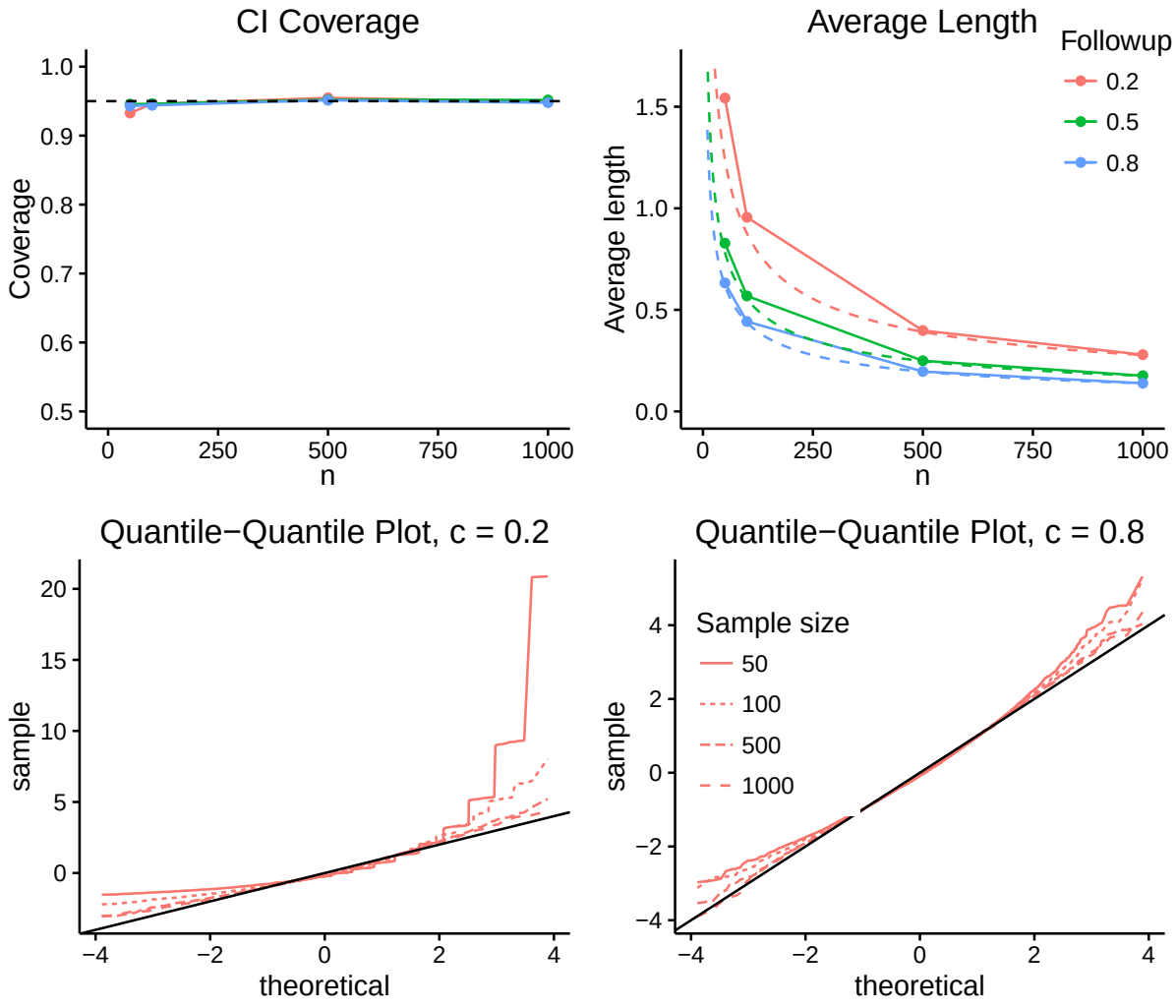

Figure 4.4: Upper: Coverage and average length of 95% confidence intervals of μ for the exponential distribution for different censoring quantiles $c = 0.2, 0.5, 0.8$ and different sample sizes $n = 50, 100, 500, 1000$. The dashed lines indicate the theoretical confidence interval lengths. Lower: Quantile-quantile plots of the shortest and longest follow-up times.

We can now do the same thing for a log-normal distribution. This will be done numerically

using `optim` in `R`. For simplicity we reparametrize the log-normal distribution by replacing μ and σ with the mean

$$m = e^{\mu + \frac{1}{2}\sigma^2},$$

and the coefficient of variation

$$cv = \sqrt{e^{\sigma^2} - 1}.$$

Now we simulate from a log-normal with $\mu = -\frac{1}{2}\log 2$ and $\sigma = \log 2$, that is, the mean and the variance is the same as for the exponential distribution with mean 1. The coverage and average length of the 95% confidence intervals can be seen in Figure 4.5. The same conclusions as for the exponential example hold for this case. We can also see the quantile-quantile plot of the shortest and longest follow-up. Here the asymptotic result certainly does not hold for the shortest follow-up and sample sizes of 50 and 100 observations. The rest are closer, but the asymptotics are much slower here than for the exponential distribution.

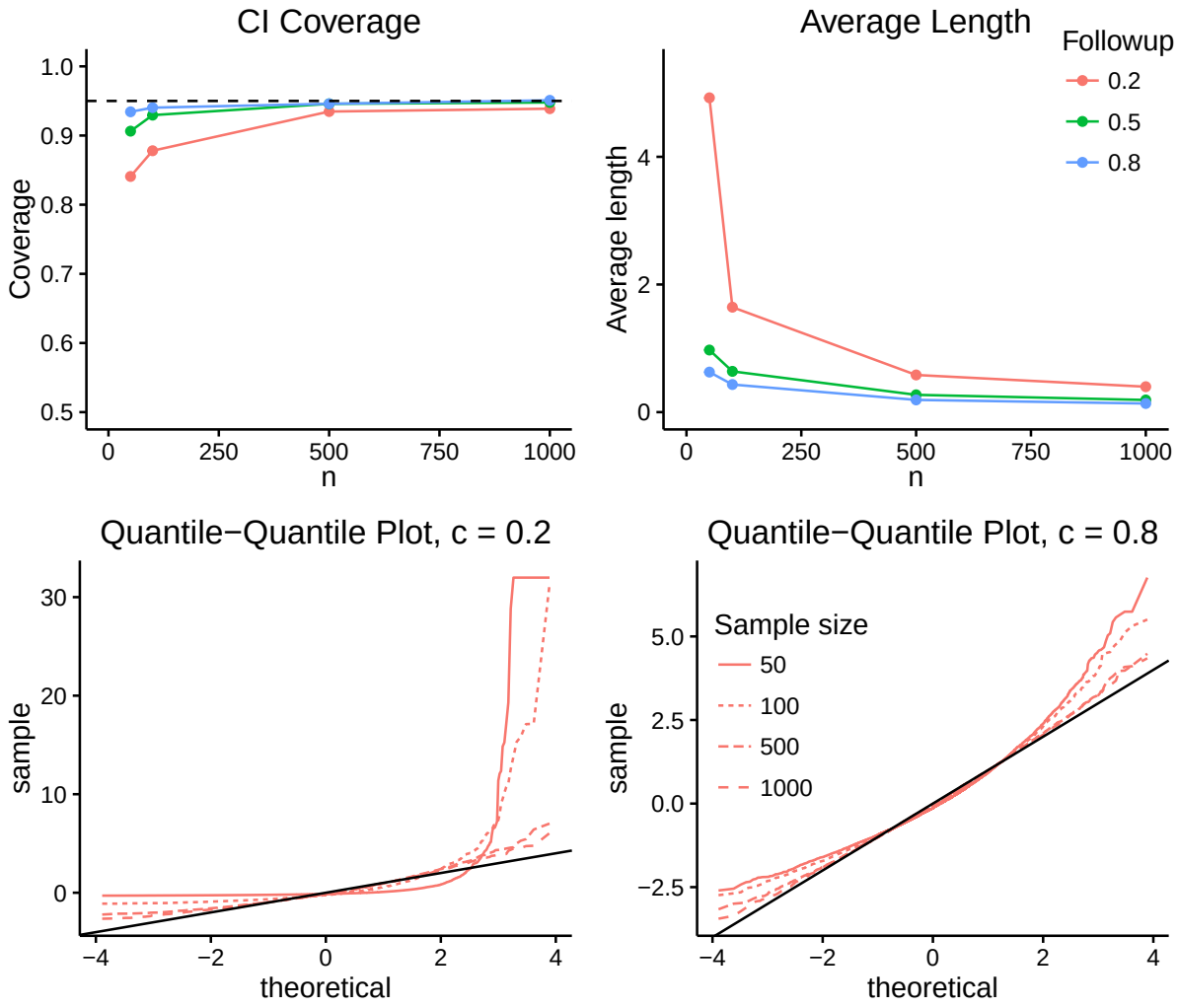

Figure 4.5: Upper: Coverage and average length of 95% confidence intervals of μ for the log-normal distribution for different censoring quantiles $c = 0.2, 0.5, 0.8$ and different sample sizes $n = 50, 100, 500, 1000$. Lower: Quantile-quantile plots of the shortest and longest follow-up times.

4.3 Simulation Result

In this section we present some of the results of the simulation study described in Section 4.2. The total run-time for the simulations was approximately 16 hours. In tables B.1-B.6 in Appendix B summary tables for the simulations can be seen, one for each imposed follow-up. They contain the mean, median, 2.5% and 97.5% quantiles of the biases in the mean estimates, coverage of 95% confidence intervals and their mean and median lengths. Furthermore they contain biases of the estimators for the 25%, 50% and 75% quantiles of the distributions. Finally they have the number of observations and the proportion of these that are not censored. In Figures 4.6-4.8 kernel density estimates, see Appendix A.5, of the biases for each combination of true model, fitted model and follow up time can be seen. From these figures we see that selecting the model by minimum AIC gives estimates of the mean with seemingly low bias, unless we have a very short follow-up time.

Royston-Parmar: The RP model captures the true population mean quite well when the exponential is the true model. Although for the shortest follow-up the range of the biases are quite wide. When the RP model is the true model, fitting the RP model does seem to perform quite well, except again for the two shortest follow-ups which yield biased estimates of size slightly below a quarter of the true mean. It performs the worst if the true model is the truncated log-normal. We get a median bias of slightly below half the true expected value for the two shortest follow-ups. By looking at Figures 4.6-4.8 we see that with sufficient follow-up the RP models seem flexible enough to not be as severely biased as a misspecified standard parametric model can be. Furthermore, when the true model does not lie within the space of fitted models, the RP models can be a good choice.

Not surprisingly, a longer follow-up should always be desired for accuracy. It not only gives tighter distributions of the biases, but it also decreases the amount of bias for models that are misspecified.

Rule-of-thumb: One interesting question is if we can make a rule-of-thumb for how much follow-up is needed for accurate results when choosing model by minimum AIC. We will now take a look at the performance for different follow-ups in terms of their relative position to the median. For the below interpretations we make use of tables B.1-B.6. We will look at the median biases and median confidence interval lengths, as opposed to the mean values. This since AIC can pick a log-normal with an unreasonably high mean as the best model, giving extreme values that we would most likely not trust in reality. Therefore, the median values are probably more informative in the sense of what we would see in real data modeling with visual tools. In the following paragraphs we will be referring to the median length of 95% confidence intervals when we say length.

For the **truncated log-normal** the true median is 46 weeks and the true mean is 109 weeks. With 26 weeks of follow-up we have a 68% coverage of the true mean with a length of the confidence interval (CI) of 141 weeks. The median bias in the estimate of the expected value

is 20 weeks, which is approximately a fifth of the true mean. With a follow-up of 52 weeks the coverage is 84% with a length of the CI of 90 weeks. The median bias is less than one week. With longer follow-up than this we get a slightly higher coverage and smaller lengths of the CIs while keeping approximately the same amount of bias. Thus it seems as if 52 weeks of follow-up is sufficient to have unbiasedness, although more follow-up is desirable to shrink the CIs. With 104 weeks of follow-up and longer the 2.5% and 97.5% quantiles of the biases are within 40 weeks. Thus the "worst case" scenarios yield biases of slightly more than a third of the true mean. With the longest tested follow-up of 260 weeks the worst case biases are approximately within a half year, or approximately a fifth of the true mean.

For the **exponential** the true median is 86 weeks and the true mean is 123 weeks. With the two shortest follow-ups the median biases of the estimators of the true mean are slightly below 10% of the true mean. With longer follow-up the absolute value of the median bias is always below one week. The coverages of the CIs are always close to 90% for all follow-up times, while the lengths shrink with additional follow-up. For the two shortest follow-ups the lengths are 143 and 68 weeks, respectively. With 104 weeks of follow-up the length of the CI has decreased to 35 weeks, while for the longest follow-up it is 16 weeks. Thus it seems as if 104 weeks of follow-up is necessary to achieve unbiasedness and reasonable accuracy.

For the **RP model** the true median is 33 weeks and the true mean is 55 weeks. The coverages of the CIs are in general worse when the RP model is the true model compared to the above cases. We have coverages of 72%, 52% and 71% for the three shortest follow-ups, respectively. Once we reach 156 weeks we have 86% coverage and the coverage increases with longer follow-up. On the other hand, the absolute value of the median bias is below 3 weeks regardless of the follow-up time. For the follow-up times, listed in increasing order, the lengths of the CIs are approximately, in relation to the true mean, of size one, one half, a quarter, slightly above a fifth, slightly below a fifth and once again slightly below a fifth. A thing one should take notice to is the fact that in this case, the true model does not lie within the space of fitted models as it did for the truncated log-normal and the exponential. To clarify, knots are placed evenly spaced on the quantiles of the observed log survival times. The true knot is positioned at approximately 30 weeks and thus it would not be possible with the shortest follow-up to fit the true model.

In conclusion it seems as if a rule-of-thumb concerning the amount of follow-up needed could be that we need approximately the length of the true median to get unbiased estimates of the mean and not excessively wide confidence intervals. This will be investigated further in the analysis of the real world data in Chapter 5.

Quantiles: With censoring the mean is of course difficult to estimate, on the other hand, the quantiles of the distribution, especially those actually observed, should be easier to estimate. In tables B.1-B.6 we see that the median biases of the estimates for the 25, 50 and 75% quantiles are relatively close to zero for the minimum AIC models. The 75% quantile being the one with the most bias.

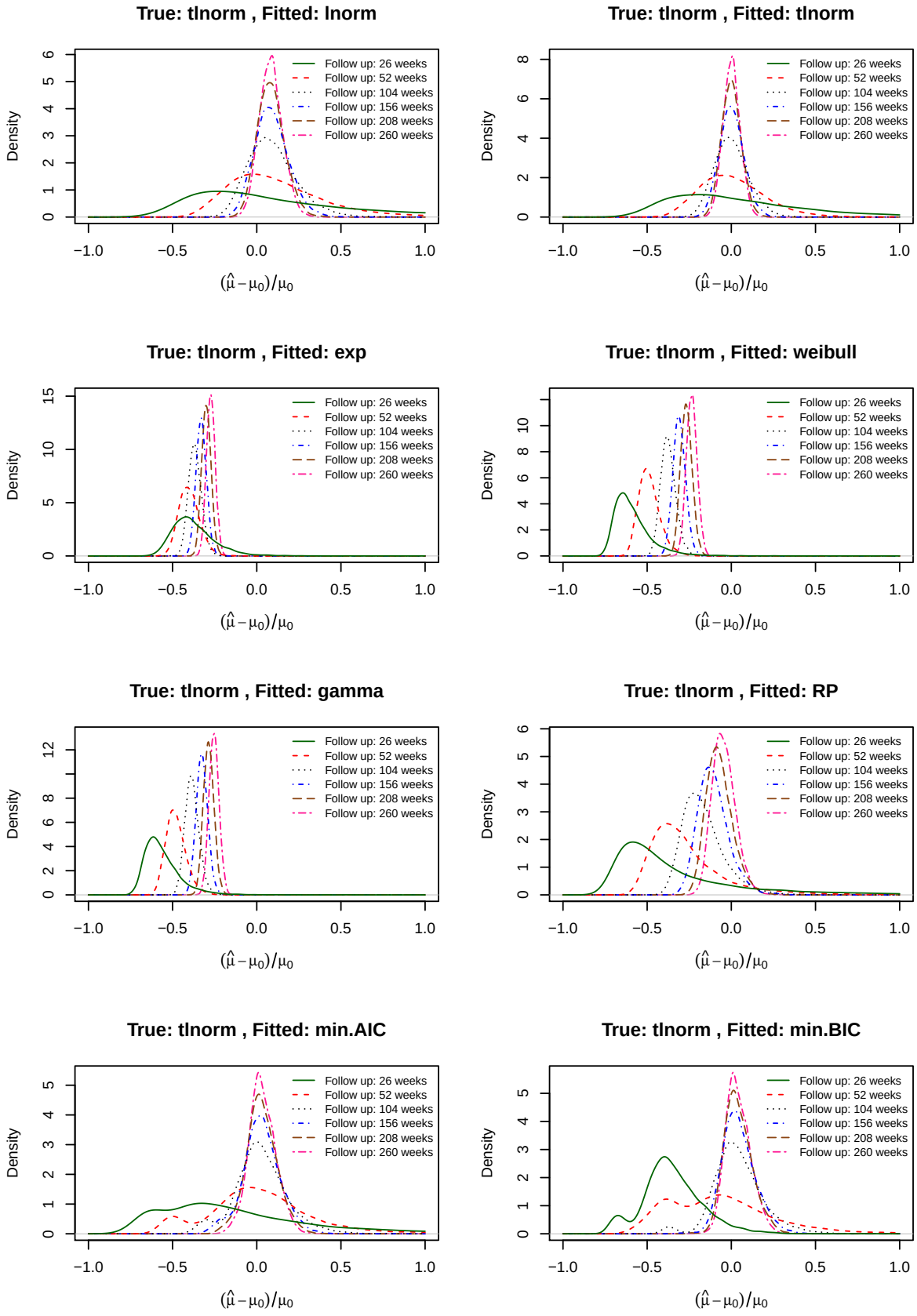

Figure 4.6: Kernel density estimates of the bias in the mean estimates in terms of the true mean. The true model is the truncated log-normal distribution. The title on each figure indicates the fitted model.

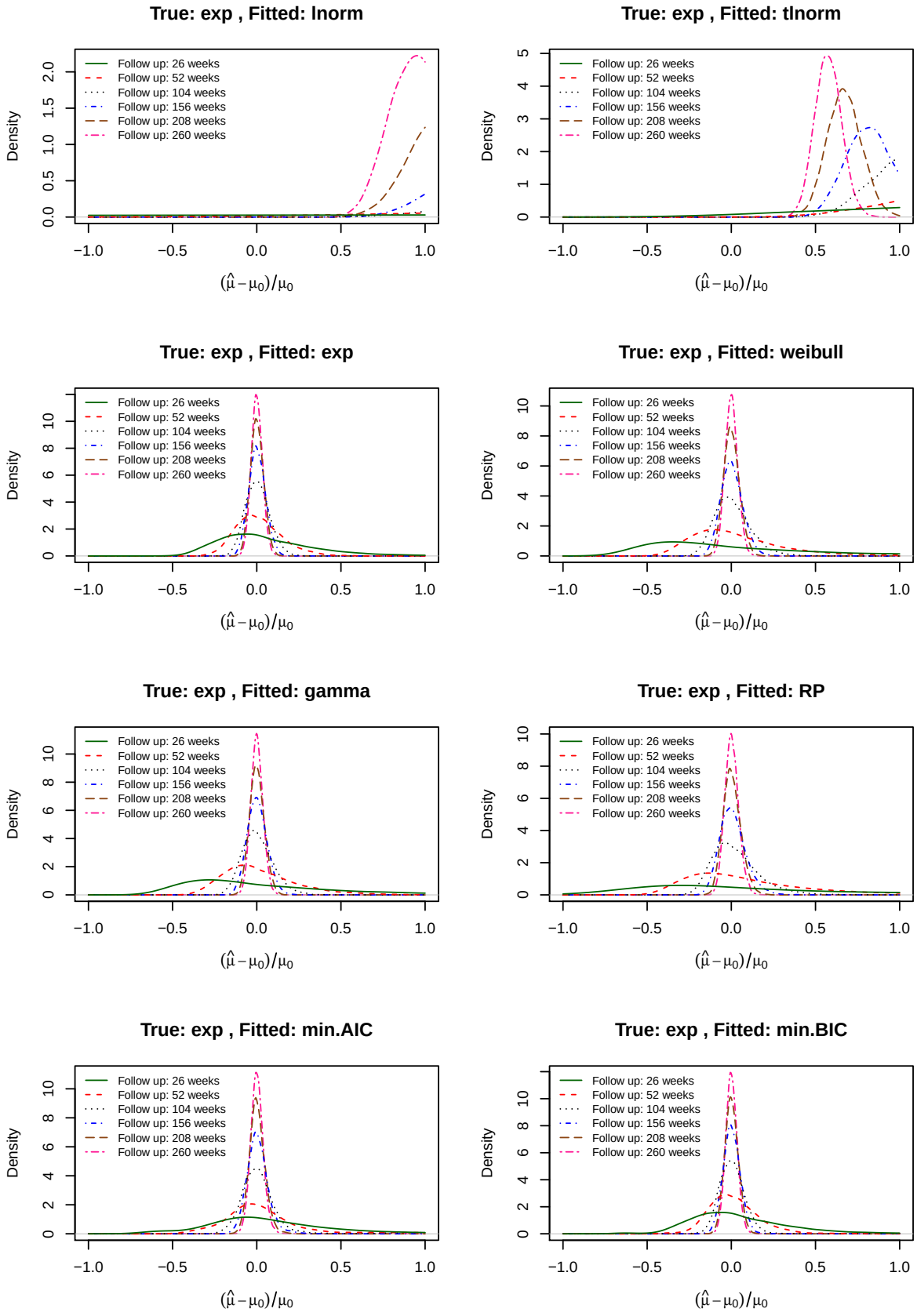

Figure 4.7: Kernel density estimates of the bias in the mean estimates in terms of the true mean. The true model is the exponential distribution. The title on each figure indicates the fitted model.

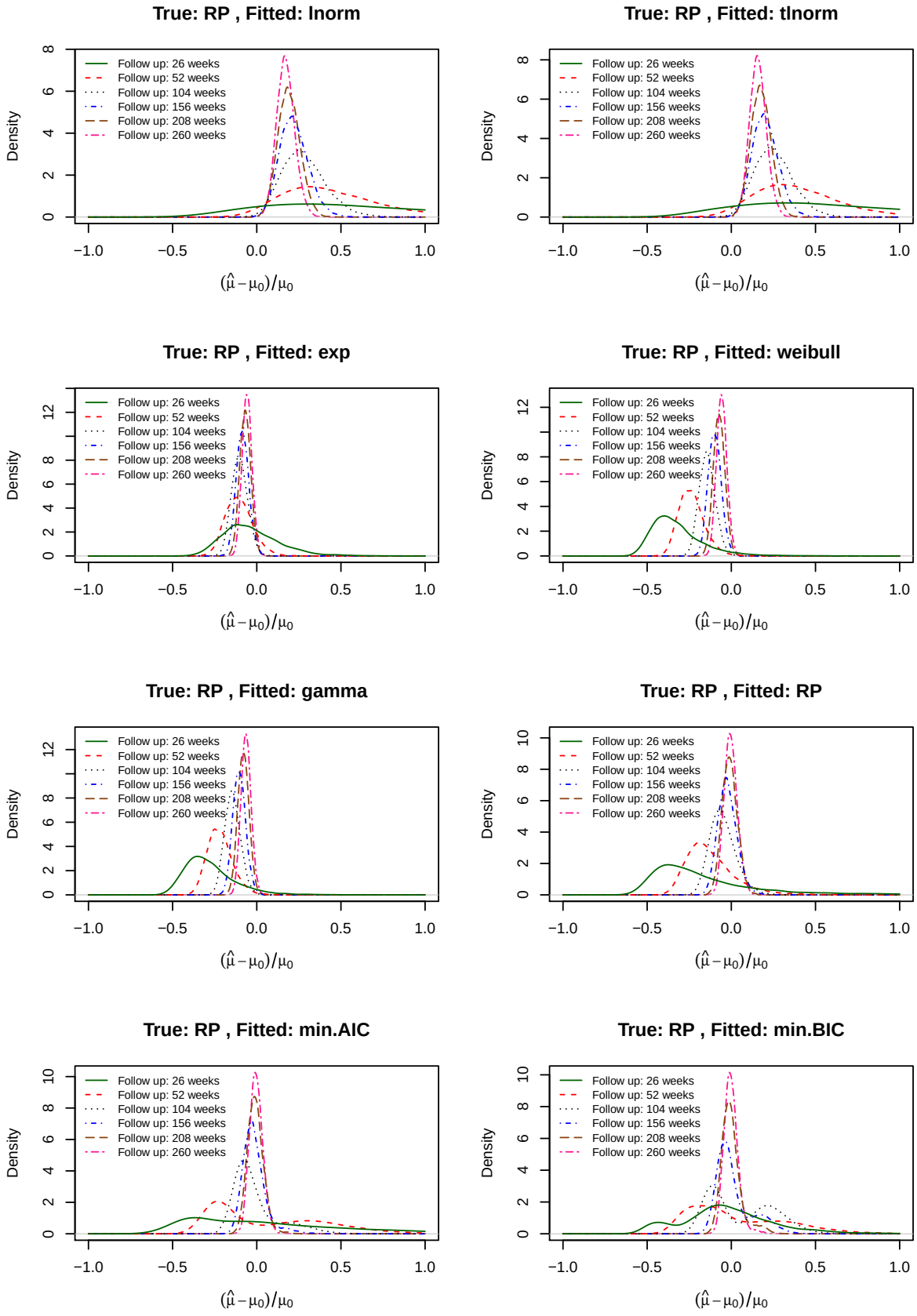

Figure 4.8: Kernel density estimates of the bias in the mean estimates in terms of the true mean. The true model is the Weibull RP model with one knot. The title on each figure indicates the fitted model.

Chapter 5

Data Analysis

In this chapter we will analyze the Swedish multiple myeloma dataset described in Chapter 3. We begin the following sections by examining how much follow-up is needed to accurately fit a distribution to the data. After that we take a look at modeling with explanatory variables using the Cox proportional hazards model and RP models and try to compare these. For simplicity, we will look at overall survival and time to next treatment for the first line of treatment.

5.1 Swedish Data

We begin by splitting the data into two subsets according to the subjects receiving new or old therapies. We then use minimal AIC and BIC to choose models for overall survival and time to next treatment for subjects in their first line of treatment, with different imposed follow-up times. In the model space we have the log-normal, truncated log-normal, log-logistic, exponential, gamma, Weibull, Weibull RP with 1, 2 and 3 knots and log-logistic RP with 1, 2 and 3 knots. In figures 5.1 and 5.2 the results can be seen for time to next treatment with new and old therapies, respectively. In figures 5.3 and 5.4 the results can be seen for overall survival with new and old therapies, respectively. The median survivals, estimated from the Kaplan-Meier curves, are 81, 47, 248 and 129 weeks, respectively. In all figures, it seems to be the case that we can catch the functional form of the data accurately as soon as we have approximately a follow-up time greater than the median survival time. Note that for subjects receiving new therapies we have limited follow-up time for overall survival and it is thus hard to compare the fitted models to some form of truth.

Figure 5.5 shows the Kaplan-Meier curves with models chosen by minimum AIC and minimum BIC using the whole data for TTNT and OS stratified on old and new therapies. For overall survival the exponential distribution seems to be adequate for both old and new therapies, which would indicate that the relationship between old and new therapies are on the proportional hazards scale. Thus the Cox proportional hazards model or the Weibull RP model might be appropriate for this case. For time-to-next treatment the two therapies seem to be on different scales and one would probably want to do a stratified analysis. In the following we will thus be

satisfied with modeling the overall survival with explanatory variables.

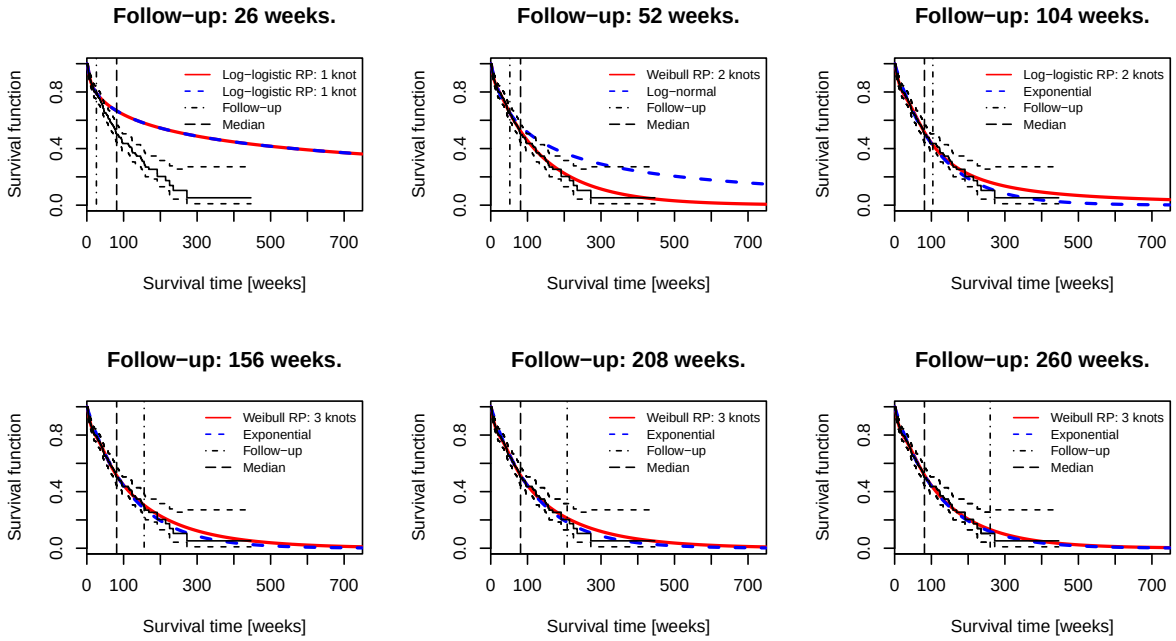

Figure 5.1: Time to next treatment for patients given new therapies for different imposed follow-up times. The median event time is 81 weeks. Models are chosen by minimum AIC (solid line) and minimum BIC (dashed line) and plotted together with the Kaplan-Meier curve for the full data.

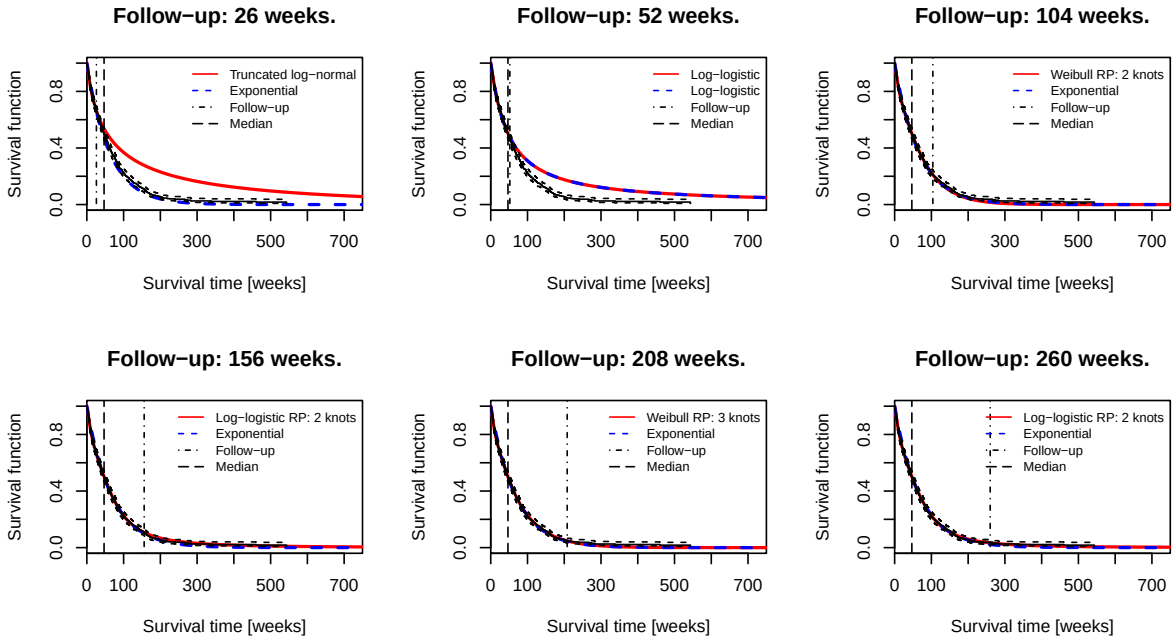

Figure 5.2: Time to next treatment for patients given old therapies for different imposed follow-up times. The median event time is 47 weeks. Models are chosen by minimum AIC (solid line) and minimum BIC (dashed line) and plotted together with the Kaplan-Meier curve for the full data.

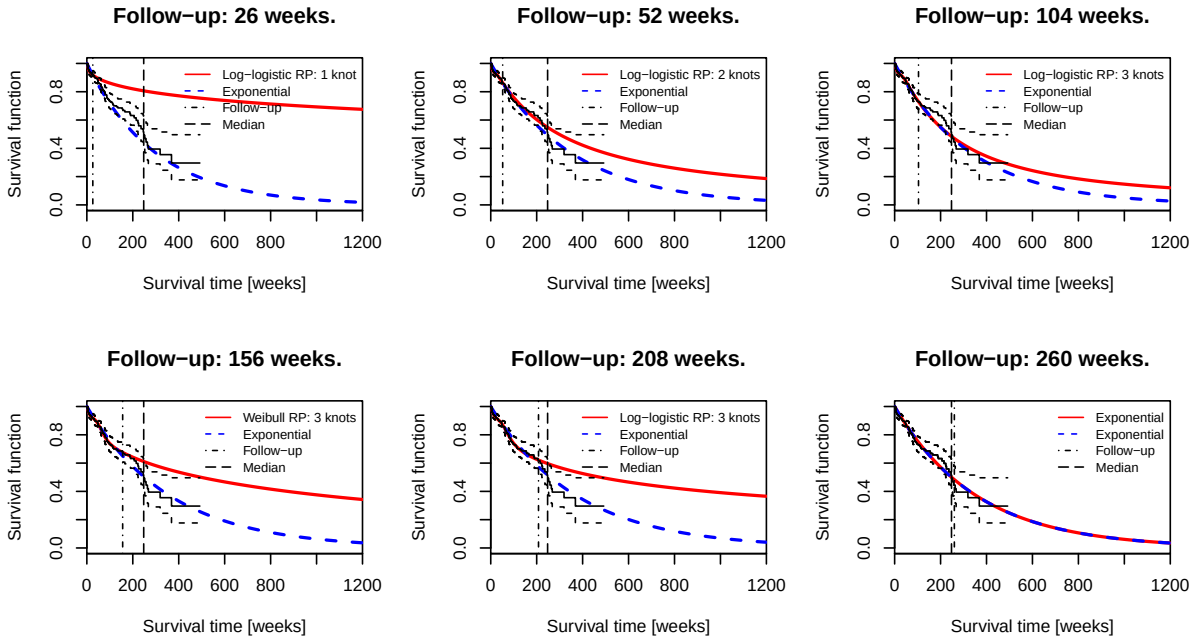

Figure 5.3: Overall survival for patients given new therapies for different imposed follow-up times. The median survival time is 248 weeks. Models are chosen by minimum AIC (solid line) and minimum BIC (dashed line) and plotted together with the Kaplan-Meier curve for the full data.

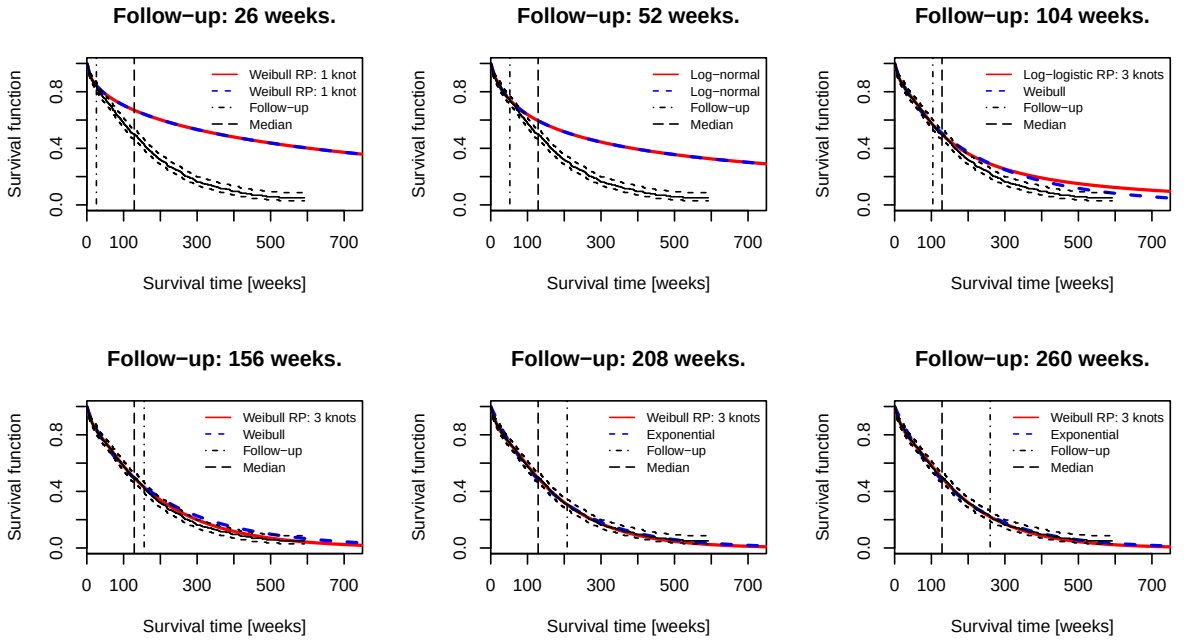

Figure 5.4: Overall survival for patients given old therapies for different imposed follow-up times. The median survival time is 129 weeks. Models are chosen by minimum AIC (solid line) and minimum BIC (dashed line) and plotted together with the Kaplan-Meier curve for the full data.

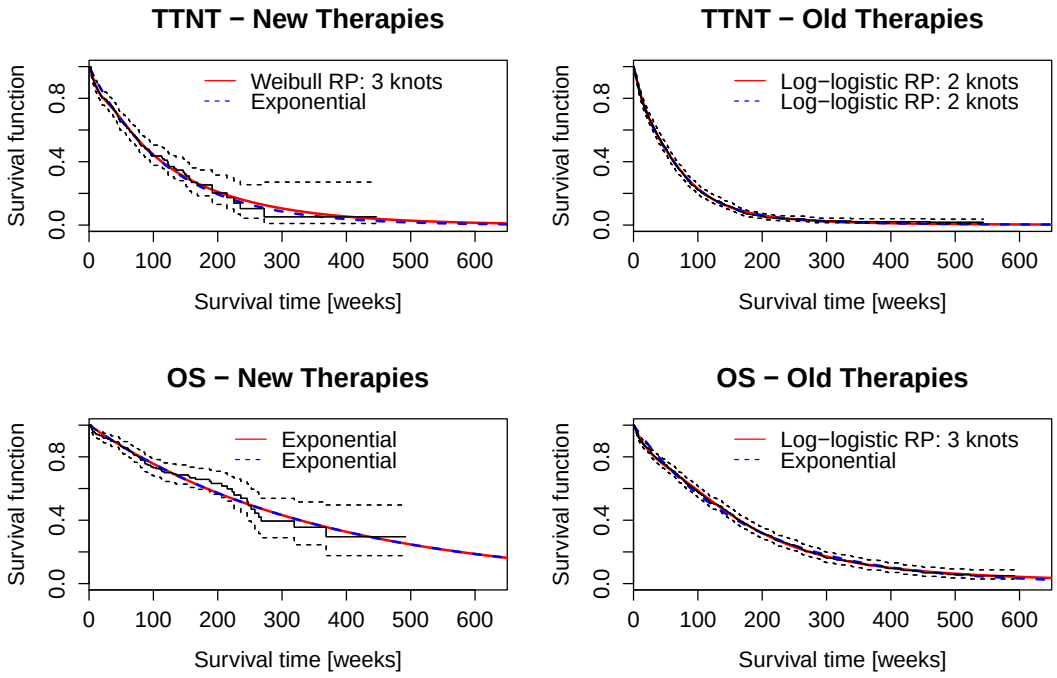

Figure 5.5: Time to next treatment in line 1 and overall survival for patients given old or new therapies. Models are chosen by minimum AIC (solid line) and minimum BIC (dashed line) and plotted together with the Kaplan-Meier curve.

5.1.1 Modeling with Explanatory Variables

We will now model the overall survival with explanatory variables. The explanatory variables considered here are type of therapy, age and stage of MM. We start by looking at each variable separately in univariable Cox proportional hazards models. The results can be seen in Table 5.1. The table contains the estimated coefficients, $\hat{\beta}$, for the Cox proportional hazards models as well as Wald p-values of the test of the null hypothesis $H_0 : \beta = 0$. In addition it contains the correlation coefficient $\hat{\rho}$ together with p-values of the test $H_0 : \rho = 0$, as described in Subsection 2.8.3 to test the proportional hazards assumption. For stage 3 and missing stage we reject the null hypothesis, but the estimated correlation coefficients are quite small and we do not necessarily have a great deviation from the proportional hazards assumption. We move on to include all three variables in a multivariable Cox proportional hazards model, see Table 5.2 for the result. In stage "missing" we see a decrease of the estimated coefficient while for new therapy we see an increase, compared to the univariable models. This is to be expected since 14% of patients with missing stage get new therapies while 43, 38 and 35% get new therapies for stage 1, 2 and 3, respectively. We do not know why it is the case that the patients with missing stage tend to get older therapies. It could be that the year of diagnosis for the patients with missing stage lie further back in time, when new therapies were not as widely available. The median year of diagnosis is 2006 for these patients while it is 2007 for patients with available stage. It could also be that these patients are older. The median age at diagnosis of patients with missing stage is 76, while it is 74 for those that have a measured stage. For the test of the proportional hazards assumption the same conclusions as for the univariable analysis hold. The Schoenfeld residuals

against time can be seen in Figure 5.6. There seems to be slight deviations from a random pattern over time for some of the variables, but not enough to warrant any great concern in the validity of the proportional hazards assumption. The solid line is a local polynomial regression line, where the fit for a point x is made using points in a neighborhood of x. These neighboring points are weighted by their distance from x. For this we have used the `loess` function in `R`. This line seems to have approximately a slope of 0 for all variables, indicating that the residuals are evenly distributed over time.

Table 5.1: Univariable Cox proportional hazards models for overall survival.

Variable	$\hat{\beta}$	$\exp(\hat{\beta})$	Wald p-value	$\hat{\rho}$	PH test p-value
New therapy	-0.71	0.49	$9 \cdot 10^{-11}$	-0.01	0.74
Age	0.04	1.04	3.2^{-13}	0.04	0.22
Stage 2	0.61	1.84	$9.8 \cdot 10^{-5}$	-0.06	0.10
Stage 3	1.24	3.46	$4.1 \cdot 10^{-15}$	-0.15	10^{-4}
Stage missing	0.94	2.56	$3.1 \cdot 10^{-10}$	-0.11	$5.3 \cdot 10^{-3}$

Table 5.2: Multivariable Cox proportional hazards models for overall survival.

Variable	$\hat{\beta}$	$\exp(\hat{\beta})$	Wald p-value	$\hat{\rho}$	PH test p-value
New therapy	-0.64	0.53	$1.3 \cdot 10^{-8}$	-0.02	0.69
Age	0.03	1.03	$2.1 \cdot 10^{-10}$	0.07	0.03
Stage 2	0.54	1.72	$5.6 \cdot 10^{-4}$	-0.06	0.11
Stage 3	1.22	3.39	$1.4 \cdot 10^{-14}$	-0.13	$7.8 \cdot 10^{-4}$
Stage missing	0.76	2.14	$4.1 \cdot 10^{-7}$	-0.09	0.02

The next step is to fit an RP model with these covariates. As seen in Figure 5.5 the best model, according to BIC, for overall survival stratified on received therapy was an exponential model. We will therefore look at fitting a Weibull RP model with the covariates added linearly and the number of knots determined by AIC or BIC. For completeness we will also investigate a log-logistic RP model.

When using AIC to determine the number of knots for the Weibull RP model we get a model with 6 knots. This model has quite a complex shape and it does seem to be overfitted to the data. Using BIC to determine the number of knots we get a model with 3 knots, which seems a bit more appropriate for our data. For the log-logistic RP model we settle on 2 knots, determined by BIC. The log-logistic RP model has higher AIC and BIC than the Weibull RP models. In Figure 5.7 the estimated hazard for the Weibull RP model with 3 knots can be seen for the average covariate values together with a non-parametric estimator of the hazard. The non-parametric estimator is the one used in the function `pehaz` in the package `muhaz`. It divides time into equally spaced bins and estimates the hazard by the number of events in each bin divided by the width of the bin.

In Table 5.3 the estimates of the fitted RP models with 0, 3 and 6 knots, respectively, can be seen. The coefficients for our explanatory variables are the same in the latter two models. Comparing these with the ones from the multivariable Cox proportional hazards model in Table 5.2 shows that they are basically the same here as well. There is a slight difference in the estimates when

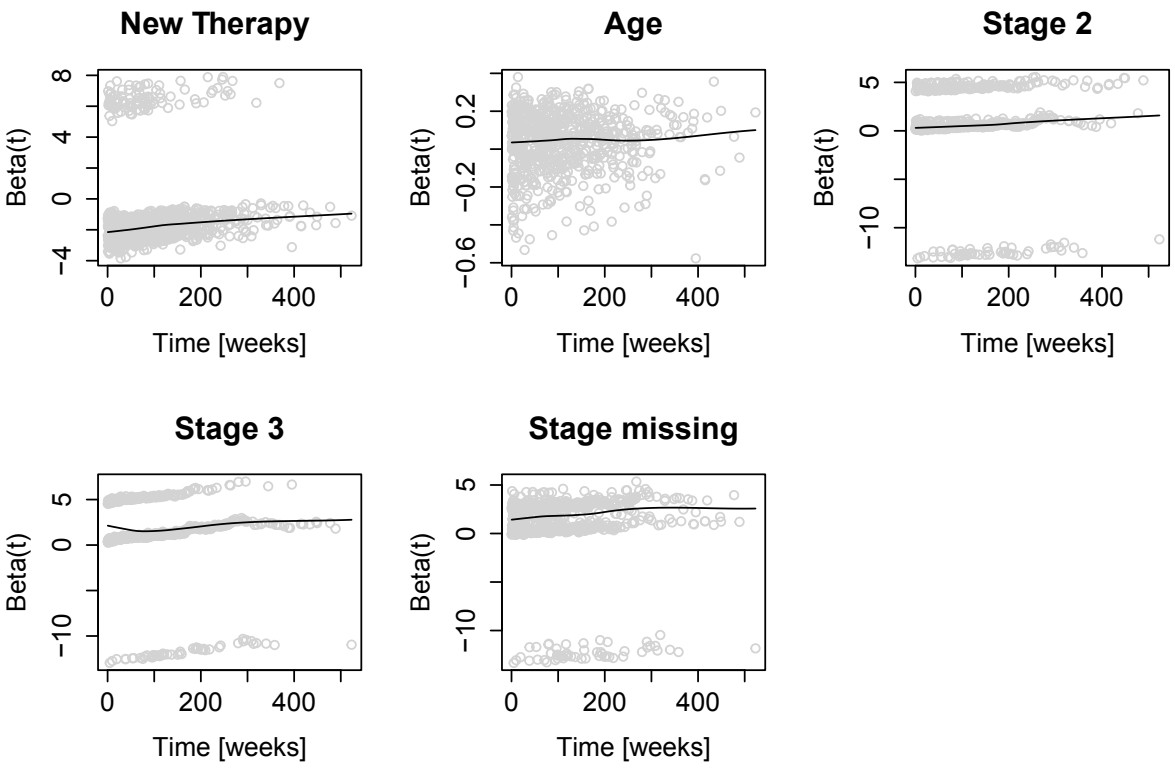

Figure 5.6: Schoenfeld residuals plotted against time in weeks for each covariate of the multivariable Cox proportional hazards model. The solid line is a local polynomial regression line.

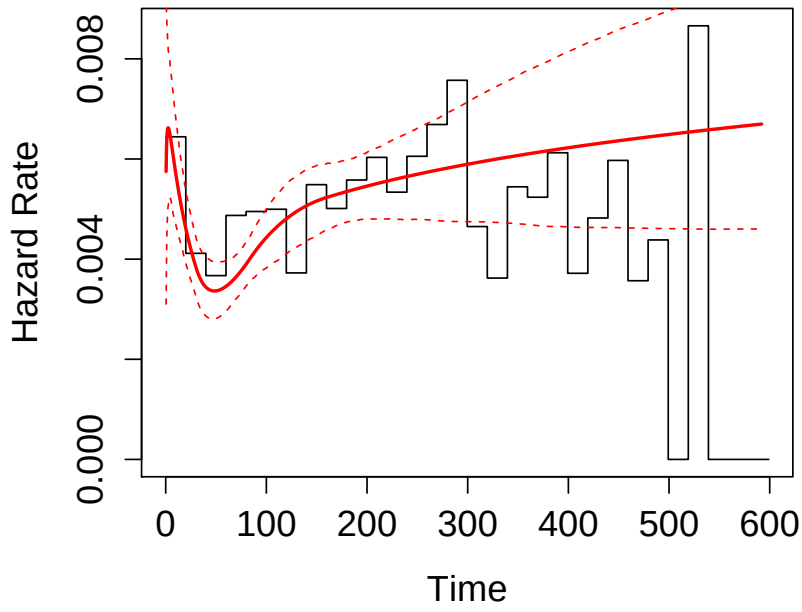

Figure 5.7: The estimated hazard for the Weibull RP model with 3 knots plotted over a non-parametric estimator of the hazard.

using 0 knots. The reason for the latter two being as similar as they are to the Cox proportional hazards model might be that when modeling with splines we capture the shape of the data to a greater extent, which would be similar to the way the Cox model captures the shape of the data perfectly, in a sense, since it technically uses a non-parametric estimator for the baseline hazard. Note that the RP model with 0 knots is just a Weibull model with covariates.

It is difficult to talk about causality when dealing with registry data and its inherent selection bias. Nevertheless, we see in Table 5.3 for the Weibull RP model with 3 knots, that patients receiving new therapies have approximately half the hazard rate compared to patients receiving old therapies, holding all other variables fixed. The age at diagnosis seems to have the impact of increasing the hazard with 3% per additional year of age. For the variable stage of MM there is an increase of hazard for each step in the stage. Patients with missing staging have 114% greater hazard compared to stage 1. This falls between stage 2 and stage 3, which has an increase of 72% and 239%, respectively, when keeping all other variables fixed.

In Table 5.3 the log-logistic RP model can be seen as well. Comparison with the coefficients of the Weibull RP models is not trivial since they are on different scales, that is, proportional hazards and proportional odds. In figures 5.8 and 5.9 Martingale residuals, as explained in Section 2.10.1, are plotted against covariate values for the Weibull RP and log-logistic RP models, respectively. The two figures are quite similar and there does not seem to be any reason to suspect that the assumptions are violated. That is, the residuals are evenly distributed around 0 and there does not seem to be any systematic differences for different covariate values. Here again, the solid line is a local polynomial regression line.

Table 5.3: Estimated coefficients for the Weibull RP models with 0, 3 and 6 knots, respectively, and the log-logistic RP model with 2 knots for overall survival. The columns p-val indicate the Wald p-values of the test $H_0 : \beta = 0$.

	Weibull RP									Log-logistic RP		
	0 Knots			3 Knots			6 Knots			2 Knots		
Variable	$\hat{\beta}$	$\exp(\hat{\beta})$	p-val	$\hat{\beta}$	$\exp(\hat{\beta})$	p-val	$\hat{\beta}$	$\exp(\hat{\beta})$	p-val	$\hat{\beta}$	$\exp(\hat{\beta})$	p-val
New therapy	-0.67	0.51	$2 \cdot 10^{-9}$	-0.63	0.53	$2 \cdot 10^{-8}$	-0.63	0.53	$1 \cdot 10^{-8}$	-0.79	0.45	$1 \cdot 10^{-7}$
Age	0.03	1.03	$2 \cdot 10^{-9}$	0.03	1.03	$1 \cdot 10^{-10}$	0.03	1.03	$1 \cdot 10^{-10}$	0.03	1.03	$3 \cdot 10^{-6}$
Stage 2	0.53	1.69	$8 \cdot 10^{-4}$	0.55	1.73	$5 \cdot 10^{-4}$	0.55	1.73	$5 \cdot 10^{-4}$	0.82	2.28	$2 \cdot 10^{-4}$
Stage 3	1.18	3.25	$9 \cdot 10^{-14}$	1.22	3.39	$1 \cdot 10^{-14}$	1.22	3.39	$1 \cdot 10^{-14}$	1.89	6.59	$4 \cdot 10^{-17}$
Stage missing	0.74	2.1	$8 \cdot 10^{-7}$	0.76	2.14	$4 \cdot 10^{-7}$	0.76	2.14	$4 \cdot 10^{-7}$	1.17	3.22	$3 \cdot 10^{-8}$

5.2 Additional Data

In this section we will take a look at the dataset `lung` in the package `survival` in `R`. It contains survival data for patients with advanced lung cancer from the North Central Cancer Treatment Group (Loprinzi et. al., 1994). The median survival, estimated from the Kaplan-Meier curve, is 310 days. In Figure 5.10 different follow-up times have been imposed and then the minimum AIC and BIC models have been fitted to the data. In this case it seems like once we reach a follow-up close to the median survival time, we start to get accurately fitted models. Here, a follow-up of 312 days seems to be the shortest follow-up we need for accurate model fitting.

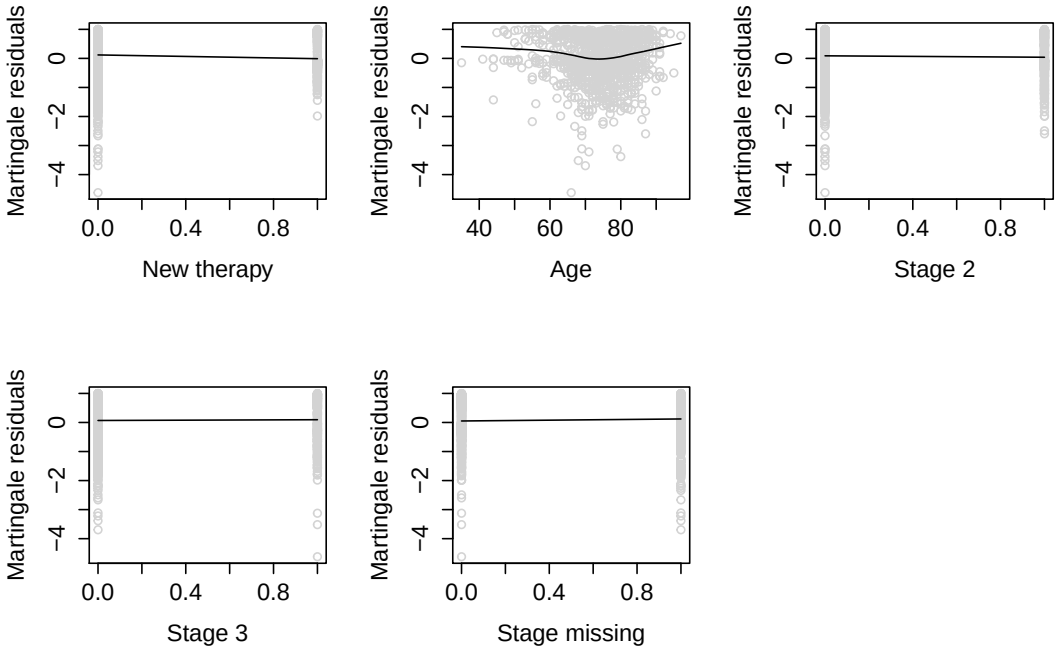

Figure 5.8: Martingale residuals plotted against covariate values for the Weibull RP model with 3 knots. For new therapy, stage 2, stage 3 and stage missing a value of 1 indicated "yes". The solid line is a local polynomial regression line.

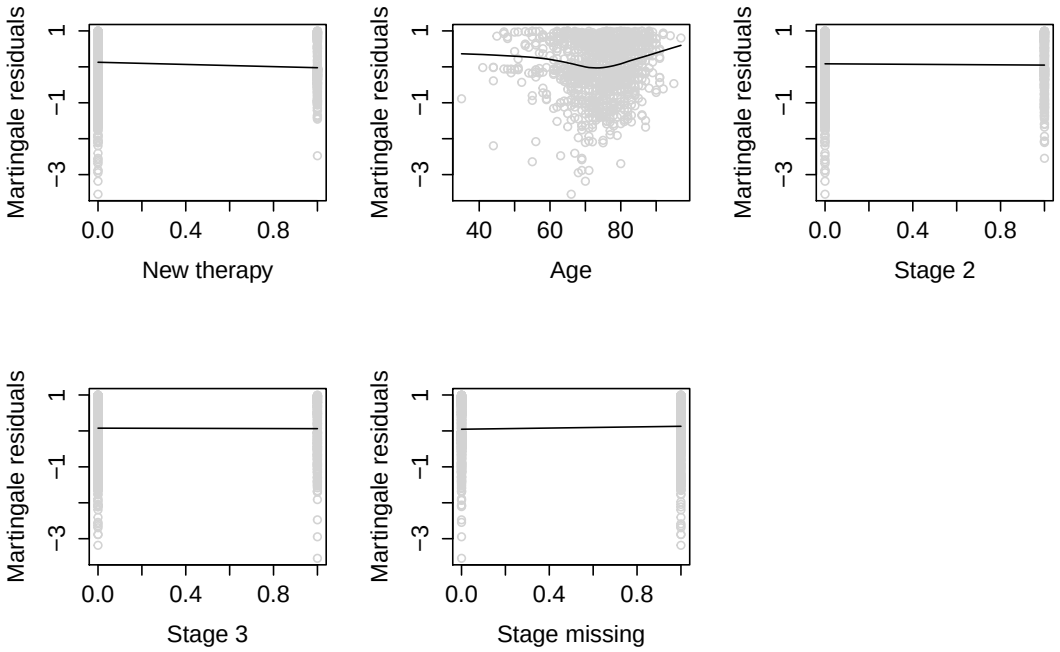

Figure 5.9: Martingale residuals plotted against covariate values for the log-logistic RP model with 2 knots. For new therapy, stage 2, stage 3 and stage missing a value of 1 indicated "yes". The solid line is a local polynomial regression line.

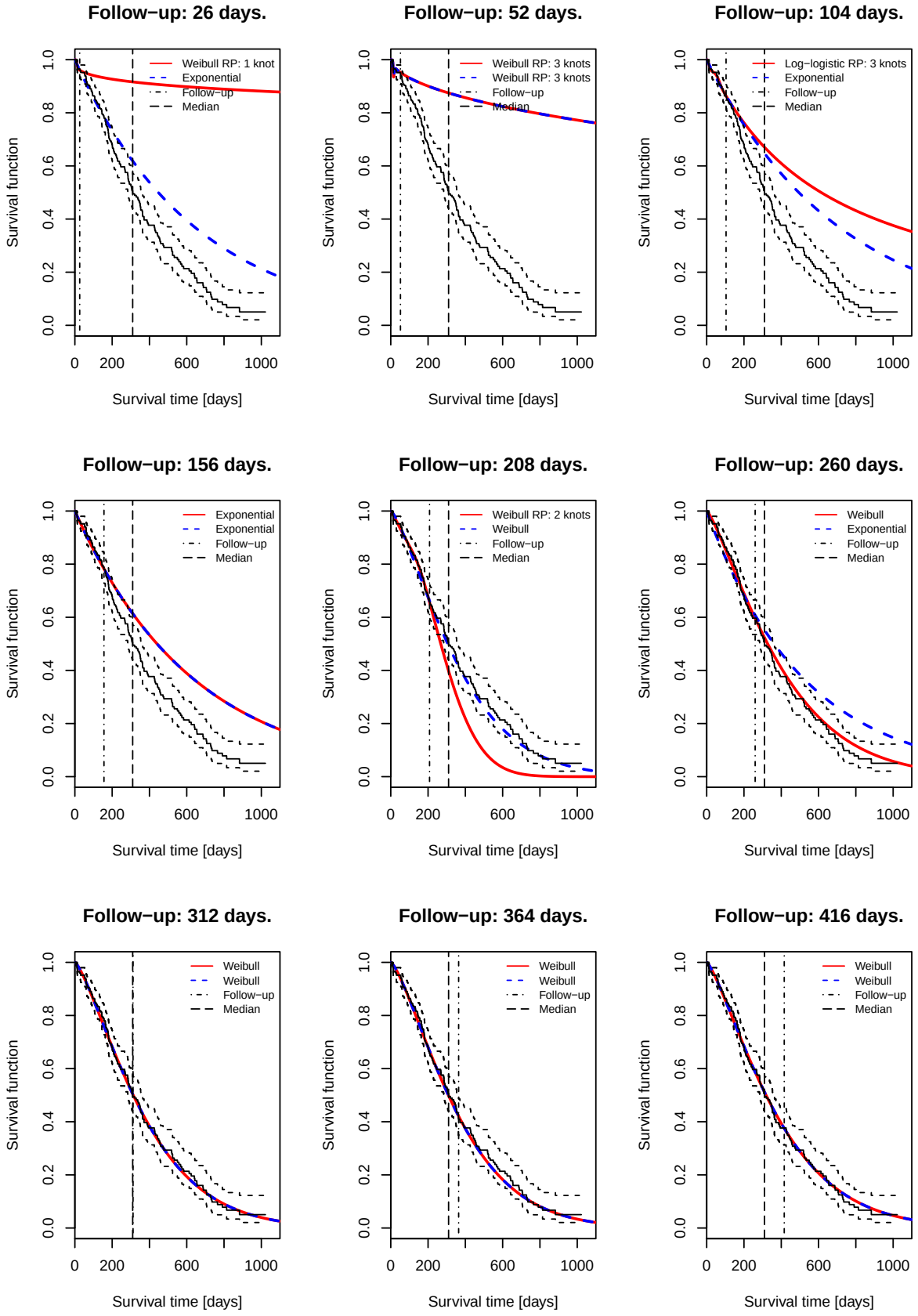

Figure 5.10: Overall survival for advanced lung cancer patients from the North Central Cancer Treatment Group for different imposed follow-up times. The median survival time is 310 days. Models are chosen by minimum AIC (solid line) and minimum BIC (dashed line) and plotted together with the Kaplan-Meier curve for the full data.

Chapter 6

Summary and Discussion

In this thesis we have delved into some of the theory of survival analysis. We looked at a promising but uncommonly used method of adding flexibility to modeling survival curves, namely the Royston-Parmar models. The aim of this thesis was to examine the performance of some of the standard parametric models and the Royston-Parmar models in terms of capturing the true mean time-to-event of the data generating distribution. Furthermore we wanted to explore ways of modeling with explanatory variables.

At our disposal we had a Swedish dataset containing registry data of 1606 multiple myeloma patients between January, 2000 and November, 2011. We analysed a subset of 1098 patients who did not receive high-dose chemotherapy followed by a stem cell transplant. We fitted Cox proportional hazards models and Royston-Parmar models to the data with overall survival as response. As explanatory variables we had old vs new therapies, age at diagnosis and stage of multiple myeloma. The best model according to BIC, among the Royston-Parmar models, was a Weibull Royston-Parmar model with 3 knots, while according to AIC it was a Weibull Royston-Parmar model with 6 knots. We observed that the Weibull Royston-Parmar models with 3 and 6 knots gave almost identical coefficient estimates as the Cox proportional hazards model. This can probably be explained by the increased flexibility of the Royston-Parmar models, which mimic, in a sense, non-parametric models. All coefficients in both the Cox model and the Royston-Parmar models were significantly different from 0 at all standard significance levels. Furthermore, the proportional hazards assumption seemed to hold, see Section 5.1.1 for the coefficient estimates.

A simulation study was performed to investigate the performance of standard parametric models and Royston-Parmar models in terms of capturing the true mean time-to-event of the data generating distribution. In addition we investigated how well choosing models, blindly, by minimal AIC or BIC performs. Here we saw that standard parametric models are very sensitive to misspecification. This is also true for the Royston-Parmar models, but with sufficient follow-up these are flexible enough to not be as severely biased as the standard parametric models. Also, when the true model does not lie within the space of models fitted to the data, it seems as if the Royston-Parmar models are a good choice. Choosing model by minimum AIC or BIC performs quite well, especially when the true model is one of the fitted models. We found that

a rule-of-thumb for the amount of follow-up needed could be that once we have follow-up above the median, the bias of the mean estimate is close to zero. Furthermore, the confidence intervals start to be of reasonable width. This should be used carefully since this is but one simulation study with a certain set of conditions. We investigated this rule-of-thumb on the datasets we had at hand and it looked to work quite well. That is, most curves fitted with follow-up above the median seemed to fit within the 95% pointwise confidence intervals of the Kaplan-Meier curves. While with follow-ups shorter than the median, this did not seem to be the case. One should be aware though that both datasets this was tested on had shapes of the survival curve that went smoothly towards zero. In the case of diseases with cure, for instance, this rule-of-thumb would most likely not hold. However, in that case fitting regular parametric distributions to the data might be difficult in and of itself. One should probably look at modeling curative treatment or treatments with a profoundly different time-to-event using mixture modeling and relative survival, where you take into account the general populations hazard and model the excessive hazard from the disease.

One problem with the simulation study is that to be able to perform a large amount of simulations, we cannot use visual methods to ascertain the validity of a model. Thus it might be the case that choosing a model based on minimal AIC together with visual inspection and reality checks would perform better. The simulation study is quite limited. This is since running the simulations was time consuming. The simulations presented in this thesis took approximately 16 hours to run. The simulation study thus only considers three data generating distributions and only a small number of different follow-up times, which are spaced by quite some distances. Furthermore, since the sample size is dependent on the length of follow-up it is difficult to determine the effect of having a longer follow-up for a fixed sample size, and the other way around. One might want to set up a grid of follow-up times and a grid of sample sizes and test each combination, as done in the simple simulation example in subsection 4.2.2, but with denser grids. With a better computer one might increase the number of models, both simulated from and fitted, as well as have more tightly spaced follow-up times. This to better be able to build an image of how the different parameters of the setup affect our ability to find the true mean.

Often one does not only have baseline characteristics at diagnosis. Certain lab values might be taken at more than one point in time. Thus using some forms of longitudinal methods might help the analysis of such data.

In our analyses we have used a few crude explanatory variables since the purpose has not been to do deep analyses of the data. Furthermore, we do not know anything about the causality of the different explanatory variables. A next step would be to fit models with more of the available variables, in collaboration with experts within multiple myeloma oncology. Since a lot of the variables have a non-negligible amount of missing values, one should also look into methods of imputation or something similar.

In Section 2.11 we mentioned some survival analysis methods that were not used in the simulation study or the data analysis. It would be interesting to investigate how mixture models and models from a Bayesian model averaging context perform as well as investigating discrete time methods,

such as the logistic regression model mentioned in subsection 2.11.4.

In conclusion, we have compared the performance of standard parametric models with Royston-Parmar models through a simulation study as well as through real world observational data. We found that standard parametric models are very sensitive to misspecification. This is also true for the Royston-Parmar models, but with sufficient follow-up these are flexible enough to not be as severely biased as the standard parametric models. Thus it would make sense for health economists to add the RP models to their repertoire of models.

Appendix A

Supplementary Theory

In this appendix we go through some parametric distributions as well as some mathematical methods and results used in the thesis. In Section A.1 a short description of each of the parametric distributions considered in this thesis is given. In Section A.2 we describe the delta method, in Section A.3 we go through how numerical approximations of the gradient are found and in Section A.4 we explain rejection sampling.

A.1 Parametric Distributions

Exponential Distribution The exponential distribution is the only distribution with a constant hazard function over time. This means that the probability of an event in a given time interval is independent of how much time has passed up until that interval. This is called the memoryless property.

Weibull Distribution The Weibull distribution is a generalization of the exponential distribution which lets the hazard depend on time through a power function (Kalbfleisch & Prentice, 2011). It includes the exponential and the Rayleigh distributions as special cases and it is itself a special case of the generalized extreme value distribution (Johnson, Kotz & Balakrishnan, 1994).

The hazard function is a monotonically decreasing function of time when $a < 1$ and a monotonically increasing function of time when $a > 1$. If $a = 1$ the Weibull distribution simplifies to the exponential distribution, where $\beta = 1/b$, and if $a = 2$ and $b = \sqrt{2}\sigma$ it simplifies to the Rayleigh distribution with scale parameter σ.

Fréchet Distribution Another special case of the generalized extreme value distribution is the Fréchet distribution, also called the inverse Weibull distribution (Khan, Pasha & Pasha, 2008). The hazard function increases from 0 at $t = 0$ to a maximum and then it decreases towards 0 as $t \to \infty$.

Gompertz Distribution There are three special cases of the generalized extreme value distribution. The third and final is the Gumbel distribution. However, it has support on the whole real line and thus is not applicable to our problem. Instead we turn to the Gompertz distribution, which is a zero-truncated Gumbel distribution (Bauckhage, 2014).

The Gompertz distribution has a monotonically increasing hazard function. The PDF shown in Table 2.1 is of the same form as can be found in the package `eha` in `R`.

Log-Normal Distribution The log-normal distribution has a non-monotonic hazard function, starting at 0 for $t = 0$, increasing to a maximum and then decreasing towards 0 as $t \to \infty$.

Log-Logistic Distribution Another distribution with a non-monotonic hazard function, similar to the log-normal distribution but with heavier tails (Kalbfleisch & Prentice, 2011), is the log-logistic distribution.

The scale parameter b is the median of the distribution. For $a > 1$ the hazard function starts at 0 for $t = 0$, increases to a maximum at $t = b(a-1)^a$ and after that decreases to 0 as $t \to \infty$ (Kalbfleisch & Prentice, 2011). For $a \leq 1$ the hazard is monotonically decreasing.

Due to the heavier tails one should check that the tails of the distribution yield reasonable survival estimates.

Gamma Distribution If $a > 1$ then the hazard function is monotonically increasing from 0 and if $a < 1$ it is monotonically decreasing from ∞ (Kalbfleisch & Prentice, 2011). In both cases it approaches $1/s$ as $t \to \infty$. If $a = 1$ then the gamma distribution simplifies to the exponential distribution and if a is a positive integer it simplifies to the Erlang distribution (Johnson, Kotz & Balakrishnan, 1994).

Generalized Gamma Distribution The generalized gamma distribution of Stacy (1962) generalizes the 2-parameter gamma distribution to a 3-parameter distribution. In this thesis we will use the parametrization of Prentice (1974) which is implemented in the `R`-package `flexsurv`. The reason for using this parametrization is that in the original parametrization, the log-normal distribution is a limiting special case ($k \to \infty$), while in this one it is simply a special case ($q = 0$).

We omit the survival and hazard functions here since neither is very informative as to the shape and behavior of the distribution.

The generalized gamma distribution in this parametrization simplifies to the gamma distribution if $q = \sigma$, to the Weibull distribution if $q = 1$, the exponential distribution if $q = \sigma = 1$ and the case $q = 0$ is the log-normal distribution (Cox, Chu, Schneider & Muñoz, 2007).

Generalized F Distribution The generalized F distribution, in the parametrization of Prentice (1975), has the location parameter μ, scale parameter $\sigma > 0$ and shape parameters q and $p > 0$. This is the parametrization implemented in the `R`-package `flexsurv`. The density is not very illuminating and is thus left out. Nonetheless, the reason for including the generalized F distribution is that it contains the generalized gamma distribution as a limiting special case when $p = 0$ and furthermore it contains the log-logistic distribution as a special case when $q = 0$ and $p = 1$ (Cox, 2008). Thus we have an umbrella distribution that contains most of the considered distributions in this thesis.

A.2 Delta Method

Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimator of the parameter vector $\boldsymbol{\theta}$. We know then that $\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{D} N(0, I^{-1}(\boldsymbol{\theta}))$, (Miller, Gong & Muñoz, 1981). We can use this to get an asymptotic distribution for the mean of a distribution by applying the delta method. Now we state the delta method (Hunter, 2006). If $\boldsymbol{g} : R^k \to R^l$ is continuously differentiable at $\boldsymbol{a} \in R^k$, that is, the jacobian $\boldsymbol{J} : R^k \to R^l$ exists, and

$$\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{a}) \xrightarrow{D} N(0, \Sigma),$$

then

$$\sqrt{n}(\boldsymbol{g}(\boldsymbol{X}_n) - \boldsymbol{g}(\boldsymbol{a})) \xrightarrow{D} N(0, J^T \Sigma J).$$

A.3 Numerically Finding the Gradient

If there is no closed form equation for the gradient, then a numerical approximation is carried out by Richardson extrapolation (Richardson & Gaunt, 1927).

The method used is the method used in the `R` function `jacobian` of the package `numDeriv`. The approach is to approximate the derivative with respect to x_i as

$$f_i'(x) \approx \frac{f(x_1, ..., x_i + d, ..., x_n) - f(x_1, ..., x_i + d, ..., x_n)}{2d},$$

iterating for smaller and smaller d and then using Richardson extrapolation.

The following is a brief description of Richardson extrapolation. We want to approximate some constant A with some approximation $A(h)$ depending on some small step size h such that

$$A(h) = A + ch^n + o(h^{n+1}),$$

for some constant c. The Richardson extrapolation of $A(h)$ is defined as

$$R(h, k) = \frac{k^n A(h) - A(kh)}{k^n - 1}.$$

This approximation has a higher-order error term in comparison to $A(h)$ which is seen by expanding $A(h)$ and $A(kh)$,

$$\begin{aligned} R(h,k) &= \frac{k^n\left(A + ch^n + o(h^{n+1})\right) - \left(a + c(kh)^n + o(h^{n+1})\right)}{k^n - 1} \\ &= A + o(h^{n+1}). \end{aligned}$$

A.4 Rejection Sampling

In the package `flexsurv` in `R` there is a function `rsurvspline` which simulates observations from an RP model. It does this by using inverse transform sampling. That is, it samples n uniform random numbers between 0 and 1 and then inserts these into the quantile function, or inverse cumulative distribution function, to get a sample from the RP models. The good thing about this is that it is possible to sample from any RP model without putting to much thought into it. The bad thing is that it is very slow, this since the quantile function `qsurvspline` uses numerical root-finding. Thus for our simulation study we need another approach since this would take too much time.

What we instead use is rejection sampling (Casella, Robert & Wells, 2004). Say we want to sample from some distribution f. To use rejection sampling we find a distribution g from which we can simulate observations and for which we can find a lower bound of f/g. Say that this lower bound is $1/\epsilon$. One step of the rejection sampling algorithm is then

1. Simulate $X \sim g$ and $U \sim U(0,1)$, independently.
2. If $U \leq \epsilon\frac{f(X)}{g(X)}$, accept $X \sim f$,
3. else, reject X and return to step 1.

Iterating n times yields a sample $\boldsymbol{X}$ from f of size n.

A.5 Kernel Density Estimation

In `R` we can do kernel density estimation with the function `density` in the package `stats`. Kernel density estimation is a non-parametric method to estimate the probability density function of random variables. The idea is to use

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),$$

where $K(\cdot)$ is the kernel and $h > 0$ is the bandwidth. The kernel is a non-negative real-valued integrable function that integrates to 1 over its support and which satisfies $K(-x) = K(x)$ for all values of x. The bandwidth has to be selected and this is a whole subject in an of itself. In this thesis we will keep it simple and use the standard normal density function as kernel function

and we will choose the bandwidth based on the rule

$$h = 1.06n^{-1/5} \cdot \min\left\{\hat{\sigma}, IQR/1.34\right\},$$

where n is the number of observations, $\hat{\sigma}$ is the estimated standard deviation of the sample and IQR is the interquartile range. This rule is derived by minimizing the mean integrated square error (MISE),

$$MISE(f, \hat{f}_h) = \mathrm{E}\left[\int (f(x) - \hat{f}_h(x))^2 dx\right],$$

under the assumption of the data being normal. See Hollander, Chicken & Wolfe (2013) for the above and more on kernel density estimation.

Appendix B

Simulation Results

In this appendix the summary tables of the simulation study are shown. Below we explain the columns of the tables.

- True: Model simulated from.
- Fitted: Model fitted to the simulated data.
- μ_0: Mean of the true model.
- $\lambda_{50\%}$: Median of the true model.
- $B(\hat{\mu})_{\text{Mean}}$: Mean bias of the mean estimates.
- $B(\hat{\mu})_{50\%}$: Median bias of the mean estimates.
- $B(\hat{\mu})_{2.5\%}$: 2.5% bias of the mean estimates.
- $B(\hat{\mu})_{97.5\%}$: 97.5% bias of the mean estimates.
- CI Cov.: Coverage of 95% confidence intervals of the mean estimates.
- Mean len.: Mean length of the 95% confidence intervals.
- Med. len.: Median length of the 95% confidence intervals.
- $B(\hat{\lambda}_{25\%})$: Mean bias in the estimates of the 25% quantile of the true distribution.
- $B(\hat{\lambda}_{50\%})$: Mean bias in the estimates of the 50% quantile of the true distribution.
- $B(\hat{\lambda}_{75\%})$: Mean bias in the estimates of the 75% quantile of the true distribution.
- n: Mean number of observations simulated.
- p: Mean proportion of the simulated observations that are not censored.

Table B.1: Summary table of the simulation study with (see text for an explanation of the columns): Followup: 26 weeks. Rate of new patients: 6 per week.

True	Fitted	μ_0	$\lambda_{50\%}$	$B(\hat{\mu})_{\text{Mean}}$	$B(\hat{\mu})_{50\%}$	$B(\hat{\mu})_{2.5\%}$	$B(\hat{\mu})_{97.5\%}$	CI Cov.	Mean len.	Med. len.	$B(\hat{\lambda}_{25\%})$	$B(\hat{\lambda}_{50\%})$	$B(\hat{\lambda}_{75\%})$	n	p
tlnorm	lnorm	109.0	45.5	-39.4	-3.3	-325.4	56.2	0.91	403.0	224.8	0.1	1.2	4.1	156	0.17
tlnorm	tlnorm	109.0	45.5	-5.2	4.8	-118.9	56.2	0.90	175.9	164.5	0.1	1.3	4.6	156	0.17
tlnorm	exp	109.0	45.5	40.8	43.1	8.2	61.7	0.21	52.9	49.6	-1.1	-0.2	23.8	156	0.17
tlnorm	Weibull	109.0	45.5	63.7	66.6	32.1	79.3	0.05	39.5	33.3	-0.0	10.6	56.2	156	0.17
tlnorm	gamma	109.0	45.5	61.0	63.5	31.1	76.7	0.07	47.3	40.9	0.1	10.1	52.8	156	0.17
tlnorm	RP	109.0	45.5	0.8	48.6	-360.2	77.5	0.57	488.7	129.0	-0.1	3.2	31.8	156	0.17
tlnorm	min.AIC	109.0	45.5	-24.4	20.3	-354.5	75.7	0.68	467.8	141.2	0.0	2.4	15.4	156	0.17
tlnorm	min.BIC	109.0	45.5	20.9	39.7	-16.8	75.4	0.41	194.0	60.3	-0.4	2.5	25.3	156	0.17
exp	lnorm	123.4	85.6	-68668965.4	-1392.2	-483783.1	-18.4	1.00	3140637041.9	8869.5	-7.8	-89.3	-537.9	156	0.10
exp	tlnorm	123.4	85.6	-234.9	-230.7	-490.7	-7.0	0.98	26023003365.3	534.8	-6.8	-69.4	-316.5	156	0.10
exp	exp	123.4	85.6	-8.7	-0.7	-106.7	43.7	0.94	141.3	124.5	-0.2	-0.5	-1.0	156	0.10
exp	Weibull	123.4	85.6	-67.4	7.4	-593.5	74.7	0.78	857.3	193.3	0.9	4.0	10.4	156	0.10
exp	gamma	123.4	85.6	-32.9	6.2	-356.2	67.7	0.87	437.0	224.9	0.8	3.5	8.9	156	0.10
exp	RP	123.4	85.6	-330.0	-20.9	-3528.2	77.3	0.88	2487.1	540.4	-1.4	-8.5	-26.8	156	0.10
exp	min.AIC	123.4	85.6	-22618420.8	-10.8	-28956.9	70.0	0.90	1285419333.5	143.1	-0.6	-3.2	-11.7	156	0.10
exp	min.BIC	123.4	85.6	-469037.1	-0.8	-144.2	47.0	0.93	13997048.5	125.7	-0.2	-0.4	-0.8	156	0.10
RP	lnorm	55.4	32.5	-60.3	-35.6	-265.1	12.0	0.99	256.9	157.0	0.1	-3.6	-20.7	156	0.21
RP	tlnorm	55.4	32.5	-39.2	-30.8	-133.5	12.1	0.95	132.5	123.3	0.1	-3.6	-20.5	156	0.21
RP	exp	55.4	32.5	1.8	3.1	-20.1	16.7	0.91	37.3	35.6	-0.4	-3.7	-3.2	156	0.21
RP	Weibull	55.4	32.5	17.6	19.4	-4.1	29.4	0.30	28.4	24.9	-0.3	3.2	19.4	156	0.21
RP	gamma	55.4	32.5	15.4	16.9	-5.1	27.5	0.44	33.0	29.6	-0.2	2.8	16.5	156	0.21
RP	RP	55.4	32.5	2.1	13.4	-83.5	28.8	0.70	135.7	58.0	-0.1	1.0	11.1	156	0.21
RP	min.AIC	55.4	32.5	-14.3	-0.4	-133.1	27.7	0.72	132.5	48.3	-0.1	-0.1	-0.5	156	0.21
RP	min.BIC	55.4	32.5	-0.3	2.3	-34.4	27.4	0.82	62.0	38.6	-0.2	-1.4	-1.0	156	0.21

Table B.2: Summary table of the simulation study with (see text for an explanation of the columns): Followup: 52 weeks. Rate of new patients: 6 per week.

True	Fitted	μ_0	$\lambda_{50\%}$	$B(\hat{\mu})_{\text{Mean}}$	$B(\hat{\mu})_{50\%}$	$B(\hat{\mu})_{2.5\%}$	$B(\hat{\mu})_{97.5\%}$	CI Cov.	Mean len.	Med. len.	$B(\hat{\lambda}_{25\%})$	$B(\hat{\lambda}_{50\%})$	$B(\hat{\lambda}_{75\%})$	n	p
tlnorm	lnorm	109.0	45.5	-13.9	-7.9	-96.1	32.4	0.96	125.0	112.4	-0.0	0.2	0.7	312	0.31
tlnorm	tlnorm	109.0	45.5	-0.9	1.2	-47.9	33.6	0.93	81.2	80.1	-0.0	0.2	1.2	312	0.31
tlnorm	exp	109.0	45.5	43.7	44.2	29.4	55.5	0.00	26.1	25.7	-0.8	0.5	25.3	312	0.31
tlnorm	Weibull	109.0	45.5	53.1	53.8	36.8	64.4	0.00	24.2	23.2	-1.6	3.4	38.3	312	0.31
tlnorm	gamma	109.0	45.5	52.3	52.9	37.4	63.2	0.00	25.7	24.9	-1.6	3.7	37.7	312	0.31
tlnorm	RP	109.0	45.5	27.5	34.4	-45.4	58.3	0.54	101.8	75.6	0.2	0.0	14.8	312	0.31
tlnorm	min.AIC	109.0	45.5	-2.8	0.3	-89.5	58.4	0.84	112.8	90.1	-0.1	0.3	2.5	312	0.31
tlnorm	min.BIC	109.0	45.5	6.2	9.6	-79.1	57.1	0.70	92.4	78.8	-0.3	0.8	10.5	312	0.31
exp	lnorm	123.4	85.6	-1176.5	-694.5	-4843.3	-137.1	1.00	4120.7	1922.2	-1.3	-46.7	-308.2	312	0.18
exp	tlnorm	123.4	85.6	-186.4	-183.9	-313.2	-74.9	0.05	260.3	218.3	-1.5	-41.3	-226.0	312	0.18
exp	exp	123.4	85.6	-1.6	0.2	-40.4	26.9	0.95	65.6	63.9	0.1	0.1	0.3	312	0.18
exp	Weibull	123.4	85.6	-5.5	2.1	-97.8	42.8	0.87	108.4	91.8	0.1	1.3	3.0	312	0.18
exp	gamma	123.4	85.6	-3.0	2.1	-72.2	37.8	0.92	107.8	97.7	0.1	1.2	3.0	312	0.18
exp	RP	123.4	85.6	-24.5	-3.2	-232.3	46.4	0.92	259.5	169.9	0.1	-0.8	-3.9	312	0.18
exp	min.AIC	123.4	85.6	-101.6	-3.9	-1041.1	41.0	0.85	378.6	67.8	0.1	-1.9	-5.4	312	0.18
exp	min.BIC	123.4	85.6	-25.1	-0.3	-75.4	28.8	0.94	152.8	64.3	0.1	-0.1	-0.4	312	0.18
RP	lnorm	55.4	32.5	-25.9	-22.6	-70.8	0.6	0.85	64.1	58.5	0.6	-1.0	-11.2	312	0.38
RP	tlnorm	55.4	32.5	-22.1	-20.3	-54.9	0.9	0.72	51.4	50.0	0.6	-1.0	-11.1	312	0.38
RP	exp	55.4	32.5	5.8	6.1	-3.8	13.7	0.69	17.8	17.7	0.4	-1.6	1.0	312	0.38
RP	Weibull	55.4	32.5	13.0	13.4	3.6	20.1	0.15	15.0	14.6	-0.9	-0.1	10.7	312	0.38
RP	gamma	55.4	32.5	12.1	12.4	3.0	19.2	0.21	15.9	15.6	-0.8	0.1	10.1	312	0.38
RP	RP	55.4	32.5	6.7	8.2	-13.7	18.3	0.70	32.5	27.9	-0.0	-0.3	4.4	312	0.38
RP	min.AIC	55.4	32.5	-3.7	2.5	-45.0	18.5	0.52	34.0	26.0	-0.1	-0.3	2.0	312	0.38
RP	min.BIC	55.4	32.5	-2.1	2.9	-38.8	18.4	0.63	30.6	19.5	0.1	-0.6	0.2	312	0.38

Table B.3: Summary table of the simulation study with (see text for an explanation of the columns): Followup: 104 weeks. Rate of new patients: 6 per week.

True	Fitted	μ_0	$\lambda_{50\%}$	$B(\hat{\mu})_{\text{Mean}}$	$B(\hat{\mu})_{50\%}$	$B(\hat{\mu})_{2.5\%}$	$B(\hat{\mu})_{97.5\%}$	CI Cov.	Mean len.	Med. len.	$B(\hat{\lambda}_{25\%})$	$B(\hat{\lambda}_{50\%})$	$B(\hat{\lambda}_{75\%})$	n	p
tlnorm	lnorm	109.0	45.5	-10.2	-8.4	-46.2	15.8	0.96	60.8	59.1	-0.0	0.1	0.2	624	0.48
tlnorm	tlnorm	109.0	45.5	-0.1	0.6	-23.4	19.3	0.94	42.3	42.1	-0.0	0.1	0.5	624	0.48
tlnorm	exp	109.0	45.5	40.0	40.2	31.3	47.6	0.00	15.7	15.6	-1.9	-2.2	19.9	624	0.48
tlnorm	Weibull	109.0	45.5	40.8	41.1	30.1	49.5	0.00	16.2	16.0	-2.2	-2.2	21.0	624	0.48
tlnorm	gamma	109.0	45.5	41.9	42.1	32.2	50.0	0.00	16.6	16.4	-2.5	-2.0	22.6	624	0.48
tlnorm	RP	109.0	45.5	18.1	20.2	-14.4	38.3	0.58	53.0	48.5	0.1	0.1	2.7	624	0.48
tlnorm	min.AIC	109.0	45.5	-2.5	-1.9	-40.5	34.2	0.90	53.5	47.9	0.0	0.1	0.4	624	0.48
tlnorm	min.BIC	109.0	45.5	-4.2	-3.0	-41.3	33.3	0.93	51.5	47.6	-0.0	0.1	0.8	624	0.48
exp	lnorm	123.4	85.6	-349.0	-314.6	-745.8	-142.3	0.00	467.9	410.7	3.3	-16.2	-150.9	624	0.32
exp	tlnorm	123.4	85.6	-131.0	-129.4	-189.2	-79.2	0.00	96.4	96.0	3.0	-16.3	-132.1	624	0.32
exp	exp	123.4	85.6	-0.5	-0.1	-18.6	15.3	0.95	34.3	34.1	-0.0	-0.1	-0.2	624	0.32
exp	Weibull	123.4	85.6	-1.1	0.4	-31.5	21.2	0.91	43.5	42.0	-0.1	0.1	0.6	624	0.32
exp	gamma	123.4	85.6	-0.7	0.4	-25.7	19.1	0.95	44.3	43.3	-0.1	0.1	0.6	624	0.32
exp	RP	123.4	85.6	-3.4	-0.6	-45.9	22.8	0.95	67.6	62.5	0.1	-0.0	-0.6	624	0.32
exp	min.AIC	123.4	85.6	-5.2	-0.3	-106.9	20.1	0.88	39.7	34.8	0.0	-0.2	-0.4	624	0.32
exp	min.BIC	123.4	85.6	-2.0	-0.2	-22.8	16.1	0.93	35.5	34.1	-0.0	-0.1	-0.3	624	0.32
RP	lnorm	55.4	32.5	-15.5	-14.8	-31.5	-3.3	0.36	27.0	26.4	0.9	0.5	-5.2	624	0.58
RP	tlnorm	55.4	32.5	-14.0	-13.5	-27.6	-2.9	0.35	23.9	23.7	0.9	0.5	-5.2	624	0.58
RP	exp	55.4	32.5	5.7	5.7	0.3	10.5	0.41	10.3	10.2	0.4	-1.9	0.5	624	0.58
RP	Weibull	55.4	32.5	7.9	8.0	2.3	12.8	0.14	8.9	8.9	-0.9	-2.3	3.3	624	0.58
RP	gamma	55.4	32.5	8.0	8.1	2.6	12.7	0.15	9.6	9.5	-1.1	-2.1	3.9	624	0.58
RP	RP	55.4	32.5	2.8	3.2	-7.0	10.0	0.82	16.5	15.8	0.1	-0.3	0.0	624	0.58
RP	min.AIC	55.4	32.5	0.7	2.8	-19.4	10.2	0.71	16.9	16.0	0.2	-0.3	-0.2	624	0.58
RP	min.BIC	55.4	32.5	-3.2	-0.2	-23.5	11.3	0.43	17.4	17.5	0.3	-0.4	-0.9	624	0.58

Table B.4: Summary table of the simulation study with (see text for an explanation of the columns): Followup: 156 weeks. Rate of new patients: 6 per week.

True	Fitted	μ_0	$\lambda_{50\%}$	$B(\hat{\mu})_{\text{Mean}}$	$B(\hat{\mu})_{50\%}$	$B(\hat{\mu})_{2.5\%}$	$B(\hat{\mu})_{97.5\%}$	CI Cov.	Mean len.	Med. len.	$B(\hat{\lambda}_{25\%})$	$B(\hat{\lambda}_{50\%})$	$B(\hat{\lambda}_{75\%})$	n	p
tlnorm	lnorm	109.0	45.5	-9.8	-9.1	-33.4	9.8	0.92	43.0	42.5	-0.0	-0.0	-0.3	936	0.58
tlnorm	tlnorm	109.0	45.5	-0.2	-0.0	-16.3	14.3	0.95	30.5	30.4	-0.0	-0.1	-0.1	936	0.58
tlnorm	exp	109.0	45.5	35.9	36.0	29.2	42.4	0.00	12.3	12.3	-3.1	-5.1	14.0	936	0.58
tlnorm	Weibull	109.0	45.5	33.7	33.8	25.3	41.5	0.00	13.1	13.0	-2.1	-4.7	11.5	936	0.58
tlnorm	gamma	109.0	45.5	35.5	35.6	27.8	42.6	0.00	13.5	13.5	-2.9	-5.1	13.4	936	0.58
tlnorm	RP	109.0	45.5	11.5	12.5	-11.8	28.4	0.69	40.3	38.5	-0.0	0.3	-0.7	936	0.58
tlnorm	min.AIC	109.0	45.5	-3.0	-2.8	-28.4	24.2	0.89	37.8	34.7	-0.0	0.0	-0.2	936	0.58
tlnorm	min.BIC	109.0	45.5	-5.0	-3.9	-29.3	13.4	0.93	37.0	34.3	-0.0	-0.1	-0.2	936	0.58
exp	lnorm	123.4	85.6	-211.2	-203.9	-353.3	-116.4	0.00	191.6	183.1	5.0	-4.5	-94.8	936	0.43
exp	tlnorm	123.4	85.6	-101.5	-101.1	-137.2	-69.2	0.00	61.0	61.0	4.7	-5.4	-89.2	936	0.43
exp	exp	123.4	85.6	-0.3	-0.1	-12.9	11.3	0.95	24.2	24.1	-0.0	-0.0	-0.1	936	0.43
exp	Weibull	123.4	85.6	-0.5	0.0	-17.5	14.1	0.91	27.3	26.9	-0.1	-0.0	0.0	936	0.43
exp	gamma	123.4	85.6	-0.3	-0.0	-15.2	13.0	0.95	28.2	28.0	-0.1	-0.0	-0.0	936	0.43
exp	RP	123.4	85.6	-1.2	-0.3	-22.6	15.1	0.95	37.3	36.0	-0.0	-0.0	-0.3	936	0.43
exp	min.AIC	123.4	85.6	-0.7	-0.1	-18.5	13.4	0.92	26.2	24.4	-0.0	-0.0	-0.2	936	0.43
exp	min.BIC	123.4	85.6	-0.4	-0.1	-13.3	11.7	0.94	24.2	24.1	-0.0	-0.0	-0.1	936	0.43
RP	lnorm	55.4	32.5	-12.3	-11.9	-22.4	-3.9	0.19	18.1	17.9	0.9	1.1	-3.0	935	0.69
RP	tlnorm	55.4	32.5	-11.2	-10.9	-20.3	-3.5	0.20	16.5	16.4	0.9	1.1	-3.0	935	0.69
RP	exp	55.4	32.5	4.7	4.7	0.4	8.7	0.36	7.9	7.8	0.0	-2.6	-0.9	935	0.69
RP	Weibull	55.4	32.5	5.4	5.5	0.9	9.5	0.19	6.7	6.7	-0.7	-3.1	-0.0	935	0.69
RP	gamma	55.4	32.5	5.8	5.8	1.4	9.7	0.19	7.5	7.5	-1.1	-3.2	0.7	935	0.69
RP	RP	55.4	32.5	1.2	1.4	-5.4	6.8	0.89	11.9	11.7	0.1	-0.1	-0.5	935	0.69
RP	min.AIC	55.4	32.5	0.8	1.4	-10.0	6.8	0.86	12.0	11.7	0.1	-0.1	-0.5	935	0.69
RP	min.BIC	55.4	32.5	-0.8	1.0	-15.4	7.0	0.72	12.4	11.8	0.2	-0.0	-0.7	935	0.69

Table B.5: Summary table of the simulation study with (see text for an explanation of the columns): Followup: 208 weeks. Rate of new patients: 6 per week.

True	Fitted	μ_0	$\lambda_{50\%}$	$B(\hat{\mu})_{\text{Mean}}$	$B(\hat{\mu})_{50\%}$	$B(\hat{\mu})_{2.5\%}$	$B(\hat{\mu})_{97.5\%}$	CI Cov.	Mean len.	Med. len.	$B(\hat{\lambda}_{25\%})$	$B(\hat{\lambda}_{50\%})$	$B(\hat{\lambda}_{75\%})$	n	p
tlnorm	lnorm	109.0	45.5	-9.2	-8.8	-27.6	6.6	0.87	34.2	33.8	0.0	0.0	-0.1	1248	0.64
tlnorm	tlnorm	109.0	45.5	-0.1	0.1	-12.7	11.6	0.95	24.5	24.5	-0.0	-0.0	0.0	1248	0.64
tlnorm	exp	109.0	45.5	32.6	32.7	26.5	38.3	0.00	10.6	10.6	-4.1	-7.4	9.4	1248	0.64
tlnorm	Weibull	109.0	45.5	29.1	29.2	21.5	36.0	0.00	11.3	11.2	-1.9	-6.0	5.8	1248	0.64
tlnorm	gamma	109.0	45.5	31.1	31.2	24.2	37.5	0.00	11.8	11.8	-2.9	-7.0	7.3	1248	0.64
tlnorm	RP	109.0	45.5	7.6	8.5	-11.7	22.1	0.77	33.5	32.5	-0.1	0.6	-0.9	1248	0.64
tlnorm	min.AIC	109.0	45.5	-3.1	-2.7	-24.3	17.4	0.88	30.1	28.1	-0.0	0.1	-0.1	1248	0.64
tlnorm	min.BIC	109.0	45.5	-4.5	-3.5	-24.8	10.8	0.91	29.4	27.3	-0.0	0.0	-0.1	1248	0.64
exp	lnorm	123.4	85.6	-152.1	-148.8	-228.0	-95.5	0.00	110.5	107.8	5.7	1.6	-65.0	1248	0.52
exp	tlnorm	123.4	85.6	-83.0	-82.6	-108.4	-59.5	0.00	44.8	44.8	5.5	0.6	-64.0	1248	0.52
exp	exp	123.4	85.6	-0.2	-0.0	-9.9	8.9	0.95	19.1	19.1	-0.0	-0.0	-0.1	1248	0.52
exp	Weibull	123.4	85.6	-0.2	0.0	-12.3	10.5	0.91	19.9	19.8	-0.0	-0.0	0.1	1248	0.52
exp	gamma	123.4	85.6	-0.1	0.0	-11.2	9.8	0.95	21.1	21.0	-0.0	-0.0	0.1	1248	0.52
exp	RP	123.4	85.6	-0.5	-0.1	-14.6	11.1	0.95	25.5	25.0	-0.0	0.0	0.0	1248	0.52
exp	min.AIC	123.4	85.6	-0.3	0.0	-12.4	10.2	0.93	19.9	19.2	-0.0	-0.0	0.0	1248	0.52
exp	min.BIC	123.4	85.6	-0.2	-0.0	-10.1	9.0	0.95	19.1	19.1	-0.0	-0.0	-0.0	1248	0.52
RP	lnorm	55.4	32.5	-10.6	-10.5	-18.0	-4.0	0.11	14.2	14.1	0.9	1.3	-1.9	1248	0.75
RP	tlnorm	55.4	32.5	-9.8	-9.7	-16.5	-3.6	0.12	13.1	13.1	0.9	1.3	-1.9	1248	0.75
RP	exp	55.4	32.5	3.8	3.8	0.3	7.2	0.40	6.6	6.6	-0.2	-3.2	-2.2	1248	0.75
RP	Weibull	55.4	32.5	4.0	4.0	0.3	7.6	0.27	5.6	5.5	-0.5	-3.5	-1.9	1248	0.75
RP	gamma	55.4	32.5	4.4	4.4	0.8	7.9	0.27	6.4	6.4	-1.1	-3.8	-1.3	1248	0.75
RP	RP	55.4	32.5	0.5	0.6	-4.6	5.1	0.93	9.7	9.5	0.1	-0.1	-0.5	1248	0.75
RP	min.AIC	55.4	32.5	0.4	0.6	-5.1	5.1	0.92	9.7	9.5	0.1	-0.1	-0.5	1248	0.75
RP	min.BIC	55.4	32.5	0.0	0.5	-9.9	5.2	0.87	9.8	9.6	0.1	-0.0	-0.5	1248	0.75

Table B.6: Summary table of the simulation study with (see text for an explanation of the columns): Followup: 260 weeks. Rate of new patients: 6 per week.

True	Fitted	μ_0	$\lambda_{50\%}$	$B(\hat{\mu})_{\text{Mean}}$	$B(\hat{\mu})_{50\%}$	$B(\hat{\mu})_{2.5\%}$	$B(\hat{\mu})_{97.5\%}$	CI Cov.	Mean len.	Med. len.	$B(\hat{\lambda}_{25\%})$	$B(\hat{\lambda}_{50\%})$	$B(\hat{\lambda}_{75\%})$	n	p
tlnorm	lnorm	109.0	45.5	-9.2	-9.0	-24.6	4.6	0.81	29.0	28.9	0.0	0.0	-0.2	1560	0.69
tlnorm	tlnorm	109.0	45.5	-0.2	-0.2	-11.0	10.2	0.95	20.9	21.0	-0.0	-0.1	-0.2	1560	0.69
tlnorm	exp	109.0	45.5	29.7	29.8	24.2	35.2	0.00	9.5	9.5	-4.9	-9.5	5.3	1560	0.69
tlnorm	Weibull	109.0	45.5	25.6	25.6	18.6	32.2	0.00	10.1	10.1	-1.7	-6.9	1.5	1560	0.69
tlnorm	gamma	109.0	45.5	27.7	27.7	21.2	33.8	0.00	10.7	10.7	-2.9	-8.4	2.5	1560	0.69
tlnorm	RP	109.0	45.5	4.8	5.3	-11.4	17.7	0.84	29.4	28.7	-0.2	0.6	-0.8	1560	0.69
tlnorm	min.AIC	109.0	45.5	-3.6	-3.0	-21.5	13.3	0.87	25.5	24.2	-0.0	0.0	-0.2	1560	0.69
tlnorm	min.BIC	109.0	45.5	-4.5	-3.7	-21.9	9.2	0.88	25.0	23.1	0.0	-0.0	-0.2	1560	0.69
exp	lnorm	123.4	85.6	-120.8	-119.3	-170.1	-81.7	0.00	76.0	75.0	6.1	5.3	-47.7	1560	0.58
exp	tlnorm	123.4	85.6	-71.1	-70.9	-90.9	-52.5	0.00	35.7	35.7	6.0	4.3	-48.6	1560	0.58
exp	exp	123.4	85.6	-0.1	0.0	-8.4	7.9	0.95	16.1	16.0	0.0	0.0	0.0	1560	0.58
exp	Weibull	123.4	85.6	-0.1	-0.0	-9.5	8.6	0.91	15.9	15.8	-0.0	-0.0	0.0	1560	0.58
exp	gamma	123.4	85.6	-0.1	0.0	-9.0	8.4	0.95	17.2	17.1	-0.0	0.0	0.0	1560	0.58
exp	RP	123.4	85.6	-0.3	-0.1	-10.9	9.1	0.95	19.6	19.4	-0.0	0.1	-0.1	1560	0.58
exp	min.AIC	123.4	85.6	-0.1	-0.0	-9.7	8.6	0.93	16.4	16.1	-0.0	0.0	0.0	1560	0.58
exp	min.BIC	123.4	85.6	-0.1	0.0	-8.5	8.0	0.94	16.1	16.0	0.0	0.0	0.0	1560	0.58
RP	lnorm	55.4	32.5	-9.5	-9.4	-15.7	-4.0	0.07	11.9	11.8	1.0	1.5	-1.0	1559	0.80
RP	tlnorm	55.4	32.5	-8.8	-8.7	-14.4	-3.6	0.08	11.1	11.1	1.0	1.5	-1.1	1559	0.80
RP	exp	55.4	32.5	3.2	3.2	-0.0	6.3	0.43	5.8	5.8	-0.4	-3.6	-3.0	1559	0.80
RP	Weibull	55.4	32.5	3.2	3.2	-0.3	6.3	0.33	4.8	4.8	-0.3	-3.6	-3.0	1559	0.80
RP	gamma	55.4	32.5	3.5	3.6	0.2	6.7	0.35	5.7	5.7	-1.0	-4.1	-2.5	1559	0.80
RP	RP	55.4	32.5	0.2	0.3	-4.2	4.1	0.94	8.3	8.2	0.1	-0.0	-0.3	1559	0.80
RP	min.AIC	55.4	32.5	0.2	0.3	-4.3	4.1	0.94	8.3	8.2	0.1	-0.0	-0.3	1559	0.80
RP	min.BIC	55.4	32.5	0.1	0.3	-5.3	4.1	0.93	8.3	8.2	0.1	-0.0	-0.3	1559	0.80

Bibliography

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 701-726.

Aalen, O., Borgan, O., & Gjessing, H. (2008). *Survival and event history analysis: a point process point of view.* Springer Science & Business Media.

Agresti, A. (2013). *Categorical Data Analysis* - 3rd ed., John Wiley & Sonds, Inc.

Andersson, T., Dickman, P., Eloranta, S., & Lambert, P. (2011). Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Med Res Methodol, 11*(1), 96.
http://dx.doi.org/10.1186/1471-2288-11-96

Bauckhage, C. (2014). Characterizations and Kullback-Leibler Divergence of Gompertz Distributions. *arXiv preprint arXiv:1402.3193*.

Breslow, N. E. (1972). Contribution to the discussion of the paper by DR Cox. *Journal of the Royal Statistical Society, Series B* 34.2 (1972): 216-217.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological methods & research,* 33(2), 261-304.

Casella, G., Robert, C. P., & Wells, M. T. (2004) Generalized Accept-Reject Sampling Schemes. *Lecture Notes-Monograph Series*, 342-347. http://dx.doi.org/10.1214/lnms/1196285403.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (methodological), 34*(2), 187-220.

Cox, C., Chu, H., Schneider, M. F., & Muñoz, A. (2007). Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in medicine, 26*(23), 4352-4374.

Cox, C. (2008). The generalized F distribution: an umbrella for parametric survival analysis. *Statistics in medicine, 27*(21), 4301-4312.

Crowder, M. J. (2001). *Classical competing risks.* CRC Press.

Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika, 81*(3), 515-526.

Greenwood, M. (1926). The natural duration of cancer, Reports of public health and related subjects, vol. 33.

Hoeting, J. A., Madigan, D., Raftery, A., E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical sicence,* 382-401.

Hollander, M., Chicken, E., & Wolfe, D. (2013). *Nonparametric Statistical Methods, 3rd Edition.* John Wiley & Sons.

Hosmer Jr, D. W., & Lemeshow, S. (1999). Applied Survival Analysis: Regression Modelling of Time to Event Data (1999).

Hunter, D. R. (2006). Statistics 553: asymptotic tools. *Penn State University (Fall 2011).*

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions Vol. 1.* New York: Wiley.

Kalbfleisch, J. D., & Prentice, R. L. (2011). *The statistical analysis of failure time data* (Vol. 360). John Wiley & Sons.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association, 53*(282), 457-481.

Khan, M. S., Pasha, G. R., & Pasha, A. H. (2008). Theoretical analysis of Inverse Weibull distribution. *WSEAS Transactions on Mathematics, 7*(2), 30-38.

Kleinbaum, D. G., & Klein, M. (2005). Survival analysis: A self-learning approach. *Springer, New York, USA.*.

Latimer, N. (2011). NICE DSU technical support document 14: survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data. *Sheffield: Report by the Decision Support Unit.*

Lehmann, E. L. & Casella, G. (1998). *Theory of Point Estimation, Second Edition.* Springer.

Loprinzi, C. L., Laurie, J. A., Wieand, H. S., Krook, J. E., Novotny, P. J., Kugler, J. W., Bartel, J., Law, M., Bateman, M. & Klatt, N. E. (1994). Prospective evaluation of prognostic variables from patient-completed questionnaires. North Central Cancer Treatment Group. *Journal of Clinical Oncology, 12*(3), 601-607.

Melnikov, A., & Romaniuk, Y. (2006). Evaluating the performance of Gompertz, Makeham and Lee-Carter mortality models for risk management with unit-linked contracts. *Insurance: Mathematics and Economics, 39*(3), 310-329.

Nelson, W.B. (2005). *Applied life data analysis* (Vol. 577). John Wiley & Sons.

Miller, R., Gong, G., & Muñoz, A. (1981). *Survival analysis.* New York: Wiley.

Nelson, W. (1972). Theory and Applications of Hazard Plotting for Censored Failure Data. *Technometrics*, 945-966.

Peterson Jr, A. V. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *Journal of the American Statistical Association, 72*(360a), 854.858.

Prentice, R. L. (1974). A log gamma model and its maximum likelihood estimation. *Biometrika, 61*(3), 539-544.

Prentice, R. L. (1975). Discrimination among some parametric models. *Biometrika, 62*(3), 607-614.

Richardson, L. F., & Gaunt, J. A. (1927). The deferred approach to the limit. Par I. Single lattice. Part II. Interpenetrating lattices. *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character, 226*, 299-361. http://dx.doi.org/10.1098/rsta.1927.0008.

Royston, P., & Lambert, P. (2011). *Flexible parametric survival analysis using Stata.* College Station, TX: Stata Press.

Royston, P., & Parmar, M. K. (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), 2175-2197. http://dx.doi.org/10.1002/sim.1203.

Royston, P., & Parmar, M. K. (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC medical research methodology, 13*(1), 152. http://dx.doi.org/10.1186/1471-2288-13-152.

Stacy, E. W. (1962). A generalization of the gamma distribution. *Annals of Mathematical Statistics* 1187-1192. http://dx.doi.org/10.1214/aoms/1177704481.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika, 69*(1), 239-241.

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: extending the Cox model.* Springer Science & Business Media.

R Package Bibliography

Broström, G. (2012). eha: event history analysis. R package, version 2.4-3.

Gilbert, P. & Varadhan, R. (2015). numDeriv: Accurate Numerical Derivatives. R package version 2014.2-1. https://CRAN.R-project.org/package=numDeriv.

Grolemund, G. & Wickham, H. (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL http://www.jstatsoft.org/v40/i03/.

S original by Hess, K. & R port by Gentleman, R. (2014). muhaz: Hazard Function Estimation in Survival Analysis. R package version 1.2.6.
https://CRAN.R-project.org/package=muhaz.

Jackson, C. (2015). flexsurv: Flexible Parametric Survival and Multi-State Models. R package version 0.7. https://CRAN.R-project.org/package=flexsurv.

R Core Team (2015). R: A language and enviroment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Therneau, T. M. (2015). survival: Survival Analysis. R package version 2.38.
https://CRAN.R-project.org/package=survival.

Wickham, H. & Francois, R. (2015). dplyr: A Grammar of Data Manipulation. R package version 0.4.3. https://CRAN.R-project.org/package=dplyr