Stockholms universitet

Survival Analysis and its Application to Childhood Cancer Data

Ida Hed Myrberg

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2020:2
Matematisk statistik
Februari 2020

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2020:2**
http://www.math.su.se

Survival Analysis and its Application to Childhood Cancer Data

Ida Hed Myrberg*

February 2020

Abstract

Survival analysis denotes a collection of statistical methods, where time to one or several events of interest is considered, for example death, or the onset of a disease. In this thesis, essential concepts and quantities within the field of survival analysis, such as censoring, hazard and survival functions, are introduced. A number of selected methods are presented in detail; the non-parametric Nelson-Aalen and Kaplan-Meier estimators, the semi-parametric Cox proportional hazards model, the fully parametric accelerated failure time, proportional hazards and proportional odds models, the flexible parametric Royston and Parmar proportional hazards and proportional odds models, and the distribution-free quantile regression model. In order to estimate the power of detecting deviations from the proportional hazards assumption, Monte Carlo simulations are used, assuming underlying Weibull distributions. Large deviations are detectable at a high power even for moderate sample sizes, while small deviations are hard to detect even for large sample sizes. The Type I error rates are accurate when the proportional hazards assumption is fulfilled, for all investigated sample sizes and censoring proportions. Furthermore, Cox, Weibull, and Royston and Parmar proportional hazards models are compared, given an underlying Weibull distribution, and given that the proportional hazards assumption is fulfilled, using Monte-Carlo simulations. The three methods show comparable estimates and standard errors. Average coefficient estimates, standard errors, and confidence interval coverage of quantile regression models are evaluated using Monte-Carlo simulations, showing accurate coefficients estimates, but too low confidence interval coverage for small and moderate sample sizes. The proposed methods are used to investigate the association between the so called event-free survival time of children with acute lymphoblastic leukemia (ALL) and a variety of risk factors in a heavily right-censored dataset. Fully parametric distributions do not fit the data well, but coefficient estimates are comparable to semi-parametric and flexible parametric models. Some covariates do not fulfill the proportional hazards assumption, and are better modeled dependent on time in Royston and Parmar models. Quantile regression only works for small probabilities, since the proportion censored observations is high. Even so, this method provides a different perspective that could be useful in a clinical setting.

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

Abstract

Survival analysis denotes a collection of statistical methods, where time to one or several events of interest is considered, for example death, or the onset of a disease. In this thesis, essential concepts and quantities within the field of survival analysis, such as censoring, hazard and survival functions, are introduced. A number of selected methods are presented in detail; the non-parametric Nelson-Aalen and Kaplan-Meier estimators, the semi-parametric Cox proportional hazards model, the fully parametric accelerated failure time, proportional hazards and proportional odds models, the flexible parametric Royston and Parmar proportional hazards and proportional odds models, and the distribution-free quantile regression model. In order to estimate the power of detecting deviations from the proportional hazards assumption, Monte Carlo simulations are used, assuming underlying Weibull distributions. Large deviations are detectable at a high power even for moderate sample sizes, while small deviations are hard to detect even for large sample sizes. The Type I error rates are accurate when the proportional hazards assumption is fulfilled, for all investigated sample sizes and censoring proportions. Furthermore, Cox, Weibull, and Royston and Parmar proportional hazards models are compared, given an underlying Weibull distribution, and given that the proportional hazards assumption is fulfilled, using Monte-Carlo simulations. The three methods show comparable estimates and standard errors. Average coefficient estimates, standard errors, and confidence interval coverage of quantile regression models are evaluated using Monte-Carlo simulations, showing accurate coefficients estimates, but too low confidence interval coverage for small and moderate sample sizes. The proposed methods are used to investigate the association between the so called event-free survival time of children with acute lymphoblastic leukemia (ALL) and a variety of risk factors in a heavily right-censored dataset. Fully parametric distributions do not fit the data well, but coefficient estimates are comparable to semi-parametric and flexible parametric models. Some covariates do not fulfill the proportional hazards assumption, and are better modeled dependent on time in Royston and Parmar models. Quantile regression only works for small probabilities, since the proportion censored observations is high. Even so, this method provides a different perspective that could be useful in a clinical setting.

Acknowledgments

I would like to thank Paolo Frumento, my external supervisor, for introducing me to the interesting field of quantile regression, for interesting discussions, and for taking the time to co-supervise me on this project. I would also like to thank Mathias Millberg Lindholm, my supervisor from Stockholm University, for his time and valuable input on all aspects of this thesis. Thank you to Adrian Levitsky for highly appreciated comments on the abstract, introduction and discussion. A big thank you to Mats Heyman and the Nordic Society of Paediatric Haematology and Oncology (NOPHO) for providing me with patient data from children suffering from acute lymphoblastic leukemia. Finally, I would like to thank my colleagues at the Childhood Cancer Research Unit at the Department of Women's and Children's Health, Karolinska Institutet, for giving me the opportunity to be a part of lifesaving research and for awakening my interest in survival analysis.

Contents

Abbreviations

AFT	Accelerated failure time
AIC	Akaike information criterion
ALL	Acute lymphoblastic leukemia
BIC	Bayesian information criterion
CDF	Cumulative distribution function
CI	Confidence interval
CNS	Central nervous system
CSF	Cerebrospinal fluid
EFS	Event-free survival
HM	Hazard model
HR	Hazard ratio
IQR	Inter-quartile range
KM	Kaplan-Meier
NA	Nelson-Aalen
NOPHO	Nordic Society of Paediatric Haematology and Oncology
NPHM	Non-proportional hazards model
NPOM	Non-proportional odds model
OM	Odds model
OS	Overall survival
PCH	Piecewise constant hazards
PH	Proportional hazards
PHA	Proportional hazards assumption
PHM	Proportional hazards model
PO	Proportional odds
POA	Proportional odds assumption
POM	Proportional odds model
QR	Quantile regression
RCS	Restricted cubic splines
RG	Risk group
RP	Royston and Parmar

1 Introduction

The theme of this thesis is the thriving field of survival analysis, used widely in medical statistics. Survival analysis has been an important part of the medical science throughout the 20th century up until today, and the two go hand in hand – medical researchers use the methods developed by statisticians, and statisticians get ideas and data from medical researchers for developing new methods. With an application to childhood cancer data, this master thesis in mathematical statistics is in itself a manifestation of this relationship.

It is very common that medical researchers prefer non-parametric methods such as Kaplan-Meier curves and the log-rank test, introduced in Sections 3.2.2 and 3.2.3, or semi-parametric methods such as the Cox proportional hazards model, introduced in Section 3.4. This can partly be explained by small sample sizes, interpretability, and habit. However, each method has its own assumptions, and deviations from these can be modeled more accurately using other methods.

In this thesis, the reader is given an overview of classical methods, such as the ones mentioned above, along with more recent developments in survival analysis. Monte Carlo simulations are performed to evaluate some of the methods, and all methods are applied to a dataset containing patient data from children diagnosed with acute lymphoblastic leukemia (ALL).

1.1 Objectives

The main objectives of this thesis are to

1. Review the non-parametric survival analysis methods Nelson-Aalen and Kaplan-Meier estimators, the semi-parametric Cox proportional hazards models, the fully parametric Weibull and log-logistic models, the flexible parametric Royston and Parmar models, and the distribution-free quantile regression model, in terms of model assumptions, estimation, inference, and interpretation.

2. Evaluate various properties of the methods using Monte-Carlo simulations.

3. Apply the methods to investigate the association between survival time in children with ALL, and a number of risk factors.

1.2 Outline

First, time-to-event data and important concepts in survival analysis are introduced to the reader in Sections 2.1 and 2.2. A brief introduction to childhood cancer and ALL is given in Section 2.3. The methods used in the thesis are thoroughly presented in Section 3. In Section 4, results from Monte-Carlo simulations used to evaluate some of the methods are presented. Results of the application of the selected methods to the childhood ALL data are presented in Section 5, starting with a description of the data in Section 5.1. A summary and a discussion is given in Section 6, followed by conclusions in Section 7. Some supplementary theory and simulation results are given in Appendices A, B and C.

2 Background

In this section, the reader is given an introduction to key concepts in survival analysis in Sections 2.1 and 2.2, along with a brief overview of childhood cancer in Section 2.3 and acute lymphoblastic leukemia in Section 2.3.1.

2.1 Time-to-event data

Time-to-event data, also known as survival data or failure time data, measures time from some clearly defined start time point, to the occurrence of an event of interest. Common objectives within the field of medicine are for example time from birth until the onset of a certain disease, and time from onset of a certain disease until death. However, time-to-event data can represent a wide range of endpoints, such as time until the birth of a person's first child, the lifetime of a light-bulb, or time from unemployment to employment. By definition, time variables are strictly positive, a property that could complicate common analysis approaches such as linear regression. However, the analysis of time-to-event data is special mainly due to the presence of censoring, a type of partial information, introduced in Section 2.1.1 below.

2.1.1 Censoring and truncation

Censoring of survival data occurs when the event of interest is unobserved for some individuals during the studied time frame. There are three types of censoring; right-censoring, left-censoring, and interval censoring. Right-censoring occurs when the event of interest has not yet happened during the time in which the subject is observed. This can be due to several reasons, for example the end of study, the subject could have moved abroad, or the subject could have had a competing event that prevents the event of interest from happening. If the start time is defined individually for each subject, rather than by a fixed calendar time, for example the time of diagnosis of a certain disease, then each subject will have a different follow-up time even if everyone is followed until the end of study, provided that they are not diagnosed the exact same date. Right-censoring is the most common type of censoring, and is present in most survival data. In Figure 1, examples of survival times are given. The left panel shows calendar time, and the right shows study time. On the calendar time scale, each observation has a different entry time, while on the study time scale, the entry time is defined as the start time, and hence each observation starts at time zero.

Left-censoring occurs when the event of interest has already occurred when a subject is entered into the study, but the exact event time is unknown. In interval censoring, it is only known that the event occurred in some time interval, but here too, the exact time-point is unknown.

Left truncation occurs when only subjects who have not yet experienced the event are entered into the study. For example, if studying persons suffering from heart attacks, and only including persons admitted to a hospital, persons who die before hospital admission are not included in the study. Hence, persons need to have survived long enough to be admitted to the hospital. For this reason, truncation is also called *length-biased sampling* (Hosmer, Lemeshow, & May, 2008).

In this thesis, only right-censoring without truncation is considered. To read more about censoring and truncation, the interested reader is referred to Klein and Moeschberger (2003, Ch.

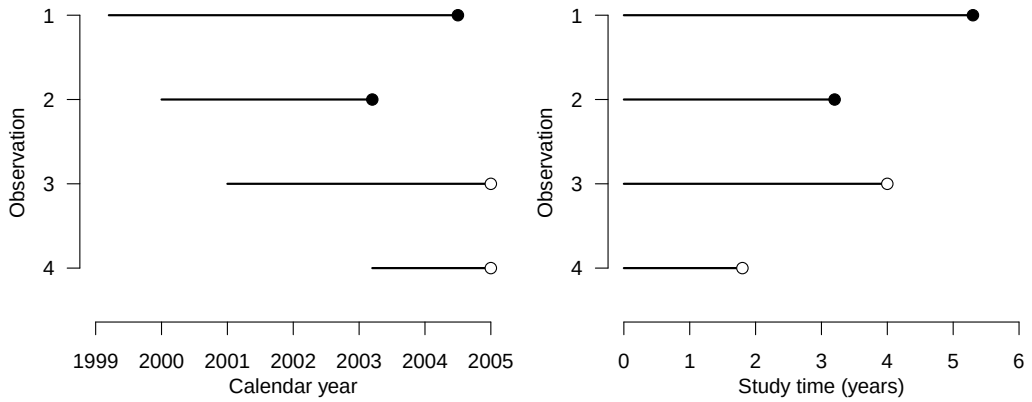

Figure 1: Example of survival times. An open circle indicates a censoring, and a filled circle indicates an event.

3).

For right-censored data, the random variable C is introduced, denoting the time of censoring, and δ, an indicator or *status* variable for whether or not the event was observed before the time of censoring, i.e. $\delta = 1$, if the event was observed before the time of censoring, and $\delta = 0$, otherwise. In many standard survival analysis methods, the time to censoring, C, and the time to event, T, are assumed (conditionally) independent.

2.1.2 Typical data structure

In right-censored survival data, apart from the survival time, a status variable is needed, indicating whether or not the event was observed before censoring. The survival time should be given in an appropriate unit of choice, e.g. seconds, days, weeks, months, or years. In Table 1, an example of survival data structure is given. The first and second observations experienced the event of interest at time points 5.3 and 3.2, respectively. The third and fourth observations were censored at time points 4.0 and 1.8, respectively. Along with the time and status variables, an arbitrary number of covariates, $x_1, ..., x_p$, may be provided. The data structure can be extended to incorporate different types of events, time-varying covariates, recurrent events, interval censoring, and truncation.

Table 1: Example of survival data structure

Observation	Time	Status
1	5.3	1
2	3.2	1
3	4.0	0
4	1.8	0

The time to event is most often assumed to be continuous, but in practice there will be rounding errors due to inexact measurement. For example, in clinical applications, survival time is often measured in number of days. Inexact measuring of survival time can lead to so called ties, where two or more subjects experience the event of interest at the same time point. On a continuous time scale this is not a problem, but with rounded units of measurement, this occurs more or less frequently. In different applications, there are different ways of handling ties. In ties between one event time and one or more censoring times, the convention is to consider

the event time occurring prior to the censoring time. Ties between censoring times do not cause problems in calculations of estimates (Aalen, Borgan, & Gjessing, 2008, Ch. 3.1.2, p. 84).

2.2 Important quantities in survival analysis

Some important quantities in survival analysis are listed below. For now, the quantities are introduced unconditional on covariates $\boldsymbol{x} = (x_1, ..., x_p)'$.

$$
\begin{aligned}
T &= \text{Time to event (random variable)} \\
t &= \text{Time to event (observed)} \\
C &= \text{Time to censoring (random variable)} \\
c &= \text{Time to censoring (observed)} \\
Y &= \min(T, C) \\
y &= \min(t, c) \\
\delta &= \mathrm{I}(t < c) \\
f(t) &= \text{Probability density function} \\
F(t) &= \text{Cumulative distribution function} \\
S(t) &= \text{Survival function} \\
h(t) &= \text{Hazard function} \\
H(t) &= \text{Cumulative hazard function} \\
\boldsymbol{x} &= (x_1, ..., x_p)' = \text{Column vector of covariates} \\
\boldsymbol{\beta} &= (\beta_0, ..., \beta_p)' = \text{Column vector of coefficients} \\
\mathscr{R}(t) &= \text{Risk set at time } t^- \\
Y(t) &= |\mathscr{R}(t)| = \text{Number at risk at time } t^-
\end{aligned}
$$

The time to event, denoted by T, is a random variable that measures time from start (t_0, constant) until an event of interest occurs. Assume from here on that T is a non-negative continuous random variable. Let $T \sim F$, with $F(\cdot)$ being the cumulative distribution function, i.e.

$$F(t) = P(T \leq t) = \int_0^t f(s)ds,$$

where $f(\cdot)$ denotes the probability density function, $f(t) \geq 0$. The survival function $S(\cdot)$ is defined as

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

and thus denotes the probability of being event-free ("surviving") at least until time t. Since $S(t)$ is a probability, it takes values between 0 and 1. The hazard function $h(t)$ at time t is the instantaneous hazard of having the event at time t, and is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | T > t), \quad h(t) \geq 0.$$

The cumulative hazard $H(\cdot)$ is defined as

$$H(t) = \int_0^t h(s)ds, \quad H(t) \geq 0,$$

It is common to denote the hazard and cumulative hazard with $\lambda(t)$ and $\Lambda(t)$, respectively, but to avoid confusing them with the intensity processes described in Section 3.1, and with certain parameters in some of the parametric distributions presented in Section 3.5.1, $h(t)$ and $H(t)$ are used throughout this thesis. Any of the quantities $f(t), F(t), h(t), H(t), S(t)$ uniquely defines the distribution of T, and are connected through the following relations,

$$\begin{aligned}
h(t) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | T > t) \\
&= \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \cap T > t)}{\Delta t P(T > t)} \\
&= \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t P(T > t)} \\
&= \frac{f(t)}{S(t)} \\
&= \frac{-\frac{d}{dt}[1 - F(t)]}{S(t)} \\
&= -\frac{S'(t)}{S(t)}.
\end{aligned}$$

Due to the above relation, it also holds that

$$\begin{aligned}
h(t) &= -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \log[S(t)] \\
&\iff \\
\log[S(t)] &= -\int_0^t h(s)ds \\
&\iff \\
S(t) &= \exp\left\{-\int_0^t h(s)ds\right\} = \exp\{-H(t)\} \\
&\iff \\
H(t) &= -\log S(t),
\end{aligned}$$

and

$$f(t) = h(t)S(t).$$

The above results are summarized below,

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

$$F(t) = \int_0^t f(s)ds$$

$$H(t) = \int_0^t h(s)ds \tag{2}$$

$$H(t) = -\log S(t) \tag{3}$$

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

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

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

$$S(t) = \exp\{-H(t)\} = \exp\left\{-\int_0^t h(s)ds\right\}. \tag{4}$$

The above quantities $f(t), F(t), S(t), h(t), H(t)$ are properties of T, unless stated otherwise. For example, the time-to-event random variable C has the corresponding quantities $f_C(t)$, $F_C(t)$, $S_C(t)$, $h_C(t)$, $H_C(t)$. For any given individual, the following triplet of variables is observed $(y_i, \delta_i, \boldsymbol{x}_i)$. All of the above quantities can be expressed as functions of $t|\boldsymbol{x}$, which will be used when introducing regression models in Sections 3.4, 3.5, 3.6, and 3.7.

2.3 Childhood cancer

In cancer research, information on survival is collected by following up each patient from time of diagnosis until death or censoring, often also registering other events such as relapse of cancer, stem cell transplantation, secondary malignancy, toxicities such as thrombosis and seizures, surgeries, and other events specific for different cancer types, along with relevant background variables known at diagnosis, and treatment-specific factors. Death is an important outcome, and survival until death is called overall survival (OS), defined as time from onset, diagnosis, or start of treatment, to death or censoring. However, other endpoints can be of importance, and it is very common to use event-free survival (EFS) as an outcome, defined as time to death or another or several other events, such as relapse of cancer, getting another type of cancer, whichever comes first.

In Sweden, there are several cancer registries for a variety of cancer types. The different registries provide valuable data for cancer research. Different registries can be connected, and, hence, a researcher with an approved ethical application can combine different types of information from several registries for the same study cohort, enabling important research on, for example, survival in childhood cancer. The NOPHO registry is one such registry, started by the Nordic Society of Paediatric Haematology and Oncology (NOPHO), and includes survival data and important clinical data on Nordic and Baltic children with various cancer diagnoses.

Yearly, around 300 children in Sweden are diagnosed with cancer, and it is the primary cause of death in children aged 1-14 years (Bergling et al., 2019). The cause of childhood cancer is mostly unknown, as children have not been exposed for an extended period to known risk factors of cancer. The most common childhood cancers are brain tumors and leukemias (Kaatsch, 2010). Survival in childhood cancer is nowadays around 80% in developed countries,

although some cancer types are still not curable (Bergling et al., 2019).

2.3.1 Acute lymphoblastic leukemia

Acute lymphoblastic leukemia (ALL) is a cancer constituting 27% of all childhood cancer cases (Kaatsch, 2010), most common in children, with a peak incidence at 2-5 years of age (Inaba, Greaves, & Mullighan, 2013). It is not a tumor cancer, as it consists of immature white blood cells proliferating uncontrollably in the bone marrow. During the time preceding an ALL diagnosis, children with ALL can bleed or bruise easily, and be pale and tired (Hunger & Mullighan, 2015). ALL is fatal if left untreated. Survival after ALL has improved enormously during the 20th century, from being practically incurable in the 1950s and 1960s, to around 90% survival in the 21st century. This is due to combining several chemotherapeutic drugs, and stratifying the treatment intensity depending on characteristics of the patient and the patient's cancer cells, and on early treatment response (Hunger & Mullighan, 2015). ALL was one of the first malignant cancer diseases to be successfully cured (W. E. Nelson & Behrman, 1996, Ch. 449.1).

Since ALL is a rare disease, with an incidence of around 3 cases per 100 000 persons under 20 years of age (Hunger & Mullighan, 2015), hospitals, regions and countries often need to collaborate in clinical studies of ALL. The NOPHO ALL2008 protocol, which started enrolling patients in July 2008 (Toft et al., 2018), is a collaboration between the member countries of NOPHO - the Nordic countries, Estonia, and Lithuania. The treatment consists of several chemotherapeutic and steroid drugs, given with a varying intensity, depending on the risk stratification of the child. For all risk groups, the treatment length is 2.5 years. In this thesis, data from the NOPHO registry will be used. Details of the data and included patients are presented in Section 5.1.

3 Methods

This section contains a description of a number of methods from the field of survival analysis. First, a brief introduction to the counting process view of survival analysis is given in Section 3.1. Then the non-parametric Nelson-Aalen estimator, Kaplan-Meier estimator and the log-rank test are introduced in Sections 3.2.1, 3.2.2, and 3.2.3. This is followed by the general likelihood function for survival data in Section 3.3. The regression methods Cox proportional hazards model, parametric modeling, Royston and Parmar models, and quantile regression are presented in Sections 3.4, 3.5, 3.6, and 3.7. The methods are compared in Section 3.8.

3.1 Counting process approach to survival analysis

Survival analysis deals with events occurring over the course of time, making it natural to include stochastic process theory and methods. A brief introduction is given here, following Aalen et al. (2008, Ch. 1-3). Although stochastic process theory is fundamental for many results in survival analysis, this section is not crucial for the understanding of the rest of the thesis.

Let $\mathscr{F}_t$ denote the event history up until time t. A stochastic process $M = \{M(t); t \in [0, \gamma]\}$, where γ is finite, is called a martingale with respect to the history $\mathscr{F}_t$ if

$$E(M(t)|\mathscr{F}_s) = M(s) \text{ for all } s < t,$$

known as the martingale property. A sub-martingale $X = \{X(t); t \in [0, \gamma]\}$ satisfies

$$E(X(t)|\mathscr{F}_s) \geq X(s) \text{ for all } s < t.$$

An important result in martingale theory is the Doob-Meyer decomposition, stating that a sub-martingale X can be expressed as

$$X^* + M,$$

where X^* is a non-decreasing predictable process and M is a zero-mean martingale. A continuous-time counting process $N(t)$, counting the number of events until and including time t, fulfills

- $N(t) \geq 0$
- $N(t)$ is right-continuous.
- Jumps are of size 1.
- $N(t)$ is constant in between jumps.

Since $N(t)$ is non-decreasing, it is a sub-martingale. Let $\lambda(t)$ be the intensity process of $N(t)$, defined by

$$\lambda(t)dt = P(dN(t) = 1|\mathscr{F}_{t-}) = E(dN(t)|\mathscr{F}_{t-}).$$

The Doob-Meyer decomposition states that $N(t)$ can be written $N(t) = \Lambda(t) + M(t)$, where $\Lambda(t)$ is the cumulative intensity process defined by $\Lambda(t) = \int_0^t \lambda(s)ds$. In survival analysis, the multiplicative intensity model is given by $\lambda(t) = Y(t)h(t)$, where $h(t)$ is the hazard rate, and

$Y(t)$ is the number at risk just before time t. The increment $dN(t)$ can be written

$$dN(t) = \lambda(t)dt + dM(t) = Y(t)h(t)dt + dM(t).$$

Dividing with $Y(t)$ and multiplying by $I(t) = I(Y(t) > 0)$ yields

$$\frac{I(t)}{Y(t)}dN(t) = I(t)h(t)dt + \frac{I(t)}{Y(t)}dM(t).$$

Integrating both sides yields

$$\int_0^t \frac{I(s)}{Y(s)}dN(s) = \int_0^t I(s)h(s)ds + \int_0^t \frac{I(s)}{Y(s)}dM(s).$$

In an application with a finite number of event times, the left-hand side simply becomes a sum over the event times, according to

$$\int_0^t \frac{I(s)}{Y(s)}dN(s) = \sum_{t_j \leq t} \frac{I(t_j)}{Y(t_j)},$$

which is recognized as the Nelson-Aalen estimator, presented in detail in Section 3.2.1. It can be shown that a stochastic integral with respect to a mean-zero martingale is itself a mean-zero martingale, hence

$$E\left(\int_0^t \frac{I(s)}{Y(s)}dN(s) - \int_0^t I(s)h(s)ds\right) = 0.$$

The Nelson-Aalen estimator is thus an unbiased estimator of $\int_0^t I(s)h(s)ds$, which is very similar to the cumulative hazard, except for the factor $I(s)$ that restricts the non-parametric estimator of the cumulative hazard to be within the observed time frame. For a derivation of the Nelson-Aalen variance estimator, the interested reader is referred to Aalen et al. (2008, Ch. 3).

Note that, since the jumps in a counting process, $N(t)$, are of size 1, tied event times are not accounted for in the counting process theory presented here. Various ways of handling tied event times are presented in the following method sections.

3.2 Non-parametric estimators

Non-parametric estimators of cumulative hazard and survival are widely used in the field of medicine for visualization and hypothesis testing, and the most popular estimator is the Kaplan-Meier estimator of Section 3.2.2. These estimators are attractive due their straight-forward approach, their suitability for small sample sizes, their easy-to-interpret visualization, and their lack of a parametric assumptions. All formulas of estimators, variance estimators and confidence intervals of Sections 3.2.1 and 3.2.2 are taken from Aalen et al. (2008, Ch. 3), and Hosmer et al. (2008, Ch. 2.4) for Section 3.2.3, unless otherwise stated.

3.2.1 Nelson-Aalen estimator

The Nelson-Aalen estimator of the cumulative hazard, $H(t)$, was first introduced by W. Nelson (1969, 1972), and later extended by Aalen (1975, 1978) using a counting process approach. The

estimator is given by a sum over the event times $t_j, j = 1, ..., m$, with m being the number of events, as follows

$$\hat{H}_{NA}(t) = \sum_{t_j \leq t} \frac{1}{Y(t_j)},$$

where $Y(t_j)$ is the number at risk just before time t_j. The censoring is taken into account by the denominators, where censored cases are excluded from the risk set at the time of censoring. At any given time t, the Nelson-Aalen estimator is asymptotically normal as the number of events goes to infinity, i.e. the estimator is approximately normal for large enough samples. Using a counting process approach, the Nelson-Aalen estimator is derived in Section 3.1.

An estimate of the variance of $\hat{H}_{NA}(t)$ is given by

$$\widehat{\text{Var}}[\hat{H}_{NA}(t)] = \hat{\sigma}^2_{NA}(t) = \sum_{t_j \leq t} \frac{1}{Y(t_j)^2}.$$

As discussed in Section 2.1.2, tied events are present in many real-life survival datasets. There are two approaches to tied events for non-parametric estimators. The first one considers time as inherently continuous, and that ties occur due to rounding. The second one considers time as truly discrete. We now write the Nelson-Aalen estimate as

$$\hat{H}_{NA}(t) = \sum_{t_j \leq t} \Delta\hat{H}_{NA}(t_j), \tag{5}$$

where $\Delta\hat{H}_{NA}(t_j)$ are the Nelson-Aalen increments. When no ties are present, $\Delta\hat{H}_{NA}(t_j) = 1/Y(t_j)$. When ties are present, a choice has to be made between the first and the second approach to tied event times. Let d_j denote the number of events at time t_j. With the first approach, where time is considered continuous, the Nelson-Aalen increments are given by

$$\Delta\hat{H}_{NA}(t_j) = \sum_{l=0}^{d_j - l} \frac{1}{Y(t_j) - l},$$

assuming that the true survival times are not exactly the same, i.e. ties are due to rounding. With the second approach, where time is considered discrete, the Nelson-Aalen increments are given by

$$\Delta\hat{H}_{NA}(t_j) = \frac{d_j}{Y(t_j)},$$

assuming that the d_j tied events actually occurred at the same discrete time-point t_j.

Similar to Eq. (5), the variance estimate of $\hat{H}_{NA}$ can be written as

$$\hat{\sigma}^2_{NA}(t) = \sum_{t_j \leq t} \Delta\hat{\sigma}_{NA},$$

where $\Delta\hat{\sigma}_{NA} = 1/Y(t_j)^2$ when no ties are present, and

$$\Delta\hat{\sigma}_{NA} = \sum_{l=0}^{d_j - 1} 1/(Y(t_j) - l)^2,$$

when continuous time scale is assumed, and

$$\Delta\hat{\sigma}_{NA} = (Y(t_j) - d_j)d_j/Y(t_j)^3,$$

when a discrete time scale is assumed.

An approximate $(1-\alpha)\%$ confidence interval is given by

$$\hat{H}_{NA}(t) \pm z_{1-\alpha/2}\hat{\sigma}_{NA}(t),$$

of which the asymptotic properties are improved by using a log-transformation according to

$$\hat{H}_{NA}(t) \exp\left\{\pm \frac{z_{1-\alpha/2}\hat{\sigma}_{NA}(t)}{\hat{H}_{NA}(t)}\right\}.$$

Using Eq. (4), stating that $S(t) = \exp\{-H(t)\}$, the Nelson-Aalen estimator of the cumulative hazard can be used to estimate the survival function by $\hat{S}_{NA}(t) = \exp\{-\hat{H}_{NA}(t)\}$.

The estimated cumulative hazard is visualized by a step-function, with time since origin on the x-axis and the estimated cumulative hazard function on the y-axis. Examples of these plots are given in Figure 2, for varying sample sizes. The event times are simulated from a Weibull distribution, introduced in Table 2 of Section 3.5.1, with parameters $\lambda = 1, \alpha = 0.9$, and the censoring distribution is uniform on the interval (0.2,2), independent of the event times. From the figure it can be seen that for small sample sizes, the steps are large, and for large sample sizes, the curve is more smooth.

3.2.2 Kaplan-Meier estimator

In the 1950s, Edward L. Kaplan and Paul Meier each submitted a paper on an estimator for the survival function, $S(t)$, to the Journal of the American Statistical Association. Since the papers were so similar, the editor convinced the two to write a joint paper (Tobacman, 2011). This resulted in the now historic and frequently cited paper from 1958, in which Kaplan and Meier introduced what they called the product-limit estimator, but what was later to be better known as the Kaplan-Meier estimator (Kaplan & Meier, 1958). The estimator is given by a product over the event times, as follows

$$\hat{S}_{KM}(t) = \prod_{t_j \leq t} \left(1 - \frac{1}{Y(t_j)}\right),$$

where $Y(t_j)$ is the number at risk just before time t. The variance can be estimated by

$$\widehat{\mathrm{Var}}[\hat{S}_{KM}(t)] = \hat{\sigma}^2_{KM}(t) = \hat{S}_{KM}(t)^2 \sum_{t_j \leq t} \frac{1}{Y(t_j)^2}.$$

Note that the variance estimator $\hat{\sigma}^2_{KM}(t)$ is simply the Kaplan-Meier estimator $\hat{S}_{KM}(t)^2$ times the Nelson-Aalen variance estimator $\hat{\sigma}^2_{NA}(t)$.

In the case of tied event times, the continuous and discrete time approaches described in

Section 3.2.1 lead to the same Kaplan-Meier estimator

$$\hat{S}_{KM}(t) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{Y(t_j)}\right).$$

This is due to the fact that, if t_j is an event time with several tied events, the discrete time approach leads to the factor $(1 - d_j/Y(t_j))$, and the continuous time approach leads to the factor

$$\left(1 - \frac{1}{Y(t_j)}\right) \cdot \left(1 - \frac{1}{Y(t_j) - 1}\right) \cdots \left(1 - \frac{1}{Y(t_j) - (d_j - 1)}\right).$$

It can be shown with some simple algebra that these two quantities are the same. The continuous and discrete time approaches to tied events do not, however, lead to the same variance estimator. The continuous time approach leads to

$$\hat{\sigma}^2_{KM}(t) = \hat{S}_{KM}(t)^2 \hat{\sigma}^2_{NA}(t),$$

using the continuous time approach version of $\hat{\sigma}^2_{NA}(t)$ from Section 3.2.1. The discrete time approach leads to the variance estimator

$$\hat{\tilde{\sigma}}^2_{KM}(t) = \hat{S}_{KM}(t)^2 \sum_{t_j \leq t} \frac{d_j}{Y(t_j)(Y(t_j) - d_j)},$$

referred to as Greenwood's formula, which is commonly used in most applications.

An approximate $(1 - \alpha)\%$ confidence interval for $\hat{S}_{KM}(t)$ is given by

$$\hat{S}_{KM}(t) \pm z_{1-\alpha/2} \hat{\tilde{\sigma}}_{KM}(t).$$

However, the above confidence interval risks exceeding the value 1 or being less than the value 0, which is problematic, considering $S(t)$ is a probability. An alternative confidence interval, avoiding this problem, and also giving a better normal approximation, using the transformation $\log(-\log[\hat{S}_{KM}(t)])$, is given by

$$\hat{S}_{KM}(t)^{\exp\left\{\pm \frac{z_{1-\alpha/2}\hat{\tilde{\sigma}}_{KM}(t)}{\hat{S}_{KM}(t)\log[\hat{S}_{KM}(t)]}\right\}}. \tag{6}$$

In studies involving survival data, it is common to plot the estimated survival function against time, possibly separating the curves by groups, see Figure 2 for examples of Kaplan-Meier curves for varying sample sizes. Each event decreases the survival function, and each censoring is marked by a +-symbol. It is also common to include a risk table below the x-axis, showing how many subjects are still at risk of an event at selected time points.

3.2.3 The log-rank test

A common objective in survival analysis is the comparison of two or more hazard rates. The log-rank test is a non-parametric test for testing the null hypothesis of no difference in hazard rates, and thus no difference in the corresponding survival curves. In this subsection, formulas for the comparison of two hazard rates are provided. An extension for more than two groups can be found in Hosmer et al. (2008, Ch. 2, p.51ff). Denote the two groups under comparison

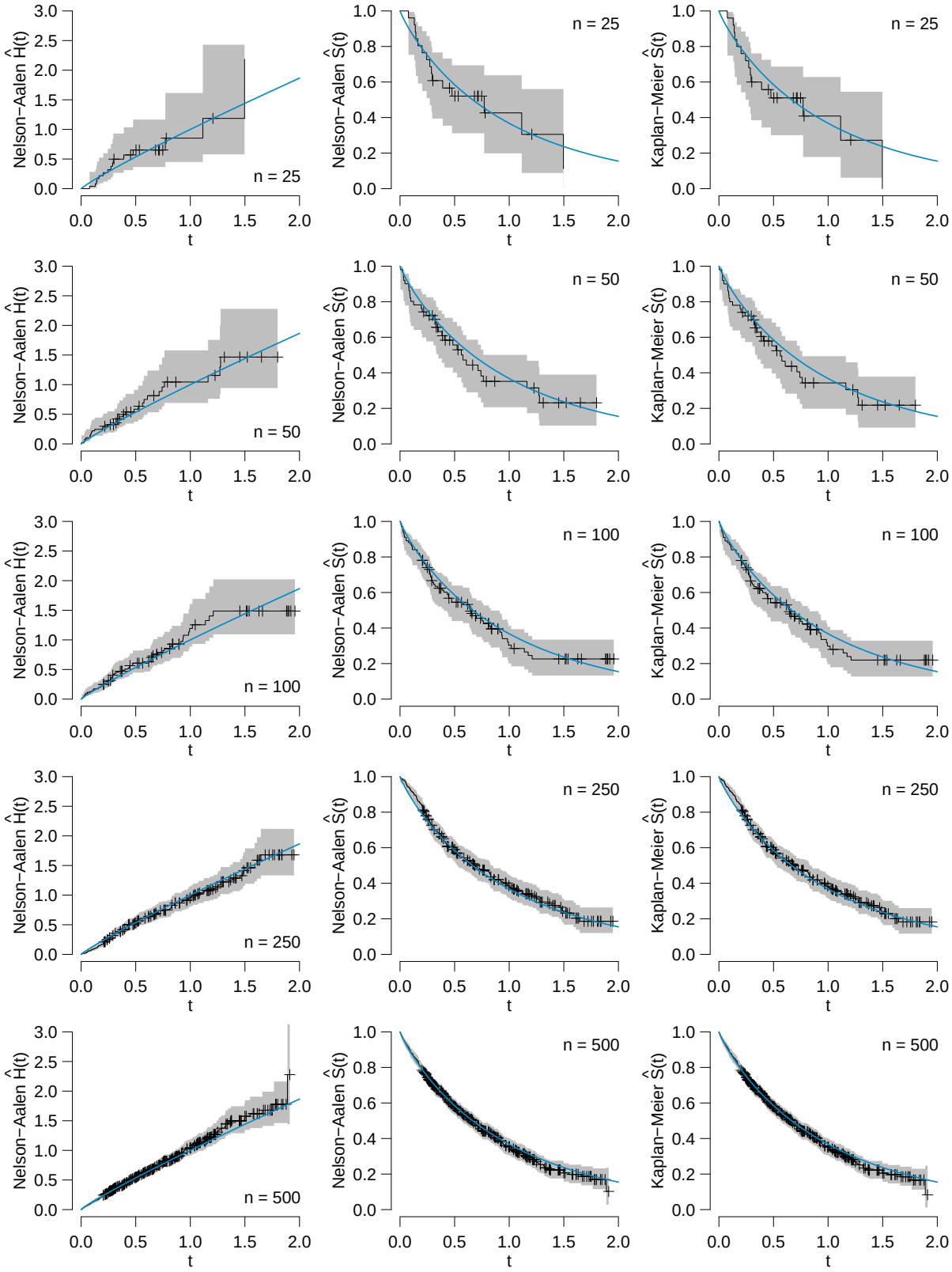

Figure 2: Example of Nelson-Aalen and Kaplan-Meier estimates for varying sample sizes. The left and middle column show Nelson-Aalen estimates of $H(t)$ and $S(t)$, respectively. The right column shows Kaplan-Meier estimates of $S(t)$. Times to event are sampled from a Weibull distribution with $\lambda = 1$ and $\alpha = 0.9$, and times to censoring from a uniform distribution $U(0.2, 2)$. Censoring times are marked with a +-symbol. The true Weibull $H(t)$ and $S(t)$ are shown in blue.

by 0 and 1. The null and alternative hypotheses are formally written

$$\begin{aligned} \mathrm{H}_0 &: h_0(t) = h_1(t), \text{ for all } t > 0 \\ \mathrm{H}_1 &: h_0(t) \neq h_1(t), \text{ for some } t > 0, \end{aligned}$$

where $h_i(t)$ denotes the hazard function of group $i, i = 0, 1$. A general test statistic is given by

$$Q = \frac{\left(\sum_{i=1}^{m} w_i[d_{1,i} - Y_1(t_i)d_i/Y(t_i)]\right)^2}{\sum_{i=1}^{m} w_i^2 \hat{v}_{1,i}},$$

where m denotes the number of observed events, d_i the observed number of events at event time t_i, $d_{1,i}$ the observed number of events in group 1 at time t_i, $Y(t_i)$ the number at risk at t_i^-, and $Y_1(t_i)$ the corresponding number at risk in group 1, w_i denotes weights, and $\hat{v}_{1,i}$ the variance estimate of $D_{1,i}$ (random version of $d_{1,i}$) at time t_i under the null. Note that $Y_1(t_i)d_i/Y(t_i)$ is the expected number of events in group 1 at time t_i under the null. The variance estimate, based on the hypergeometric distribution, is given by

$$\hat{v}_{1,i} = \frac{Y_1(t_i)Y_0(t_i)d_i[Y(t_i) - d_i]}{Y(t_i)^2[Y(t_i) - 1]}.$$

Under the null and under the assumption of similar censoring patterns in the different groups, Q is asymptotically chi-square distributed with one degree of freedom. There are a number of options for the weights, see Hosmer et al. (2008, Ch. 2.4) for some examples, and Aalen et al. (2008, Ch. 3.3.1) for some examples of weights using a counting process approach. Worth mentioning are the generalized Wilcoxon test and the Tarone-Ware test, with weights $w_i = Y(t_i)$ and $w_i = \sqrt{Y(t_i)}$, respectively, that hence put more emphasis on differences in hazard rates in the beginning of the time scale, where there are still more individuals in the risk set. Note that possible ties are already taken care of in the expressions of Q and $\hat{v}_{1,i}$, using the discrete-time approach. A continuous-time approach is not as simple when it comes to comparisons between groups, since the order of the tied events is not possible to separate between groups (Aalen et al., 2008, Ch. 3, p. 112).

The most common non-parametric test for differences in hazard rates is the log-rank test, where the test statistic is given by letting $w_i = 1, i = 1, ..., m$, yielding

$$Q_{l-r} = \frac{\left(\sum_{i=1}^{m} [d_{1,i} - Y_1(t_i)d_i/Y(t_i)]\right)^2}{\sum_{i=1}^{m} \hat{v}_{1,i}}. \tag{7}$$

In the case of crossing survival curves, the terms of the sum in Eq. (7) will first be negative (positive), and then positive (negative), resulting in a small test statistic, not detecting any difference between the survival curves at standard significance levels. It is therefore of value to inspect the survival curves visually as a complement to the hypothesis testing. For a discussion on non-parametric methods for crossing survival curves, see Li, Han, Hou, Chen, and Chen (2015). An example of two Kaplan-Meier plots comparing survival curves is given in Figure 3. The Log-rank test is able to detect a difference between the curves in the left plot, but not in the right where the curves are crossing.

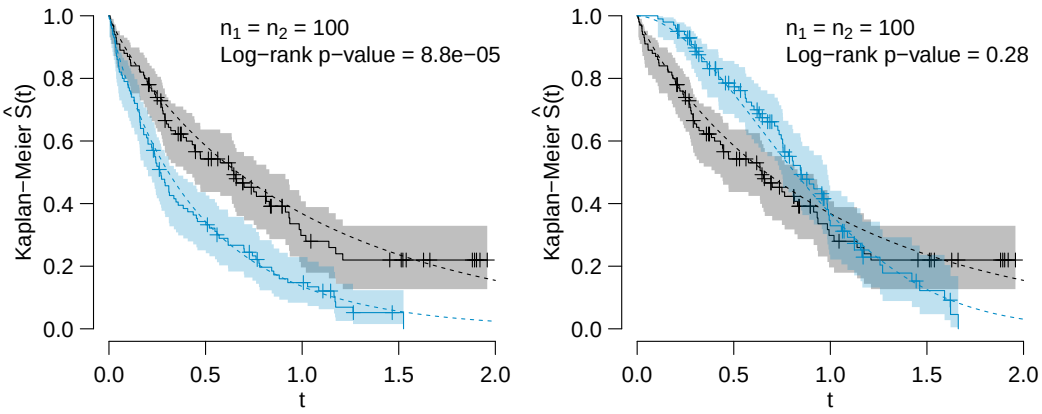

Figure 3: Example of Kaplan-Meier estimates for two groups, $n = 100$ in each group. Times to event are sampled from a Weibull distribution with $\lambda = 1$, $\alpha = 0.9$ (black line, both plots), and $\lambda = 2$, $\alpha = 0.9$ (blue line, left plot), and $\lambda = 1$, $\alpha = 1.8$ (blue line, right plot). Times to censoring are sampled from a uniform distribution $U(0.2, 2)$. Censorings are marked with a +-symbol. Shaded areas show 95% confidence intervals. Dashed lines show true Weibull survival curves.

3.3 Likelihood function for right-censored data

Although the non-parametric estimators presented above have desirable qualities for some applications, they are not regression methods. Covariates can only be taken into account by stratification, which is inconvenient even with a small number of covariates, and continuous covariates can only be incorporated by categorization. The remainder of this thesis will focus on regression methods, which, with the exception of quantile regression presented in Section 3.7, are likelihood-based. The general likelihood function for right-censored data is presented in this section, and later applied to the regression models presented in Sections 3.4, 3.5, and 3.6.

Assuming no left-censoring or truncation, each observation has a triplet of observed values $(y_i, \delta_i, \boldsymbol{x}_i)$, where y_i is the time until event or censoring, $\delta_i = I(t < c)$ is an indicator for whether the event of interest occurred before censoring, and $\boldsymbol{x}_i$ is the covariate vector, $i = 1, ..., n$. The likelihood for right-censored survival data is special due to the fact that the exact event times are not known for all observations. Censored observations contribute to the likelihood with their observed survival time, while uncensored observations contribute with their time to event. Each observation will hence contribute to the likelihood with $f(y_i)^{\delta_i} S(y_i)^{1-\delta_i}$. Under the assumption of independent observations, and that the censoring distribution is independent of the time-to-event distribution, the likelihood function is given by

$$L(\boldsymbol{\beta}|\boldsymbol{x}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i)^{\delta_i} S(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i)^{1-\delta_i}, \tag{8}$$

and the log-likelihood is given by (Hosmer et al., 2008, Ch. 3.3)

$$l(\boldsymbol{\beta}|\boldsymbol{x}) = \sum_{i=1}^{n} \left\{ \delta_i \log[f(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i)] + (1 - \delta_i) \log[S(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i)] \right\}.$$

From Eq. (1), $f(y_i|\boldsymbol{\beta}, \boldsymbol{x_i}) = h(y_i|\boldsymbol{\beta}, \boldsymbol{x_i}) S(y_i|\boldsymbol{\beta}, \boldsymbol{x_i})$. Plugging this into Eq. (8) gives

$$L(\boldsymbol{\beta}|\boldsymbol{x}) = \prod_{i=1}^{n} h(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i)^{\delta_i} S(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i)^{\delta_i} S(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i)^{1-\delta_i} = \prod_{i=1}^{n} h(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i)^{\delta_i} S(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i), \tag{9}$$

and

$$l(\boldsymbol{\beta}|\boldsymbol{x}) = \sum_{i=1}^{n} \{\delta_i \log[h(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i)] + \log[S(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i)]\}\,. \tag{10}$$

3.4 Cox proportional hazards model

The Cox proportional hazards model is a semi-parametric model that was first introduced in an article published in 1972 by the British statistician Sir David Cox (Cox, 1972). It is widely used due to the simplicity of the interpretation of the regression coefficients, and because no distribution assumption is needed. The model can be extended to incorporate time-dependent coefficients and covariates, but the focus of this thesis will be on fixed coefficients and covariates. Formulas in this section are taken from Chapters 3 and 6 in Hosmer et al. (2008), unless stated otherwise.

3.4.1 Model formulation

Let $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ be a column vector of coefficients, and $\boldsymbol{x} = (x_1, ..., x_p)'$ be a column vector of covariates. The model is given by

$$h(t|\boldsymbol{x}) = h_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}) = h_0(t)\exp\left(\sum_{k=1}^{p}\beta_k x_k\right), \tag{11}$$

where $h_0(t)$ is a baseline hazard function that is seldom of primary interest when fitting the model. Note that the baseline hazard does not depend on the covariates, only on time t. No parametric assumption is made about the baseline hazard. Also note that the second factor does not depend on time t, and that there is no intercept in the second factor. Instead, the "intercept" that corresponds to the value of the hazard function when all covariates are zero, is given by $h_0(t)$. Taking the logarithm of both sides yields linearity in the coefficients,

$$\log[h(t|\boldsymbol{x})] = \log[h_0(t)] + \boldsymbol{\beta}'\boldsymbol{x} = \log[h_0(t)] + \sum_{k=1}^{p}\beta_k x_k. \tag{12}$$

Let e_k be a column vector of length p, with a 1 in the kth position, and zeros in all other positions, i.e. $e_1 = (1, 0, ..., 0)'$. A one-unit change in x_k yields

$$\begin{aligned}
&\log[h(t|\boldsymbol{x}+e_k)] - \log[h(t|\boldsymbol{x})] \\
&= \big[\log[h_0(t)] + \boldsymbol{\beta}'(\boldsymbol{x}+e_k)\big] - \big[\log[h_0(t)] + \boldsymbol{\beta}'\boldsymbol{x}\big] \\
&= \beta_k.
\end{aligned}$$

Exponentiating both sides gives

$$\frac{h(t|\boldsymbol{x}+e_k)}{h(t|\boldsymbol{x})} = \exp(\beta_k).$$

This is called the hazard ratio (HR), and is the relative change in hazard corresponding to a one-unit increase in x_k, given that the other covariates are held constant. This ratio does not depend on t, implicating that the hazards are proportional everywhere on the time scale. This is an attractive property, simplifying the interpretation of the regression coefficients enormously.

This property is called the Proportional Hazards Assumption (PHA), and is not always valid. Deviations from it can be tested using the Schoenfeld residuals, introduced in Section 3.4.3.

3.4.2 Estimation and inference

Before deriving the likelihood function, the survival function is derived from the hazard function. From Eq. (4) it holds that,

$$S(t|\boldsymbol{x}) = \exp\left\{-H(t|\boldsymbol{x})\right\} = \exp\left\{-\int_0^t h(s|\boldsymbol{x})ds\right\}.$$

Plugging in the expression for $h(t|\boldsymbol{x})$ from Eq. (11), and noting that $\exp(\boldsymbol{\beta}'\boldsymbol{x})$ is independent of t, yields

$$S(t|\boldsymbol{x}) = \exp\left\{-\int_0^t h_0(s)\exp(\boldsymbol{\beta}'\boldsymbol{x})ds\right\} = \left[\exp\left\{-\int_0^t h_0(s)ds\right\}\right]^{\exp(\boldsymbol{\beta}'\boldsymbol{x})} = S_0(t)^{\exp(\boldsymbol{\beta}'\boldsymbol{x})}, \quad (13)$$

where $S_0(t) = \exp(-\int_0^t h_0(s)ds)$ denotes the baseline survival function, which is independent of $\boldsymbol{\beta}$ and $\boldsymbol{x}$.

Recall from Eq. (10) that the general log-likelihood can be written

$$l(\boldsymbol{\beta}|\boldsymbol{x}) = \sum_{i=1}^{n}\left(\delta_i \log[h(y_i|\boldsymbol{\beta},\boldsymbol{x}_i)] + \log[S(y_i|\boldsymbol{\beta},\boldsymbol{x}_i)]\right).$$

Plugging in the expression for $\log[h(t|\boldsymbol{x})]$ and $S(t|\boldsymbol{x})$ from Eq.s (12) and (13) yields

$$l(\boldsymbol{\beta}|\boldsymbol{x}) = \sum_{i=1}^{n}\left(\delta_i(\log[h_0(y_i)] + \boldsymbol{\beta}'\boldsymbol{x}_i) + \exp(\boldsymbol{\beta}'\boldsymbol{x}_i)\log[S_0(y_i)]\right). \quad (14)$$

However, the log-likelihood in Eq. (14) cannot be used directly to estimate $\boldsymbol{\beta}$. Instead, maximization of a partial likelihood, presented by Cox (1972), is used for estimation of the coefficient vector $\boldsymbol{\beta}$, where $h_0(t)$ and $S_0(t)$ are considered to be nuisance parameters. The partial likelihood, not to be confused with the partial likelihood usually referred to in statistical inference theory, is derived by conditioning on the risk set $\mathscr{R}(y_i)$, and is given by

$$L_p(\boldsymbol{\beta}|\boldsymbol{x}) = \prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}{\sum_{j\in\mathscr{R}(y_i)}\exp(\boldsymbol{\beta}'\boldsymbol{x}_j)}\right]^{\delta_i},$$

When $\delta_i = 0$, the only contribution to the above product is 1, and, hence, the expression can be simplified by only considering the $m \leq n$ observed events, writing

$$L_p(\boldsymbol{\beta}|\boldsymbol{x}) = \prod_{i=1}^{m}\frac{\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}{\sum_{j\in\mathscr{R}(t_i)}\exp(\boldsymbol{\beta}'\boldsymbol{x}_j)}.$$

The partial log-likelihood is given by

$$l_p(\boldsymbol{\beta}|\boldsymbol{x}) = \sum_{i=1}^{m}\left[\boldsymbol{\beta}'\boldsymbol{x}_i - \log\left(\sum_{j\in\mathscr{R}(t_i)}\exp(\boldsymbol{\beta}'\boldsymbol{x}_j)\right)\right].$$

The score function U_k for a specific $\beta_k, k = 1, ..., p$, is given by

$$U_k(\boldsymbol{\beta}) = \frac{\partial l_p(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{i=1}^{m} \left[x_{i,k} - \frac{\sum_{j \in \mathscr{R}(t_i)} x_{j,k} \exp(\boldsymbol{\beta}' \boldsymbol{x}_j)}{\sum_{j \in \mathscr{R}(t_i)} \exp(\boldsymbol{\beta}' \boldsymbol{x}_j)} \right]. \tag{15}$$

The partial log-likelihood can be maximized using Fisher scoring. Andersen and Gill (1982) showed the asymptotic normality and consistency of $\hat{\boldsymbol{\beta}}$, using a counting process approach. In a model containing only one binary covariate, the score test statistic is equal to the log-rank test statistic presented in Section 3.2.3. The covariance matrix $\text{Cov}(\hat{\boldsymbol{\beta}})$ is estimated using the observed information matrix.

When tied event times are present, the likelihood can be rewritten according to the Breslow or Efron methods, described in (Klein & Moeschberger, 2003, Ch. 8.4), not stated explicitly here. When few ties are present, these two methods lead to quite similar results.

However rarely of primary interest when using a Cox model, the baseline cumulative hazard can be estimated by using the Breslow estimator, described in (Klein & Moeschberger, 2003, Ch. 8), defined as

$$\hat{H}_0(t) = \sum_{t_i \leq t} \frac{d_i}{\sum_{j \in \mathscr{R}(t_i)} \exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_j)}, \tag{16}$$

and the corresponding estimation of the baseline survival function is given by

$$\hat{S}_0(t) = \exp\{-\hat{H}_0(t)\}.$$

For an individual with covariate vector $\boldsymbol{x}_i$, the estimated cumulative hazard function and survival function are given by

$$\hat{H}(t|\boldsymbol{x}_i) = \hat{H}_0(t) \exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_i),$$

and

$$\hat{S}(t|\boldsymbol{x}_i) = \hat{S}_0(t)^{\exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_i)}.$$

The Breslow estimator of the cumulative hazard function is used in some of the residuals presented in Section 3.4.3 below. Survival estimates from two Cox models are shown in Figure 4, using the simulated survival data from Figure 3. The left plot describes the survival data well, since the hazards are proportional, while the right plot cannot capture the crossing of the survival curves in the original data.

3.4.3 Model diagnostics

In Cox regression, unlike other regression methods, the definition of residuals is not obvious. This is partly because the outcome variable is often incomplete, and partly because the outcome variable is not modeled directly for example through a function of the mean. In 1982, Schoenfeld (1982) proposed a definition of residuals corresponding to an observation's addition to the derivative of the partial log-likelihood. Plugging in $\hat{\boldsymbol{\beta}}$ to the expression of the score function in Eq. (15) yields the Schoenfeld residual for observation i, $i = 1, ..., n$ and covariate k,

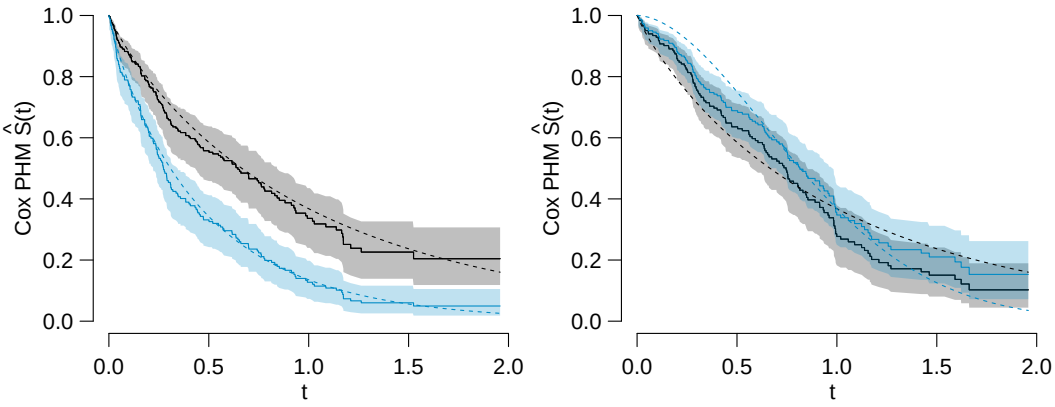

Figure 4: Example of Cox proportional hazrads (PHM) survival estimates for a model including a binary predictor indicating group, using data from Figure 3. Times to event in the left plot are sampled from Weib($\lambda = 1, \alpha = 0.9$) and Weib($\lambda = 2, \alpha = 0.9$). Times to event in the right plot are sampled from Weib($\lambda = 1, \alpha = 0.9$) and Weib($\lambda = 1, \alpha = 1.8$). Times to censoring are sampled from a uniform distribution $U(0.2,\ 2)$. Shaded areas show 95% confidence intervals. Dashed lines show true Weibull survival curves.

$k = 1, ..., p$,

$$\hat{r}_{i,k} = \begin{cases} x_{i,k} - \dfrac{\sum_{j \in \mathscr{R}(t_i)} x_{j,k} \exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_j)}{\sum_{j \in \mathscr{R}(t_i)} \exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_j)}, & \text{uncensored observation} \\ 0, & \text{censored observation.} \end{cases} \tag{17}$$

Thus, each uncensored observation generates p Schoenfeld residuals $\hat{\boldsymbol{r}}_i' = (\hat{r}_{i,1}, \hat{r}_{i,2}, ..., \hat{r}_{i,p})'$, one for each covariate. Grambsch and Therneau (1994) proposed scaling the Schoenfeld residuals using their variance according to

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

where $[\widehat{\text{Var}}(\hat{\boldsymbol{r}}_i)]^{-1}$ is approximated by $m\widehat{\text{Var}}(\hat{\boldsymbol{\beta}})$, with m being the number of events, where the inverse of the Fisher information matrix is used for estimation of $\text{Var}(\hat{\boldsymbol{\beta}})$. The PHA can be evaluated by testing the null hypothesis of zero Pearson correlation between (a transformation of) event times and the scaled Schoenfeld residuals. This test corresponds to testing

$$H_0\text{: Proportional hazards}$$
$$H_1\text{: Non-proportional hazards.}$$

Since the Pearson correlation is linear, it is important to visually inspect the plot of the scaled Schoenfeld residuals against the (possibly transformed) event times.

In Figure 5, the scaled Schoenfeld residuals of Cox PH models fitted to the data from Figure 3 are shown. The model is given by $h(t|x) = h_0(t)\exp(\beta x)$, where $x = 1$ for the blue group, and $x = 0$ for the black group. In the left plot, there is no statistically significant deviation from the PHA, while in the right plot, with data forming crossing Kaplan-Meier curves, the PHA is rejected at all standard significance levels.

If the PHA does not hold, time-varying coefficients can be incorporated, but the functional form of the relationship between the coefficient $\beta_k(t)$ and time t must be assumed, for example $\beta_k(t) = a + b\log(t)$ or $\beta_k(t) = a + bt + ct^2$, yielding $\beta_k x_k = ax_k + b\log(t)x_k$ and $\beta_k(t) = ax_k + btx_k + ct^2x_k$. The Cox model can also be extended to incorporate time-varying covariates,

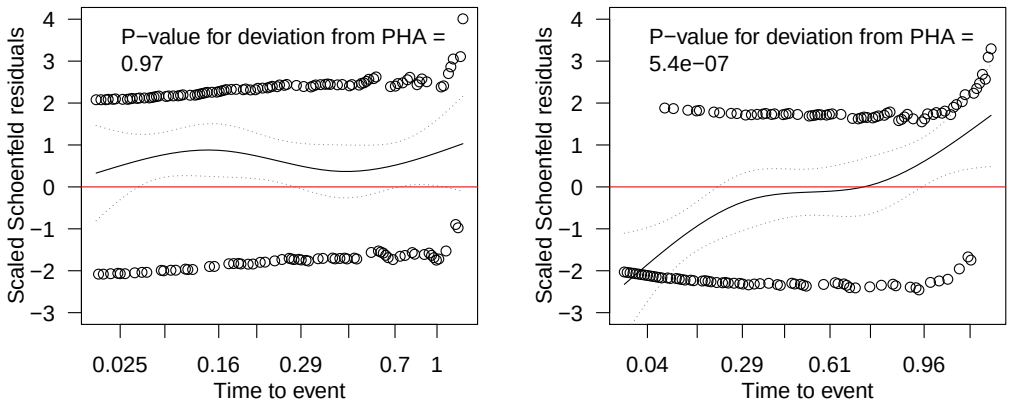

Figure 5: Scaled Schoenfeld residuals from a Cox PH model with two groups, using data from Figure 3, plotted against rank-ordered event times. The blue group is coded as 1 and the black group is coded as 0. A smooth curve for the association between (rank-ordered) time and the Scaled Schoenfeld residuals was added for illustrative purposes.

though this is not the scope of this thesis, and the interested reader is referred to Klein and Moeschberger (2003, Ch. 9).

Other types of residuals are the Cox-Snell, Martingale, and Deviance residuals, all described in (Klein & Moeschberger, 2003, Ch. 11). The Cox-Snell residuals are defined as

$$r_{CS,i} = \hat{H}(t_i|\boldsymbol{x}), \tag{18}$$

with $\hat{H}(t_i|\boldsymbol{x})$ estimated from the model as

$$r_{CS,i} = \hat{H}_0(t_i)\exp(\boldsymbol{\beta}'\boldsymbol{x}),$$

where $\hat{H}_0(t_i)$ is the Breslow estimator defined in Eq. (16). Note that the Cox-Snell residuals are defined for uncensored as well as censored observations, unlike the Schoenfeld residuals in Eq. (17). Plotted against $\hat{H}_{NA}$ of Eq. (5) of the residuals (using the residuals as the time variable), these should form a straight line through the origin with slope 1.

The Martingale residuals are defined as

$$\hat{r}_{M,i} = \delta_i - \hat{H}_0(t_i)\hat{H}(t_i|\boldsymbol{x}) = \delta_i - r_{CS,i}. \tag{19}$$

If the true $H_0(\cdot)$ and $\boldsymbol{\beta}$ were to be used, the Martingale residuals would be mean-zero martingales, which explains the name. To assess the functional form of a continuous covariate, the Martingale residuals of a Cox PHM containing all other covariates can be plotted against the covariate values. The highest possible value of the Martingale residuals is 1, while the lowest is $-\infty$. Therefore, these residuals are highly skewed.

To check overall model fit, and check for outliers, the Deviance residuals can be used, defined as

$$r_{D,i} = \text{sign}(\hat{r}_{M,i})\sqrt{-2[\hat{r}_{M,i} + \delta_i \log(\delta_i - \hat{r}_{M,i})]}. \tag{20}$$

These residuals should have an approximate normal distribution if the model fits. The deviance residuals can be plotted against the individual $\exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_i)$s to detect outliers. They can also,

similarly to the Martingale residuals, be used to assess the functional shape of a continuous covariate. With a high proportion censored observations, a high proportion of the Martingale and deviance residuals will be close to zero.

3.5 Parametric survival models

Even though the non-parametric and semi-parametric methods of Sections 3.2 and 3.4 are popular, much can be gained from a parametric model, given that the model assumptions are fulfilled. In fact, Sir David Cox himself, when asked by Nancy Reid (1994) about the popularity of his model, have said that he would prefer a parametric approach rather than a semi-parametric one.

There are several types of parametric survival models. The survival time can be modeled directly, but also the hazard function. In Section 3.5.1, some common parametric distributions used in survival analysis are introduced to the reader, followed by the model formulations of a few parametric survival models in Sections 3.5.2, 3.5.3, and 3.5.4, with a focus on the Weibull and log-logistic distributions.

3.5.1 Common parametric distributions

An informative table including a number of more or less common parametric distributions used in survival analysis is given by Klein and Moeschberger (2003, Ch. 2.5). A few of them are introduced in Table 2, and two additional ones are introduced in Table A1 and Figures A1 and A2 of Appendix A. The density function, survival function, and hazard function of the parametric distributions in Table 2 are shown in Figures 6, 7, 8, 9, 10, for varying parameter values.

Table 2: Common parametric distributions in survival analysis

Distribution	**Parameters**	$\boldsymbol{f(t)}$	$\boldsymbol{S(t)}$	$\boldsymbol{h(t)}$
Exponential $T \sim \text{Exp}(\lambda)$	$\lambda > 0$	$\lambda e^{-\lambda t}$	$e^{-\lambda t}$	λ
Weibull $T \sim \text{Weib}(\alpha, \lambda)$	$\alpha, \lambda > 0$	$\alpha\lambda t^{\alpha-1} e^{-\lambda t^\alpha}$	$e^{-\lambda t^\alpha}$	$\alpha\lambda t^{\alpha-1}$
Log-normal $T \sim \text{LN}(\mu, \sigma^2)$	$\mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{t\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\log t - \mu)^2}{\sigma^2}\right)$	$1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)(*)$	$\frac{f(t)}{S(t)}$
Log-logistic $T \sim \text{LL}(\alpha, \lambda)$	$\alpha, \lambda > 0$	$\frac{\alpha\lambda t^{\alpha-1}}{(1+\lambda t^\alpha)^2}$	$\frac{1}{1+\lambda t^\alpha}$	$\frac{\alpha\lambda t^{\alpha-1}}{1+\lambda t^\alpha}$
Generalized Gamma $T \sim \text{G}\Gamma(\alpha, \beta, \lambda)$	$\alpha, \beta, \lambda > 0$	$\frac{\alpha\lambda^\beta t^{\alpha\beta-1} \exp(-\lambda t^\alpha)}{\Gamma(\beta)}(**)$	$1 - \int_0^t f(s)ds$	$\frac{f(t)}{S(t)}$

*) $\Phi(\cdot)$ CDF for $N(0,1)$
**) $\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x)dx$. If $n \in \mathbb{N}, \Gamma(n) = (n-1)!$

The exponential distribution is memoryless, seen by the property (let $s, t > 0$)

$$P(T > s + t | T > s) = \frac{P(T > s + t, T > s)}{P(T > s)} = \frac{P(T > s + t)}{P(T > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t),$$

so T has the same distribution regardless of how long the observed survival time is. This means that the expected residual survival time is the same, even if the subject has survived for a long time. This can also be seen by the fact that the hazard function, which at time t is the instantaneous probability of having an event at time $t + dt$, given that the subject has survived at least until time t, is constant. Although an attractive mathematical property, this can be

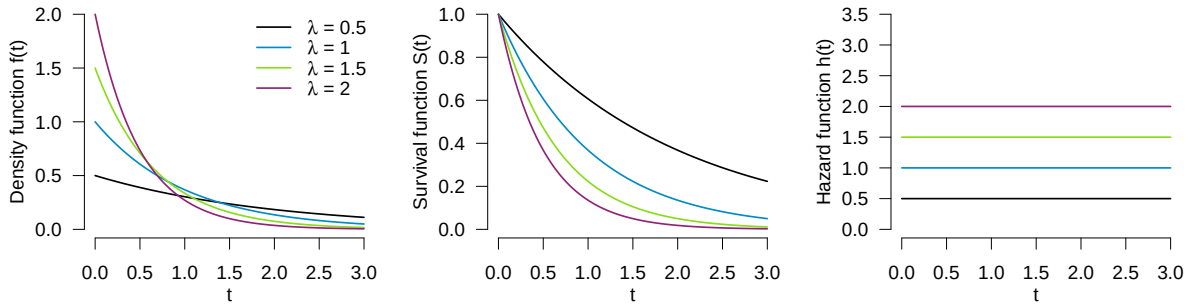

Figure 6: The Exponential distribution; density function, survival function and hazard function, for varying rate parameters λ.

more or less plausible depending on the context. It is unlikely that an 80 year old person and a 2 year old person have the same expected residual survival time.

The common way of parameterizing the Weibull distribution is by $f(t) = \frac{a}{b}\left(\frac{t}{b}\right)^{a-1} e^{-(t/b)^a}$. Table 2 presents an alternative parametrization, with $\alpha = a, \lambda = b^{-a}$, which is more suitable for proportional hazards modeling. As seen in Figure 7, for shape parameter $\alpha < 1$, the hazard function is strictly decreasing, and approaches zero as t tends to infinity, while for $\alpha > 1$ it is strictly increasing, and tends to infinity as t goes to infinity. This flexibility makes the Weibull distribution an attractive distribution in survival analysis. Note that $\alpha = 1$ produces an exponential distribution with mean $1/\lambda$. This is also visible in Figure 7, where a constant hazard function of 1 is seen for $\alpha = 1$.

The log-normal distribution is an appealing candidate due to the close connection to the well-known normal distribution. If the random variable $X \sim \text{LN}(\mu, \sigma^2)$, then $Y = \log(X) \sim N(\mu, \sigma^2)$. In Figure 8 it can be seen that hazard function $h(t)$ is first increasing and then decreasing, a property that has been deemed unlikely for many survival data applications. Recall that the hazard is the instantaneous risk of having the event at time t, given that the survival time is at least t^-. If a subject has survived a long time, it is likely that the event will occur. Though the distribution can be appropriate when the beginning of the time frame is of greater interest than later times (Klein & Moeschberger, 2003, Ch. 2.5, p. 40). For example, in childhood cancer, children not dying from the disease or treatment tend to die from other causes much later in life, and, hence, their hazard function could very well be decreasing after the first years following diagnosis.

If a random variable Y follows the logistic distribution, which resembles the normal distribution, then $X = \exp(Y)$ follows the log-logistic distribution. Similar to the log-normal hazard, the hazard function of the log-logistic distribution approaches zero as t goes to infinity (Klein & Moeschberger, 2003, Ch. 2.5, p. 41). The distribution is visualized in Figure 9.

The Generalized Gamma distribution is practical when it comes to choosing a parametric distribution and for model checking, since the exponential, Weibull, and Gamma (see Appendix A) distributions are all special cases of this distribution. An exponential distribution with mean $1/\lambda$ is produced when $\alpha = \beta = 1$, a Weibull distribution with parameters α and λ is produced when $\beta = 1$, and a Gamma distribution with parameters β and λ is produced when $\alpha = 1$. Also, when β approaches infinity, the generalized gamma distribution approaches the log-normal distribution (Klein & Moeschberger, 2003, Ch. 2.5, p. 44). The distribution is visualized in Figure 10.

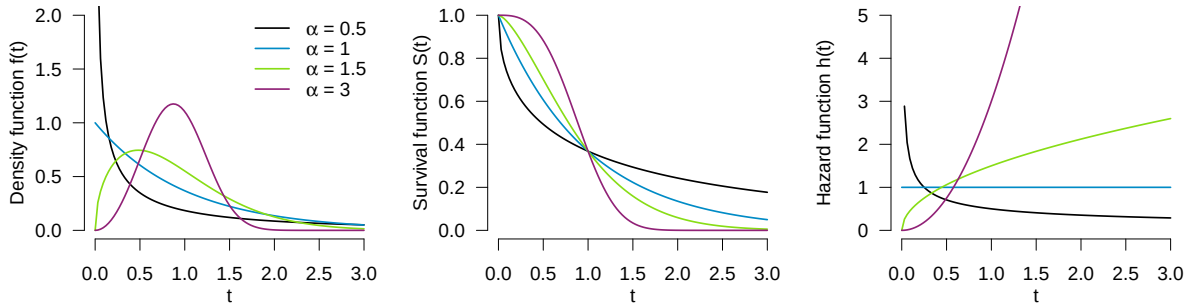

Figure 7: The Weibull distribution; density function, survival function and hazard function, for varying shape parameters α. The scale parameter $\lambda = 1$ for all α.

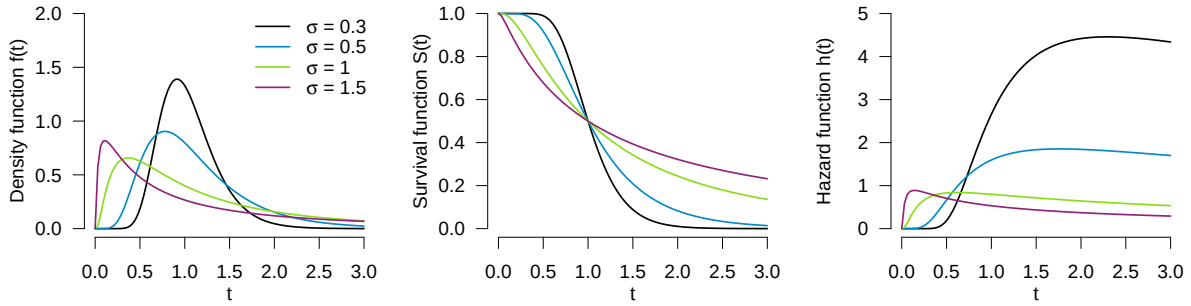

Figure 8: The Log-Normal distribution; density function, survival function and hazard function, for varying parameters, σ. $\mu = 0$ for all σ

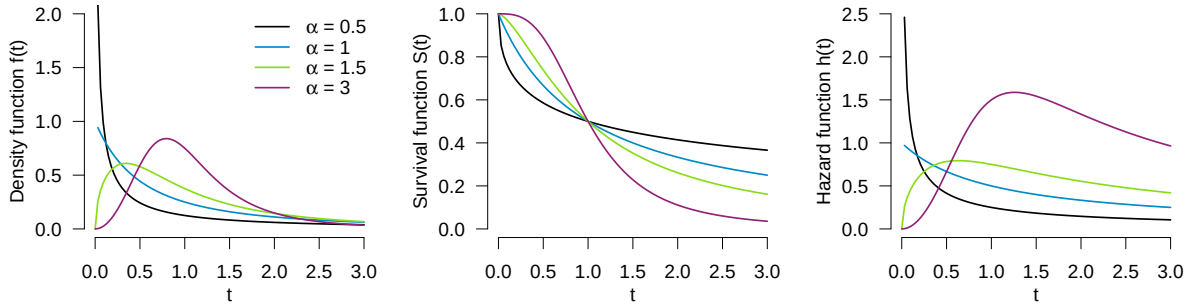

Figure 9: The Log-Logistic distribution; density function, survival function and hazard function, for varying parameters, α. $\lambda = 1$ for all α.

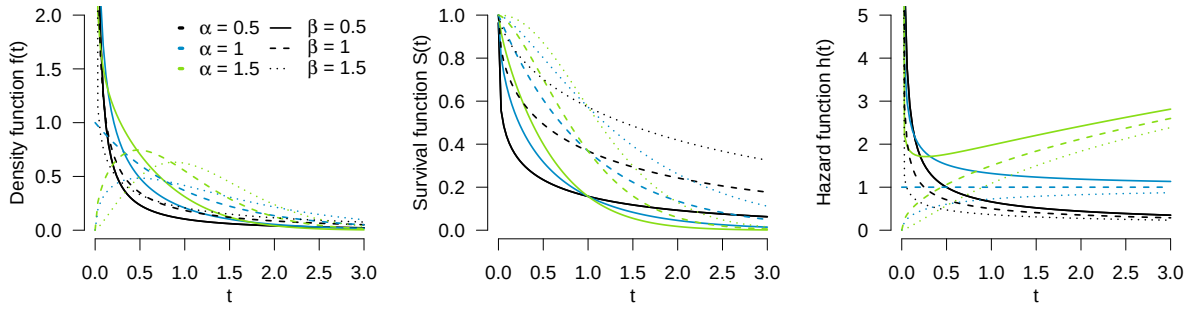

Figure 10: The Generalized Gamma distribution; density function, survival function and hazard function, for varying parameters, α, β. $\lambda = 1$ for all α, β.

3.5.2 Accelerated failure time model

This section follows Klein and Moeschberger (2003, Ch. 2 & 12). One approach to parametric regression modeling for survival data is to model the natural logarithm of the time-to-event variable T, with $V = \log(T)$, conditional on a number of covariates $\boldsymbol{x}$ in analogy with linear regression, as follows

$$V = \mu + \boldsymbol{\omega}'\boldsymbol{x} + \sigma W, \tag{21}$$

where $\boldsymbol{\omega} = (\omega_1, ..., \omega_p)'$ is a coefficient vector, and W follows some error distribution. Let $S_0(t)$ be the baseline survival function of T when $\boldsymbol{x} = \mathbf{0}$. The survival function of T given covariates $\boldsymbol{x}$ is derived as follows,

$$\begin{aligned} S(t|\boldsymbol{x}) &= P(T > t|\boldsymbol{x}) = P(V > \log(t)|\boldsymbol{x}) = P(\mu + \sigma W > \log(t) - \boldsymbol{\omega}'\boldsymbol{x}|\boldsymbol{x}) \\ &= P(\exp\{\mu + \sigma W\} > t\exp\{-\boldsymbol{\omega}'\boldsymbol{x}\}|\boldsymbol{x}) = S_0(t\exp\{-\boldsymbol{\omega}'\boldsymbol{x}\}). \end{aligned}$$

Due to $\boldsymbol{x}$ accelerating or degrading the baseline survival function, these types of parametric survival models are called *accelerated failure time* (AFT) models.

If W has a standard extreme value distribution with density

$$f_W(w) = \exp\{w - e^w\}, -\infty < w < \infty,$$

then T follows a Weibull distribution with the shape and scale parameters α and λ expressed in terms of μ and σ from Eq. (23) as $\alpha = 1/\sigma$ and $\lambda = \exp(-\mu/\sigma)$. To see this, let $T \sim \text{Weib}(\alpha = 1/\sigma, \lambda = e^{-\mu/\sigma})$, with

$$F_T(t) = 1 - \exp\{-\exp(-\mu/\sigma)t^{1/\sigma}\},$$

and $V = \log(T) = \mu + \sigma W$, so $W = [\log(T) - \mu]/\sigma$. Now,

$$\begin{aligned} F_W(w) &= P(W \leq w) = P([\log(T) - \mu]/\sigma \leq w) = P(T \leq \exp\{\mu + \sigma w\}) \\ &= F_T(\exp\{\mu + \sigma w\}) = 1 - \exp\left(-\exp\{-\mu/\sigma\}\exp\{\mu + \sigma w\}^{1/\sigma}\right) \\ &= 1 - \exp(-e^w), \end{aligned}$$

which is the CDF of a standard extreme value distribution. Using a Weibull distribution for T, the baseline survival function is given by

$$S_0(t) = \exp(-\lambda t^\alpha),$$

which yields the AFT model

$$S(t|\boldsymbol{x}) = \exp(-\lambda[t\exp(-\boldsymbol{\omega}'\boldsymbol{x})]^\alpha). \tag{22}$$

If W has a standard logistic distribution with density

$$f_W(w) = e^w/(1 + e^w)^2, -\infty < w < \infty,$$

then T follows a log-logistic distribution with $\alpha = 1/\sigma$ and $\lambda = \exp(-\mu/\sigma)$, similar to the Weibull distribution. To see this, let $T \sim \text{LL}(\alpha = 1/\sigma, \lambda = \exp\{-\mu/\sigma\}$, with

$$F_T(t) = 1 - \frac{1}{1 + \exp(-\mu/\sigma)t^{\frac{1}{\sigma}}}.$$

Similarly to the Weibull distribution, $F_W(w)$ is given by

$$F_W(w) = F_T(\exp\{\mu + \sigma w\}) = 1 - \frac{1}{1 + \exp(-\mu/\sigma)\exp(\mu + \sigma w)^{\frac{1}{\sigma}}} = 1 - \frac{1}{1 + \exp(w)},$$

which is the CDF of a standard logistic distribution. For $T \sim \text{LL}(\alpha, \lambda)$, the baseline survival function is given by

$$S_0(t) = \frac{1}{1 + \lambda t^\alpha},$$

yielding the AFT model

$$S(t|\boldsymbol{x}) = \frac{1}{1 + \lambda[t\exp(-\boldsymbol{\omega}'\boldsymbol{x})]^\alpha}. \tag{23}$$

Other possible distributions for AFT models include the log-normal and generalized gamma distributions. The interested reader is referred to Klein and Moeschberger (2003, Ch. 12), where an overview of AFT models is provided.

3.5.3 Parametric proportional hazards model

Another way to parametrically model survival data is to use a proportional hazards model, also called relative risk model or multiplicative hazards model, modeling the hazard function as follows,

$$h(t|\boldsymbol{x}) = h_0(t)c(\boldsymbol{\beta}'\boldsymbol{x}),$$

where $c(\cdot)$ is any non-decreasing function. Comparing two hazards with covariate vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively, yields

$$\frac{h(t|\boldsymbol{x}_1)}{h(t|\boldsymbol{x}_2)} = \frac{h_0(t)c(\boldsymbol{\beta}'\boldsymbol{x}_1)}{h_0(t)c(\boldsymbol{\beta}'\boldsymbol{x}_2)} = \frac{c(\boldsymbol{\beta}'\boldsymbol{x}_1)}{c(\boldsymbol{\beta}'\boldsymbol{x}_2)},$$

which is constant over time t. Hence, this is a proportional hazards model (PHM). An intuitive choice for $c(\cdot)$ is, in analogy with the Cox PHM, $\exp(\cdot)$, yielding

$$h(t|\boldsymbol{x}) = h_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}).$$

Similar to the Cox PHM, a one-unit increase in x_k yields a hazard ratio of $\exp(\beta_k)$. Note that $h_0(t)$ is not the same $h_0(t)$ as in the Cox PHM, presented in Section 3.4. Instead, $h_0(t)$ is modeled parametrically. If $T \sim \text{Weib}(\alpha, \lambda)$, then $h_0(t) = \alpha\lambda t^{\alpha-1}$, which gives the Weibull PHM

$$h(t|\boldsymbol{x}) = \alpha\lambda t^{\alpha-1}\exp(\boldsymbol{\beta}'\boldsymbol{x}). \tag{24}$$

Recall from Eq. (2) that $H(t) = \int_0^t h(s)ds$, and from Eq. (4) that $S(t) = \exp\{-H(t)\}$. For a Weibull PHM, the cumulative hazard function is given by

$$H(t) = \int_0^t \alpha\lambda s^{\alpha-1} \exp(\boldsymbol{\beta}'\boldsymbol{x})ds = \lambda t^\alpha \exp(\boldsymbol{\beta}'\boldsymbol{x}),$$

and the survival function is given by

$$S(t) = \exp\{-\lambda t^\alpha \exp(\boldsymbol{\beta}'\boldsymbol{x})\} = \exp\left\{-\lambda\left[t\exp\left(\frac{1}{\alpha}\boldsymbol{\beta}'\boldsymbol{x}\right)\right]^\alpha\right\},$$

which is on the form of the AFT model in Eq. (22), with $-\boldsymbol{\omega}' = \frac{1}{\alpha}\boldsymbol{\beta}'$. The Weibull distribution is the only one that can be written both on the proportional hazards form and on the accelerated failure time form.

3.5.4 Parametric proportional odds model

Yet another parametric model for survival data is the proportional odds model (POM), given by

$$O(t|\boldsymbol{x}) = \frac{1 - S(t|\boldsymbol{x})}{S(t|\boldsymbol{x})} = \frac{1 - S_0(t)}{S_0(t)} \exp(\boldsymbol{\beta}'\boldsymbol{x}) = O_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}), \tag{25}$$

where $O_0(t)$ is the baseline odds of an event occurring before time t. A one-unit increase in x_k results in an odds ratio (OR) of $\exp(\beta_k)$. A natural choice of distribution for this model is the log-logistic distribution, yielding

$$O_0(t) = \frac{1 - S_0(t)}{S_0(t)} = \frac{1 - \frac{1}{1+\lambda t^\alpha}}{\frac{1}{1+\lambda t^\alpha}} = \lambda t^\alpha, \tag{26}$$

and therefore

$$O(t|\boldsymbol{x}) = \lambda t^\alpha \exp(\boldsymbol{\beta}'\boldsymbol{x}).$$

Using the AFT representation in Eq. (23), the POM can be written

$$O(t|\boldsymbol{x}) = \frac{1 - \frac{1}{1+\lambda[t\exp(-\boldsymbol{\omega}'\boldsymbol{x})]^\alpha}}{\frac{1}{1+\lambda[t\exp(-\boldsymbol{\omega}'\boldsymbol{x})]^\alpha}} = \lambda t^\alpha \exp(-\alpha\boldsymbol{\omega}'\boldsymbol{x}).$$

Hence, the coefficients of the POM can be calculated from the AFT model with $\boldsymbol{\beta} = -\alpha\boldsymbol{\omega}$. The log-logistic distribution is the only one that can be written on both the POM and AFT forms.

Another approach to parametric survival data modeling is to use additive hazard models. These will not be presented here, but the interested reader is referred to (Klein & Moeschberger, 2003, Ch. 10).

3.5.5 Estimation and inference

In a Weibull AFT model, where $V = \log(T) = \mu + \boldsymbol{\omega}'\boldsymbol{x} + \sigma W$, the CDF of V is derived as

$$\begin{aligned} F_V(v) &= P(V \leq v) = P(\log(T) \leq v) = P(T \leq \exp\{v\}) = F_T(\exp\{v\}) \\ &= 1 - \exp\left(-e^{-\frac{\mu}{\sigma}}\left[e^v e^{-\boldsymbol{\omega}'\boldsymbol{x}}\right]^{\frac{1}{\sigma}}\right) = 1 - \exp\left(-e^{\frac{v-\mu-\boldsymbol{\omega}'\boldsymbol{x}}{\sigma}}\right), \end{aligned}$$

which yields

$$S_V(v) = 1 - F_V(v) = \exp\left(-e^{\frac{v-\mu-\boldsymbol{\omega}'\boldsymbol{x}}{\sigma}}\right),$$

and

$$f_V(v) = \frac{dF_V(v)}{dv} = \frac{1}{\sigma}\exp\left(\frac{v-\mu-\boldsymbol{\omega}'\boldsymbol{x}}{\sigma} - e^{\frac{v-\mu-\boldsymbol{\omega}'\boldsymbol{x}}{\sigma}}\right).$$

From Eq. (8), the likelihood function is given by, letting $v_i = \log(y_i) = \log[\min(t_i, c_i)]$,

$$\begin{aligned} L(\mu, \sigma|\boldsymbol{x}) &= \prod_{i=1}^{n} f_V(v_i)^{\delta_i} S_V(v_i)^{1-\delta_i} \\ &= \prod_{i=1}^{n}\left[\frac{1}{\sigma}\exp\left(\frac{v_i-\mu-\boldsymbol{\omega}'\boldsymbol{x}_i}{\sigma} - e^{\frac{v_i-\mu-\boldsymbol{\omega}'\boldsymbol{x}_i}{\sigma}}\right)\right]^{\delta_i}\exp\left(-e^{\frac{v_i-\mu-\boldsymbol{\omega}'\boldsymbol{x}_i}{\sigma}}\right)^{1-\delta_i}. \end{aligned}$$

The maximum likelihood estimates of μ, σ and $\boldsymbol{\omega}$ are found numerically, and the covariance matrix is estimated by the observed Fisher information matrix. Estimates of $\boldsymbol{\beta}$ from the Weibull PHM in Eq. (24), as well as α and λ are derived as $\hat{\boldsymbol{\beta}} = -\hat{\boldsymbol{\omega}}/\hat{\sigma}$, $\hat{\alpha} = 1/\hat{\sigma}$ and $\hat{\lambda} = \exp(-\hat{\mu}/\hat{\sigma})$, respectively, and their variances are estimated using the Delta method, see Klein and Moeschberger (2003, Ch. 12) for a derivation.

In a log-logistic AFT model, the CDF of V is derived as

$$F_V(v) = F_T(\exp\{v\}) = 1 - \frac{1}{1+\exp\left(\frac{v-\mu}{\sigma}\right)},$$

which yields

$$S_V(v) = 1 - F_V(v) = \frac{1}{1+\exp\left(\frac{v-\mu}{\sigma}\right)},$$

and

$$f_V(v) = \frac{dF_V(v)}{dv} = \frac{1}{\sigma}\frac{\exp\left(\frac{v-\mu}{\sigma}\right)}{\left[1+\exp\left(\frac{v-\mu}{\sigma}\right)\right]^2}.$$

The likelihood function is given by

$$L(\mu, \sigma|\boldsymbol{x}) = \prod_{i=1}^{n}\left[\frac{1}{\sigma}\frac{\exp\left(\frac{v_i-\mu}{\sigma}\right)}{\left[1+\exp\left(\frac{v_i-\mu}{\sigma}\right)\right]^2}\right]^{\delta_i}\left[\frac{1}{1+\exp\left(\frac{v_i-\mu}{\sigma}\right)}\right]^{1-\delta_i}.$$

Similarly to the Weibull AFT model, the likelihood is maximized numerically, and estimates of

$\boldsymbol{\beta}$ from the POM in Eq. (25), α and λ are estimated by $\hat{\boldsymbol{\beta}} = -\hat{\boldsymbol{\omega}}/\hat{\sigma}$, $\hat{\alpha} = 1/\hat{\sigma}$ and $\hat{\lambda} = \exp(-\hat{\mu}/\hat{\sigma})$, and their variances can be estimated with the Delta method.

3.5.6 Model diagnostics

The Generalized Gamma distribution, introduced in Section 3.5.1, can be used to check which of the log-normal, Weibull or exponential distributions fit the data best, as described in Klein and Moeschberger (2003, Ch. 12). Having fitted a Weibull AFT, a hypothesis test of $H_0 : \sigma = 1$ tests whether an exponential distribution would suffice for the data.

How well a Weibull distribution fits the data can be assessed by visually inspecting a plot of the logarithm of Nelson-Aalen estimates of $H(t)$ against $\log(t)$, since $H(t) = \lambda t^{\alpha}$, and $\log[H(t)] = \log(\lambda) + \alpha \log(t)$, which is linear in $\log(t)$.

The fit of a log-logistic distribution to the data can be assessed by visually inspecting a plot of $\log[\exp\{\hat{H}_{NA}(t)\} - 1]$ against $\log(t)$, since

$$H(t) = -\log[S(t)] = -\log\left(\frac{1}{1+\lambda t^{\alpha}}\right) \Leftrightarrow \log[\exp\{H(t)\} - 1] = \log(\lambda) + \alpha \log(t),$$

which is linear in $\log(t)$. Figure 11 shows transformed Nelson-Aalen estimates for event times simulated from a Weibull distribution and a log-logistic distribution, respectively, against $\log(t)$. The Weibull $\log[\hat{H}(t)]$ is linear in $\log(t)$, while the log-logistic distribution is not. Conversely, the log-logistic $\log[\exp\{\hat{H}(t)\} - 1]$ is linear in $\log(t)$ while the Weibull distribution is not. For small sample sizes, it may be hard to evaluate the parametric fit due to the high variability in the Nelson-Aalen estimates.

To evaluate the fit of a parametric regression model, the Cox-Snell residuals, defined in Eq. (18) as

$$r_{CS,i} = \hat{H}(t_j|\boldsymbol{x}),$$

where $\hat{H}(t_j|\boldsymbol{x})$ is estimated from the fitted model, can be plotted against $\hat{H}_{NA}$ of the residuals. If the model fits, the plot should show a straight line with slope 1. For the Weibull PHM, the Cox-Snell residuals are defined as

$$r_{CS,i} = \hat{\lambda} \exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_i) t_i^{\hat{\alpha}},$$

while for the log-logistic POM they are defined as

$$r_{CS,i} = \log\left(\frac{1}{1 + \hat{\lambda} \exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_i) t_i^{\hat{\alpha}}}\right).$$

Using the log-linear representation from Eq. (21), the so called standardized residuals, in analogy with linear regression, are given by

$$r_{S,i} = \frac{\log(t_j) - \hat{\mu} - \hat{\boldsymbol{\omega}}' \boldsymbol{x_i}}{\hat{\sigma}}.$$

If the parametric model fits, these residuals should be a censored sample from error distribution of W. As in the model diagnostics of the Cox PHM in Section 3.4.3, the Martingale residuals of Eq. (19) can be used to assess the functional form of a continuous covariate, and the Deviance

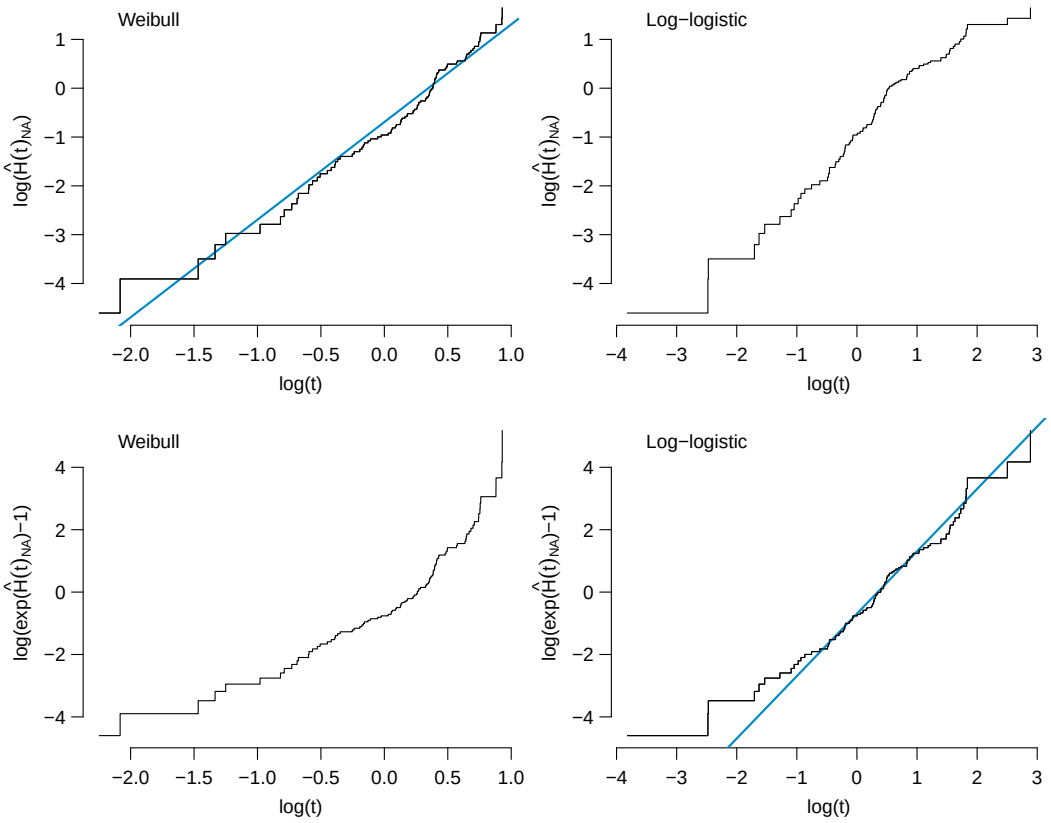

Figure 11: Transformed Nelson-Aalen estimates of $H(t)$ plotted against $\log(t)$ for $n = 100$ uncensored event times simulated from $T \sim \text{Weib}(\alpha = 2, \lambda = 0.5)$ (left plots), and $T \sim \text{LL}(\alpha = 2, \lambda = 0.5)$ (right plots). Blue line in the upper left plot shows true $\log[H(t)] = \log(\lambda) + \alpha \log(t)$ for the Weibull distribution. Blue line in the lower right plot shows true $\log[\exp\{H(t)\} - 1] = \log(\lambda) + \alpha \log(t)$ for the log-logistic distribution.

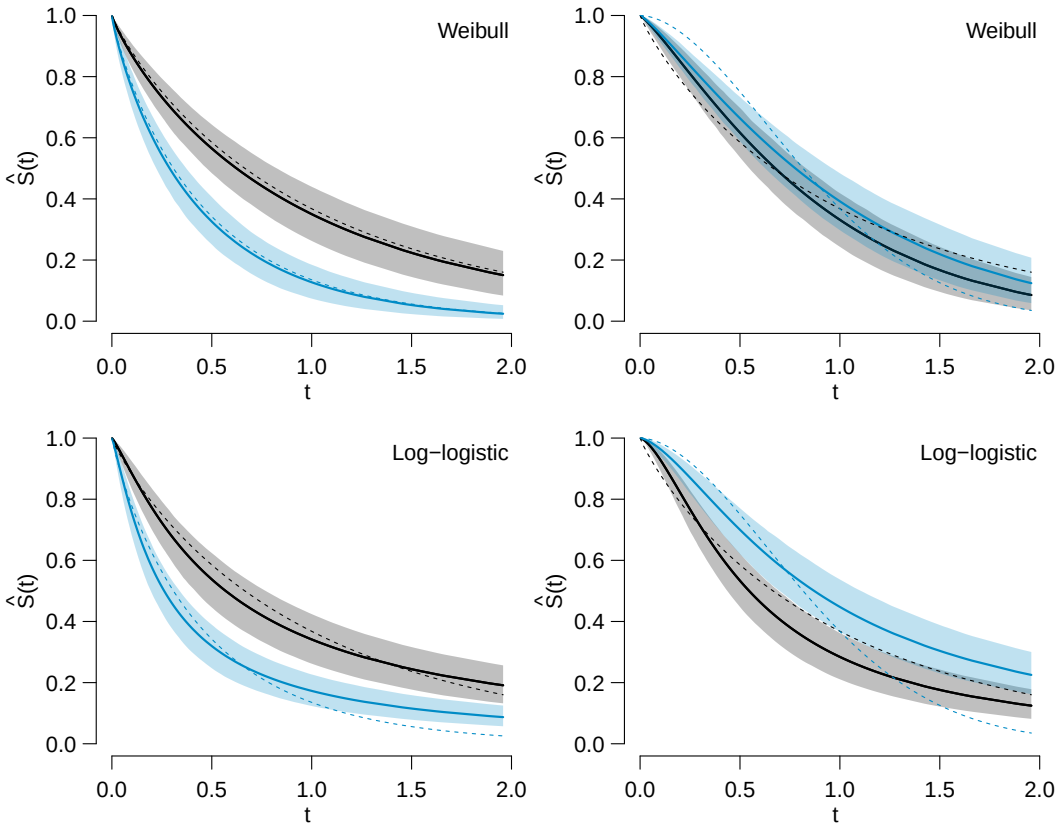

Figure 12: Example of Weibull and Log-logistic survival estimates from models including a binary predictor indicating group, using data from Figure 3. Times to event in the left column are sampled from Weib($\lambda = 1, \alpha = 0.9$) and Weib($\lambda = 2, \alpha = 0.9$). Times to event in the right column are sampled from Weib($\lambda = 1, \alpha = 0.9$) and Weib($\lambda = 1, \alpha = 1.8$). Times to censoring are sampled from a uniform distribution $U(0.2,\ 2)$. Shaded areas show 95% confidence intervals. Dashed lines show true Weibull survival curves.

residuals in Eq. (20) can be used to assess the overall model fit and to check for outliers.

In Figure 12, survival curves estimated from a Weibull and a Log-logistic model containing a single binary covariate denoting group, are shown, using data from Figure 3. None of the models can capture the crossing survival curves that are present in the right part of Figure 3.

3.6 Royston and Parmar models

In 2002, Patrick Royston and Mahesh K. B. Parmar introduced extensions of the Weibull proportional hazards and log-logistic proportional odds parametric models, arguing that their proposed method enables an easier visualization of the hazard function and a better handling of non-proportional hazards compared to the Cox model, while relaxing the parametric assumption that is inherent in fully parametric models (Royston & Parmar, 2002). The idea of Royston and Parmar models is to incorporate so called *restricted cubic splines* for the baseline log cumulative hazard and log cumulative odds. In this section, restricted cubic splines are first introduced to the reader (Section 3.6.1), after which the Royston and Parmar proportional hazards model and proportional odds model are presented (Section 3.6.2).

3.6.1 Restricted cubic splines

In various statistical regression models, the linearity assumption between (a function of) an outcome, U, and a predictor, x, may not hold, and therefore needs to be relaxed. Transforming x, for example using the natural logarithm or adding a quadratic term may be appropriate in some cases, but in other cases a more flexible transformation is needed. Using splines, i.e. piecewise polynomials with some restrictions, is a common way of dealing with this issue. Splines enable a function between U and x by dividing x into a number of intervals, joined together by so called knots, and fitting a polynomial in each interval, letting the polynomials join at the knots. If the polynomials are of degree m, then the function as a whole should have continuous $m-1$ derivatives. The simplest splines are piecewise linear (Harrell, 2015, Ch. 2.4, p. 22). These can, however, be too crude, and fitting a polynomial of higher degree may be more suitable. Cubic splines are third degree polynomials fitted between the knots, enabling high flexibility. Since higher degree polynomials can behave inappropriately at the tails, the function can be restricted to be linear before the first knot and after the last knot. Formally, restricted cubic splines (also called natural cubic splines) are calculated by selecting the number and locations of the knots, and then creating a set of extra x-variables. Let n_k denote the number of knots, including the boundary knots, and let $a_i, i = 1, ..., n_k$ be the knot locations on the x-axis. Then $n_k - 2$ extra variables are created according to (Royston & Parmar, 2002)

$$v_j(x) = (x - a_j)_+^3 - \frac{a_{n_k} - a_j}{a_{n_k} - a_1}(x - a_1)_+^3 - \frac{a_j - a_{n_k}}{a_{n_k} - a_1}(x - a_{n_k})_+^3, j = 1, ..., n_k - 2,$$

where $(z)_+ = z$ if $z > 0$ and $(z)_+ = 0$ otherwise. In a regression model, the number of fitted parameters are $n_k - 1$, excluding the intercept (Durrleman & Simon, 1989). A regression model with some link function $g(U)$, using a spline basis for x, is then written

$$g(U) = \beta_0 + \beta_1 x + \beta_2 v_1(x) + ... + \beta_{n_k-1} v_{n_k-2}(x).$$

The choice of the number and placement of knots is not obvious. With more observations, more knots can be placed, but over-fitting should be avoided. Where the x-data are sparse, less can be said about the relationship between x and U, and therefore less knots should be placed in those intervals. In the `R` package `splines` (R Core Team, 2019), the function `ns` by default places the boundary knots at the minimum and maximum x-values, respectively, and internal knots using equally spaced quantiles. The function `rcs` in `R` package `rms` also uses quantiles for knot placement, but places the boundary knots inside the data endpoints, for example at the 10%- and 90%-quantiles for 3 knots (Harrell Jr, 2019). Examples of spline fits using linear regression are shown in Figure 13. More knots are placed where the data are dense.

3.6.2 Model formulation

Royston and Parmar (2002) initially propose a proportional hazards model according to

$$H(t|\boldsymbol{x}) = H_0(t) \exp(\boldsymbol{\beta}'\boldsymbol{x}), \tag{27}$$

where $H_0(t)$ is the cumulative hazard function when $\boldsymbol{\beta} = \mathbf{0}$. Recall from Table 2 that the survival function of a Weibull distribution is $S(t) = \exp(-\lambda t^\alpha)$, and from Eq. (3) that $H(t) = -\log S(t)$.

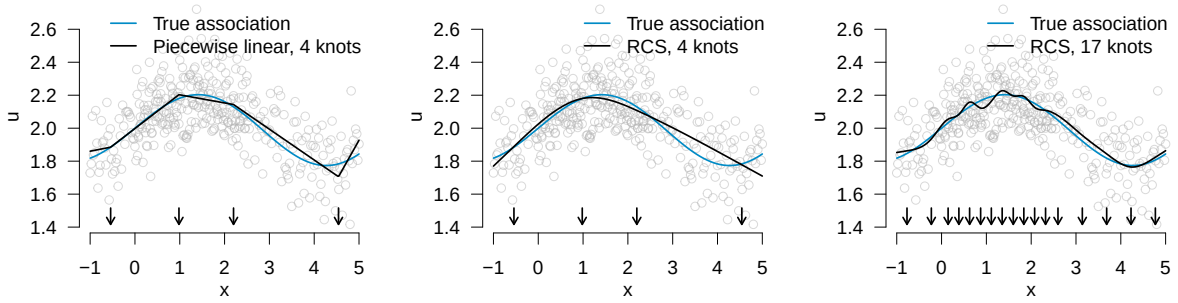

Figure 13: Example of linear regression using piecewise linear splines (left), and restricted cubic splines (RCS) (middle, right). Arrows indicate knot locations. Data simulated from $N(f(x), 0.16^2)$, where $f(x) = 2 + x/40 \cdot \cos(x) + 0.2 \cdot \sin(x)$, and x ranges from -1 to 5.

Thus, $H(t)$ for a Weibull distribution is given by

$$H(t) = -\log S(t) = -\log\left(e^{-\lambda t^\alpha}\right) = \lambda t^\alpha.$$

Taking the logarithm of $H(t)$ gives

$$\log[H(t)] = \log(\lambda) + \alpha \log(t) = \gamma_0 + \gamma_1 \log(t),$$

which is linear in $\log(t)$. In practice, T may not follow a Weibull distribution. Therefore, the non-linear function $s^*(\log(t), \boldsymbol{\gamma})$ is introduced, representing the true relationship between $\log(t)$ and $\log[H_0(t)]$. The logarithm of Eq. (27) can now be written as

$$\log[H(t|\boldsymbol{x})] = \log[H_0(t)] + \boldsymbol{\beta'x} = s^*(\log(t), \boldsymbol{\gamma}) + \boldsymbol{\beta'x}. \tag{28}$$

Restricted cubic splines are used to approximate $s^*(\log(t), \boldsymbol{\gamma})$, which is notably independent of $\boldsymbol{\beta'x}$, according to

$$s(\log(t), \boldsymbol{\gamma}) = \gamma_0 + \gamma_1 \log(t) + \gamma_2 v_1[\log(t)] + ... + \gamma_{n_k-1} v_{n_k-2}[\log(t)]. \tag{29}$$

The hazard ratio for a one-unit increase in x_k is given by

$$\frac{h(t|\boldsymbol{x} + e_k)}{h(t|\boldsymbol{x})} = \frac{\frac{d}{dt}H(t|\boldsymbol{x} + e_k)}{\frac{d}{dt}H(t|\boldsymbol{x})} = \frac{\exp(\boldsymbol{\beta'}[\boldsymbol{x} + e_k])\frac{d}{dt}H_0(t)}{\exp(\boldsymbol{\beta'x})\frac{d}{dt}H_0(t)} = \exp(\beta_k),$$

where e_k is the unit vector defined in Section 3.4.1. By definition, $H(t)$ should be monotone in t, but there is no restriction on splines to be monotone. However, Royston and Parmar argue that the spline estimation of $s(\log(t), \boldsymbol{\gamma})$ will be monotone for reasonable sample sizes.

Royston and Parmar also introduce a proportional odds model, generally defined in Eq. (25), given by

$$O(t|\boldsymbol{x}) = O_0(t) \exp(\boldsymbol{\beta'x}).$$

Using a log-logistic distribution, $T \sim \text{LL}(\alpha, \lambda)$, the baseline odds of having an event before time

t is shown in Eq. (26) to be given by $O_0(t) = \lambda t^\alpha$, yielding

$$\log[O_0(t)] = \log(\lambda) + \alpha \log(t) = \gamma_0 + \gamma_1 \log(t).$$

Similar to the baseline log cumulative hazard in Eq. (28), the baseline log odds is replaced by $s^*(\log(t), \boldsymbol{\gamma})$, which gives

$$\log[O(t|\boldsymbol{x})] = s^*(\log(t), \boldsymbol{\gamma}) + \boldsymbol{\beta}'\boldsymbol{x},$$

where $s^*(\log(t), \boldsymbol{\gamma})$ is approximated using restricted cubic splines as in Eq. (29). The odds ratio (OR) for a one-unit increase in x_k is given by $\exp(\beta_k)$.

Royston and Parmar argue that their models are robust to suboptimal knot placement, and suggest placing the boundary knots at the first and last uncensored event time, and inner knots at chosen percentiles (median for 1 inner knot, 33% and 67% percentiles for 2 inner knots, and 25%, 50% and 75% percentiles for 3 inner knots). They further suggest avoiding using more than 3 inner knots, since it may cause unstable functions.

3.6.3 Estimation and inference

The models are estimated by maximizing the likelihood functions, replacing $s^*(\log(t), \boldsymbol{\gamma})$ with $s(\log(t), \boldsymbol{\gamma})$. Let $\eta = s(\log(t), \boldsymbol{\gamma}) + \boldsymbol{\beta}'\boldsymbol{x}$. For the PHM, $S(t|\boldsymbol{x}) = \exp\{-H(t|\boldsymbol{x})\} = \exp\{-\exp(\eta)\}$. From Eq. (9), any observation, censored or uncensored, contributes to the likelihood with $h(y_i|\boldsymbol{x}_i)^{\delta_i} S(y_i|\boldsymbol{x}_i)$. Now,

$$\begin{aligned} h(t|\boldsymbol{x}) &= \frac{d}{dt}\left[H(t|\boldsymbol{x})\right] = \frac{d}{dt}\left[\exp\left\{s(\log(t), \boldsymbol{\gamma}) + \boldsymbol{\beta}'\boldsymbol{x}\right\}\right] \\ &= \frac{1}{t}\frac{d}{d\log(t)}\left[s(\log t, \boldsymbol{\gamma})\right]\exp(\eta). \end{aligned}$$

Plugging in the expressions for $h(t|\boldsymbol{x})$ and $S(t|\boldsymbol{x})$ gives the PHM likelihood function

$$\begin{aligned} L(\boldsymbol{\beta}|\boldsymbol{x}) &= \prod_{i=1}^{n} h(y_i|\boldsymbol{x}_i)^{\delta_i} S(y_i|\boldsymbol{x}_i) \\ &= \prod_{i=1}^{n} \left[\frac{1}{y_i}\frac{d}{d\log(t)}\left[s(\log(t), \boldsymbol{\gamma})\right]\Big|_{t=y_i} \exp(\eta_i)\right]^{\delta_i} \exp\{-\exp(\eta_i)\}. \end{aligned}$$

For the POM, $S(t|\boldsymbol{x}) = [1 + \exp(\eta)]^{-1}$, and

$$H(t|\boldsymbol{x}) = -\log[S(t|\boldsymbol{x})] = \log[1 + \exp(\eta)],$$

which gives

$$h(t|\boldsymbol{x}) = \frac{d}{dt}[H(t|\boldsymbol{x})] = \frac{d}{dt}[\log(1 + \exp\{\eta\})] = \frac{1}{t}\frac{d}{d\log(t)}[s(\log(t), \boldsymbol{\gamma})]\frac{\exp(\eta)}{1 + \exp(\eta)},$$

yielding the POM likelihood function

$$L(\boldsymbol{\beta}|\boldsymbol{x}) = \prod_{i=1}^{n} \left[\frac{1}{y_i}\frac{d}{d\log(t)}[s(\log(t), \boldsymbol{\gamma})]\Big|_{t=y_i} \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}\right]^{\delta_i} \frac{1}{1 + \exp(\eta_i)}.$$

In order to maximize the likelihood functions, appropriate starting values for $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ must first be found. This can be done by fitting a Cox model to the data, estimating $H(t_i|\boldsymbol{x}_i)$ with $-\log[\hat{S}(t_i|\boldsymbol{x}_i)]$ and $O(t_i|\boldsymbol{x}_i)$ with $[1-\hat{S}(t_i|\boldsymbol{x}_i)]/\hat{S}(t_i|\boldsymbol{x}_i)$, selecting starting values for $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ by fitting a linear regression model of $\log[\hat{H}(t_i|\boldsymbol{x}_i)]$ and $\log[\hat{O}(t_i|\boldsymbol{x}_i)]$, respectively, to the $\log(t_i)$, $\boldsymbol{x}_i$, and spline basis of $\log(t_i)$. After maximizing the likelihood function, standard error estimates are found from the information matrix.

3.6.4 Model extensions and diagnostics

If the proportional hazards assumption (PHA) or the proportional odds assumption (POA) is not fulfilled for a covariate x_k, then the corresponding β_k needs to be made time-dependent. Using 3 knots (2 boundary and 1 internal), the proportional hazards (PH) and proportional odds (PO) models can then be extended to

$$\log[H(t|\boldsymbol{x})] = \gamma_0 + (\gamma_{1,0} + \gamma_{1,1}x_k)\log(t) + (\gamma_{2,0} + \gamma_{2,1}x_k)v_1[\log(t)] + \boldsymbol{\beta}'\boldsymbol{x},$$

and

$$\log[O(t|\boldsymbol{x})] = \gamma_0 + (\gamma_{1,0} + \gamma_{1,1}x_k)\log(t) + (\gamma_{2,0} + \gamma_{2,1}x_k)v_1[\log(t)] + \boldsymbol{\beta}'\boldsymbol{x},$$

respectively. More generally, for n_k knots, the PH and PO models can be written

$$\log[H(t|\boldsymbol{x})] = \gamma_0 + (\gamma_{1,0} + \gamma_{1,1}x_k)\log(t) + \sum_{j=1}^{n_k-2}(\gamma_{j+1,0} + \gamma_{j+1,1}x_k)v_j[\log(t)] + \boldsymbol{\beta}'\boldsymbol{x}, \tag{30}$$

and

$$\log[O(t|\boldsymbol{x})] = \gamma_0 + (\gamma_{1,0} + \gamma_{1,1}x_k)\log(t) + \sum_{j=1}^{n_k-2}(\gamma_{j+1,0} + \gamma_{j+1,1}x_k)v_j[\log(t)] + \boldsymbol{\beta}'\boldsymbol{x}, \tag{31}$$

respectively.

The PHA can be checked in a Cox model, as described in Section 3.4.3. Otherwise a likelihood ratio test can be applied, comparing a model containing the extra $\gamma_{i,k}$s to a model assuming proportional hazards. The AIC and BIC should also reflect whether it is necessary to include non-proportionality in the models.

In Figure 14, survival curves estimated from Royston and Parmar models containing a single binary covariate denoting group, are shown, using data from Figure 3. The proportional hazards and odds models in the first row of the figure cannot capture the crossing survival curves that are present in the right part of Figure 3, while the non-proportional hazards and odds models can. For the data of the left part of Figure 3 where hazards are truly proportional, modeling the coefficient as a function of time does not make much difference.

3.7 Quantile regression

Quantile regression, where the quantiles (defined in Section 3.7.2) of an outcome variable are modeled conditional on a number of covariates, was first introduced by Koenker and Bassett (1978), and can be used for a variety of outcome variables. For survival analysis, it is natural to

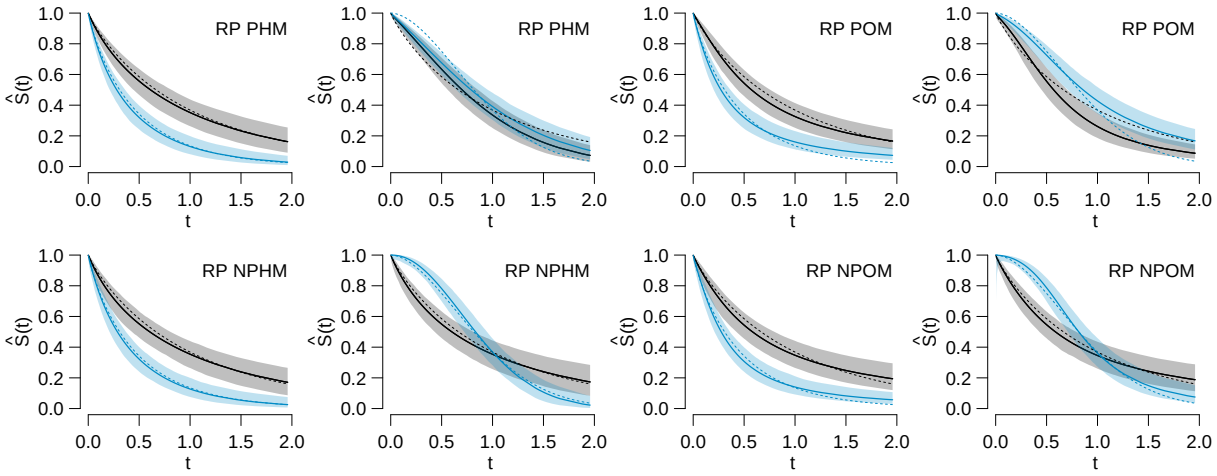

Figure 14: Example of Royston and Parmar (RP) survival estimates from models including a binary predictor indicating group, using data from Figure 3. Times to event in the first and third columns are sampled from Weib($\lambda = 1, \alpha = 0.9$) and Weib($\lambda = 2, \alpha = 0.9$). Times to event in the second and fourth columns are sampled from Weib($\lambda = 1, \alpha = 0.9$) and Weib($\lambda = 1, \alpha = 1.8$). Times to censoring are sampled from a uniform distribution $U(0.2,\ 2)$. Shaded areas show 95% confidence intervals. Dashed lines show true Weibull survival curves. PHM=Proportional hazards model, NPHM=Non-roportional hazards model, POM=Proportional oddds model, NPOM=Non-roportional odds model. In all models, 3 knots were used, including the two boundary knots.

model quantiles of T or some transformation of T. Here, the method of Frumento and Bottai (2017) (following the work of Portnoy (2003), Peng and Huang (2008), and Wang and Wang (2009)) will be introduced. First, in Section 3.7.1, piecewise constant hazards are explained, later used in the fitting of the model, after which the model formulation and estimation are presented in Sections 3.7.2 and 3.7.3.

3.7.1 Piecewise constant hazards model

A piecewise constant hazards (PCH) model is used by Frumento and Bottai (2017) when estimating the survival function, needed in the estimation of the coefficients of the quantile regression model, presented in Section 3.7.2. The time variable t is divided into A intervals, with breaks $a_1, ..., a_{A+1}$, where a_1 is the lower limit of the first interval, and a_{A+1} is the upper limit of the last interval. The hazard function is given by

$$h(t|\boldsymbol{x}) = \begin{cases} \lambda_1(\boldsymbol{x}), & a_1 < t \leq a_2 \\ \lambda_2(\boldsymbol{x}), & a_2 < t \leq a_3 \\ ... & \\ \lambda_A(\boldsymbol{x}), & a_A < t \leq a_{A+1}, \end{cases}$$

where $\lambda_k(\boldsymbol{x}), k = 1, ..., A$, are constants given $\boldsymbol{x}$. Note that piecewise constant hazards means that the survival function is piecewise exponential with $S_k(t|\boldsymbol{x}) = \exp\{-\lambda_k(\boldsymbol{x})t\}$. Poisson regression is used for modeling the logarithm of each $\lambda_k(\boldsymbol{x})$ conditional on the covariates $\boldsymbol{x}$ according to

$$\log[\lambda_k(\boldsymbol{x})] = \theta_{0,k} + \theta_{1,k}x_1 + ... + \theta_{p,k}x_p.$$

The number of fitted parameters is thus $A(p+1)$. With an increasing number of events, the number of intervals can naturally increase. A recommended number of intervals is given by $\max(\lceil m/(5p) \rceil, 5)$, where m denotes the number of events, and p the number of covariates (Frumento, 2016b). For further details on the fitting of the PCH, see Frumento (2016b).

3.7.2 Model formulation

For a random time-to-event variable T, the τ-quantile, denoted $Q(\tau)$, fulfills

$$F(Q(\tau)) = P(T \leq Q(\tau)) = \tau,$$

where τ is a probability between 0 and 1. Equivalently, the τ-quantile can be written as (Martinussen & Peng, 2014, p. 62)

$$Q(\tau) = \inf\{t : P(T \leq t) \geq \tau\}.$$

Since $S(t) = P(T > t) = 1 - F(t)$, it holds that $S(Q(\tau)) = 1 - \tau$.

An example of 0.6-quantiles, i.e. $\tau = 60\%$, in two groups, using the simulated data from the Weibull distributions given in Figure 3, is shown in Figure 15. For a Weibull distribution with survival function $S(t) = \exp\{-\lambda t^\alpha\}$, the τ-quantile is found by solving $1 - \tau = \exp\{-\lambda Q(\tau)^\alpha\}$ for $Q(\tau)$, which yields $Q(\tau) = [-\lambda^{-1} \log(1-\tau)]^{1/\alpha}$. However, the true distributions are rarely known, and a regression model may be of more use than a simple comparison between two groups.

In quantile regression, the τ-quantile of the outcome variable T is modeled conditional on a number of covariates. Martinussen and Peng (2014) define a linear model for the quantiles of $\log(T)$, enabling negative predictions that can be accurately interpreted, since, theoretically, $\log(T) \in (-\infty, \infty)$, while $T \in (0, \infty)$. However, any monotone transformation of a time-to-event random variable T can be used. An advantage of using the non-transformed T is that the regression coefficients become directly interpretable as effects on the quantiles of T. The quantile regression model for the τ-quantile of T is given by

$$Q(\tau|\boldsymbol{x}) = \boldsymbol{\beta}_\tau' \tilde{\boldsymbol{x}}_i = \beta_{0,\tau} + \beta_{1,\tau} x_1 + ... + \beta_{p,\tau} x_p, \tag{32}$$

where $\tilde{\boldsymbol{x}}_i = (1, x_1, ..., x_p)'$, and $\boldsymbol{\beta}_\tau = (\beta_{0,\tau}, \beta_{1,\tau}, ..., \beta_{p,\tau})'$. If x_k is a continuous covariate, $\beta_{k,\tau}$ represents the change in the τ-quantile for a one-unit change in x_k, when all other covariates are held constant. If x_k is a dichotomous covariate, $\beta_{k,\tau}$ represents the difference in the τ-quantile between the two groups defined by x_k. The intercept $\beta_{0,\tau}$ represents the τ-quantile when all continuous covariates are zero and all categorical covariates are at their reference level. In Figure 15, fitting the quantile regression model $Q(0.6|x_1) = \beta_{0,0.6} + \beta_{1,0.6} x_1$ using function `ctqr` in `R` package `ctqr` (Frumento, 2016a), where $x_1 = 0$ denotes the blue group and $x_1 = 1$ denotes the black group, results in the estimates $\hat{\beta}_{0,0.6} = 0.366$ and $\hat{\beta}_{1,0.6} = 0.470$. The estimated 0.6-quantile of the blue group is thus 0.366, while the corresponding number for the black group is $0.366 + 0.470 = 0.836$, and the estimated difference between the groups is 0.470, i.e. the time until 60% have experienced an event is almost half a time-unit longer for the black group.

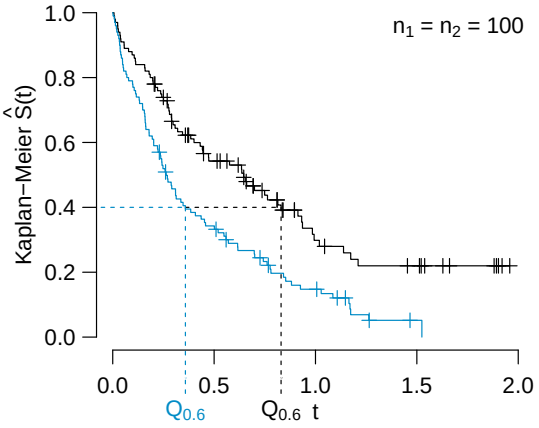

Figure 15: Kaplan-Meier survival estimates for two groups, with dashed lines marking 0.6-quantiles. Times to event are sampled from a Weibull distribution with $\lambda = 1$, $\alpha = 0.9$ (black line), and $\lambda = 2$, $\alpha = 0.9$ (blue line). The data are the same as in the left part of Figure 3.

3.7.3 Estimation and inference

Several estimating equations and fitting procedures have been proposed for quantile regression. Martinussen and Peng (2014) list methods for three different scenarios involving right-censoring: when C is always known, when C is independent of covariates, or when C is allowed to depend on covariates, where the last scenario is the most general and useful in practice. Frumento and Bottai (2017) present estimating equations and a fitting procedure for right-censored and left-truncated data. While truncation is not the focus of this thesis, the method of Frumento and Bottai (2017) have some advantages over other proposed fitting procedures. The estimating equations for when right-censoring but no truncation is present is given by

$$\boldsymbol{B}(\boldsymbol{\beta}_\tau, F) = \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{x}}_i \left[\tau - \omega_i + (1-\delta_i)\omega_i \frac{1-\tau}{S(y_i|\tilde{\boldsymbol{x}}_i)}\right], \tag{33}$$

where $\omega_i = I(y_i \leq \boldsymbol{\beta}_\tau' \tilde{\boldsymbol{x}}_i)$. The coefficient vector $\boldsymbol{\beta}_\tau$ is estimated by solving $\boldsymbol{B}(\boldsymbol{\beta}_\tau, F) = \mathbf{0}$. This has to be done in two steps, since Eq. (33) depends on the unknown $F(y_i|\boldsymbol{x}_i) = 1 - S(y_i|\tilde{\boldsymbol{x}}_i)$, which first has to be estimated and then plugged in, after which the equation $\boldsymbol{B}(\boldsymbol{\beta}_\tau, \hat{F}) = \mathbf{0}$ is solved numerically. $F(t_i|\tilde{\boldsymbol{x}}_i)$ is estimated using piecewise constant hazards, see Section 3.7.1. Frumento and Bottai (2017) show that $\hat{\boldsymbol{\beta}}_\tau$ is an unbiased estimator of $\boldsymbol{\beta}_\tau$, and that it is asymptotically normal.

There is no closed form expression of the estimator of the covariance matrix $\text{Cov}(\hat{\boldsymbol{\beta}})$. The estimation is done by incorporating the effect of the PCH estimate of $F(t_i|\boldsymbol{x}_i)$, using the formula of robust variance estimator for two-stage models of Hardin (2002). See Frumento and Bottai (2017) for details on the estimation of $\text{Cov}(\hat{\boldsymbol{\beta}})$.

3.7.4 Model diagnostics

AIC and BIC can be used for model selection in the usual sense.

Fitting quantile regression models for several values of τ, say $\tau_1, \tau_2, ..., \tau_l$, the estimates of $\beta_{k,\tau}$, $k = 1, ..., p$, can be plotted against the τ-values to check whether the $\beta_{k,\tau}$-estimates are the same across the probabilities τ. In Figure 16, $\beta_{1,\tau}$-estimates from univariable quan-

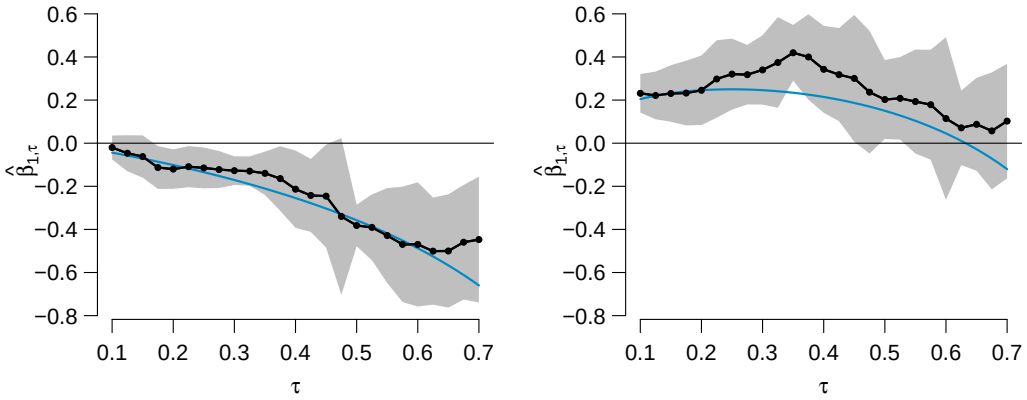

Figure 16: Plots of $\beta_{1,\tau}$-estimates against $\tau = (0.1, 0.125, 0.15, ..., 0.7)$ from quantile regression models applied to the data from Figure 3. True $\beta_{1,\tau}$s as a function of τ are plotted in blue.

Table 3: Comparison between methods. KM=Kaplan-Meier estimator, NA=Nelson-Aalen estimator, CPHM=Cox proportional hazards model, AFT=Accelerated failure time model, W PHM=Weibull proportional hazards model, LL POM=Log-logistic proportional odds model, RP PHM=Royston and Parmar proportional hazards model, RP POM=Royston and Parmar proportional odds model, QR=Quantile regression.

	KM	NA	CPHM	AFT	W PHM	LL POM	RP PHM	RP POM	QR
Non-parametric	✓	✓							
Semi-parametric			✓						
Fully parametric				✓	✓	✓			
Flexible parametric							✓	✓	
Categorical covariates	✓	✓	✓	✓	✓	✓	✓	✓	✓
Continuous covariates			✓	✓	✓	✓	✓	✓	✓
Regression model			✓	✓	✓	✓	✓	✓	✓
Hazard ratios			✓		✓		✓		
Odds ratios						✓		✓	
Extrapolation				✓	✓	✓	✓	✓	
Time-varying coefficients			✓				✓	✓	✓
$\hat{S}(t)$	✓	✓	✓	✓	✓	✓	✓	✓	
$\hat{H}(t)$	✓	✓	✓	✓	✓	✓	✓	✓	
$\hat{h}(t)$			✓	✓	✓	✓	✓	✓	

tile regression models applied to the simulated data from Figure 3 are plotted against $\tau = (0.1, 0.125, 0.15, ..., 0.7)$. In the left plot, using data from the non-crossing survival curves, one group always has less survival time than the other, and the difference is increasing in τ. In the right plot, using data from the crossing survival curves, the $\beta_{1,\tau}$-estimates are first increasing in τ and then decreasing, capturing the horizontal differences between the survival curves in Figure 3.

3.8 Method comparison

The qualities of, and what can be estimated from, the methods covered in Sections 3.2-3.7 are summarized in Table 3. The table shows that what can be estimated from fully parametric models can also be estimated by flexible parametric models, while the semi- and non-parametric models are more limited and cannot be extrapolated.

4 Simulations

In this section the results from three simulation studies are presented. The first, in Section 4.1, is an estimation of power for detecting deviations from the proportional hazards assumption (PHA). The second, in Section 4.2 compares methods assuming proportional hazards. The third, in Section 4.3, evaluates coefficient estimates, standard errors, and confidence interval coverage of quantile regression models.

4.1 Power to detect deviations from the proportional hazards assumption in a Cox proportional hazards model

In order for the coefficient estimates in a Cox PHM to be meaningful, the proportional hazards assumption (PHA) has to be fulfilled, and therefore needs careful evaluation. In this section, simulations to estimate the power in detecting deviations from the PHA are presented, with varying sample sizes, proportions of censored observations, and deviations from the PHA.

Four different Weibull time-to-event distributions were compared to a Weibull distribution with $\alpha = 0.9$, $\lambda = 1$ (the reference distribution). Figure 17 shows the Weibull survival and hazard functions, and the corresponding hazard ratios in comparison with the reference distribution. Cox PH models, defined in Eq. (11), were fitted, including a single binary covariate x, denoting group, taking values 0 or 1. When $x = 0$, the time-to-event distribution was the reference distribution. When $x = 1$, the time-to-event distribution was one of the four other Weibull distributions. When $\alpha = 0.9, \lambda = 2$, the PHA was fulfilled, as clearly seen in the right part of Figure 17, where the hazard ratios are shown. In all other cases, the PHA was not fulfilled. When $\alpha = 1$, the time-to-event distribution was exponential with $\lambda = 2$, generating a small deviation from the PHA. The cases when $\alpha = 1.2$ and $\alpha = 1.4$ corresponded to moderate and large deviations, respectively. The PHA in each fitted Cox PH model was assessed by testing the correlation between the rank-ordered event times and the scaled Schoenfeld residuals, as described in Section 3.4.3, using the function `cox.zph` in `R` package `survival` (Therneau, 2015).

The probabilities of an observation being censored before the occurrence of an event were set to $p_c = P(\delta = 0) = 0, 0.3, 0.6, 0.9$, respectively. The censoring times were uniformly distributed $U(0, \theta)$, and independent of the time-to-event distributions. For each of the simulation scenarios, θ was determined as the number generating the desired censoring probability p_c, as follows. Denote the time-to-censoring density function $g(c) = 1/\theta$. Let $f(t) = 1/2[f_1(t) + f_2(t)]$, where $f_1(t)$ are $f_2(t)$ are time-to-event distributions of the reference Weibull distribution and one of the four other Weibull distributions. Given a specific probability of censoring p_c, the upper limit

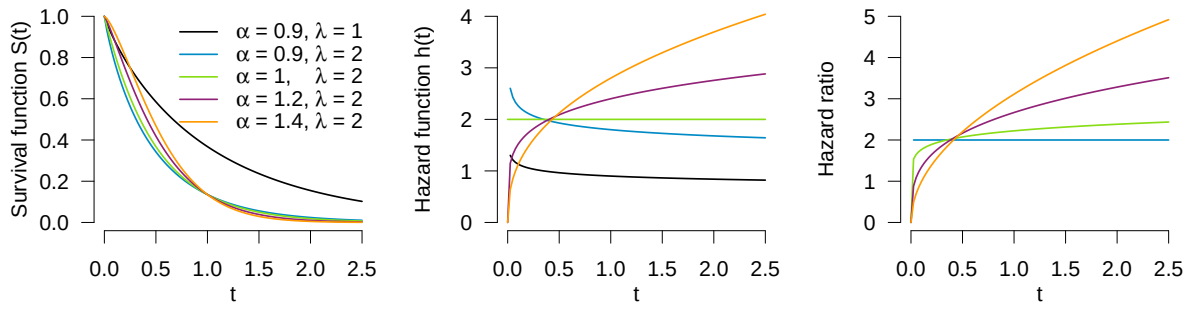

Figure 17: Survival functions, hazard functions and hazard ratios (in relation to the black line) of Weibull distributions used in simulations for estimating power in testing the PHA.

θ of the censoring distribution can be determined by solving the following equation for θ,

$$
\begin{aligned}
p_c = P(\delta = 0) &= P(0 \leq C \leq \theta, C \leq T \leq \infty) \\
&= \int_0^\theta g(c) \int_c^\infty f(t) dt dc \\
&= \int_0^\theta \frac{1}{\theta} \int_c^\infty \frac{1}{2} \left[f_1(t) + f_2(t) \right] dt dc \\
&= \int_0^\theta \frac{1}{2\theta} \int_c^\infty \left(\alpha_1 \lambda_1 t^{\alpha_1 - 1} e^{-\lambda_1 t^{\alpha_1}} + \alpha_2 \lambda_2 t^{\alpha_2 - 1} e^{-\lambda_2 t^{\alpha_2}} \right) dt dc \\
&= \int_0^\theta \frac{1}{2\theta} \left[-e^{-\lambda_1 t^{\alpha_1}} - e^{-\lambda_2 t^{\alpha_2}} \right]_{t=c}^\infty dc \\
&= \int_0^\theta \frac{1}{2\theta} \left(e^{-\lambda_1 c^{\alpha_1}} + e^{-\lambda_2 c^{\alpha_2}} \right) dc.
\end{aligned}
\tag{34}
$$

There is no closed-form expression for θ, and the above equation was therefore solved numerically using `uniroot` in `R`, for the different censoring probabilities and Weibull distributions. Note that, since the same censoring distribution is applied to both groups, the censoring proportion in the two groups combined will be on average p_c, but if $f_1(t) \neq f_2(t)$ (which is the case in these simulations), the censoring proportion in each group generally differs. This is very common in real life datasets, since one group often have events occurring faster, resulting in fewer observations being censored, while another group having events occurring slower has more censored observations. Most survival methods are applicable even if the censoring distribution is only conditionally independent of the time-to-event distribution.

The different sample sizes in each group were $n = 25, 50, 100, 250, 500, 1000$. The simulation scenarios are presented in Table 4, along with corresponding upper limits of the censoring distributions. Each scenario was simulated with 10000 Monte Carlo replications. In the scenario with no deviation from the PHA, the Type I error rate, defined as $P(\text{Reject } H_0 | H_0 \text{ is true})$, was estimated as the proportion of tests with p-values less than 5%. Similarly, in the scenarios with small to large deviations from the PHA, the statistical power of detecting deviations from the PHA, defined as $P(\text{Reject } H_0 | H_1 \text{ is true})$, was estimated as the proportion of tests with p-values less than 5%.

In Table 5, the estimated Type I error rates in the scenario of no deviation from the PHA are presented. In this scenario the null should be rejected in 5% of the cases. The table shows that without censoring, the proportion rejected null hypotheses is actually less than 5%, and

Table 4: Setup for simulation of power for testing the PHA

Scenario	Deviation from the PHA in comparison to reference	α	λ	θ for $p_c = 0.3$	θ for $p_c = 0.6$	θ for $p_c = 0.9$
Reference	-	0.9	1	-	-	-
1	None	0.9	2	2.30	0.71	0.11
2	Small	1.0	2	2.33	0.75	0.13
3	Moderate	1.2	2	2.40	0.83	0.17
4	Large	1.4	2	2.46	0.89	0.20

Table 5: Type I error rate, i.e. $P(\text{Reject } H_0 | H_0 \text{ is true})$, when testing deviations from the PHA with a significance level of 5%, using the test procedure described in Section 3.4.3, estimated using $m = 10000$ simulations for different sample sizes and proportion censored observations. The sample size in each group is denoted by n, so the total sample size is given by $2n$.

	Probability of censoring, p_c			
n	0	0.3	0.6	0.9
25	0.030	0.046	0.050	0.026
50	0.033	0.043	0.048	0.055
100	0.033	0.041	0.050	0.053
250	0.035	0.047	0.043	0.050
500	0.037	0.048	0.048	0.049
1000	0.036	0.047	0.052	0.052

around 5% in the presence of censoring. The power of detecting deviations from the PHA for small to large deviations from the PHA is presented in Table 6. The table shows that even with the relatively small sample size of $n = 100$ in each group, large deviations from the PHA can be detected with an acceptable power ($> 80\%$). Small deviations from the PHA are hard to detect even with large sample sizes. With a higher censoring probability, the power decreases, since fewer events are observed. Here, only equal group sizes are used. With unequal group sizes, the power is generally decreased, unless group sizes are inversely proportional to the proportion of censored observations in each group.

4.2 Comparing proportional hazards methods

Three of the chosen methods in this thesis work under the assumption of proportional hazards (PH) – the Cox PHM, the Weibull parametric model, and the Royston and Parmar (RP) PHM. In this section a comparison between the three methods is presented, given that the PHA holds, and that the true survival distributions are Weibull. The comparison was done by simulating Weibull event time points and Uniform censoring time points for two groups with proportional

Table 6: Power, i.e. $P(\text{Reject } H_0 | H_1 \text{ is true})$, in detecting deviations from the PHA with a significance level of 5%, using the test procedure described in Section 3.4.3, estimated using $m = 10000$ simulations for different sample sizes, proportion censored observations, and magnitudes of deviations from the PHA. The sample size in each group is denoted by n, so the total sample size is given by $2n$.

	Small deviation from the PHA				Moderate deviation from the PHA				Large deviation from the PHA			
	Probability of censoring, p_c				Probability of censoring, p_c				Probability of censoring, p_c			
n	0	0.3	0.6	0.9	0	0.3	0.6	0.9	0	0.3	0.6	0.9
25	0.046	0.063	0.058	0.032	0.136	0.147	0.104	0.040	0.289	0.273	0.167	0.039
50	0.064	0.073	0.064	0.056	0.271	0.253	0.160	0.075	0.581	0.523	0.312	0.096
100	0.098	0.102	0.081	0.062	0.534	0.469	0.274	0.094	0.891	0.819	0.559	0.148
250	0.213	0.193	0.125	0.068	0.921	0.851	0.578	0.182	1.000	0.996	0.923	0.341
500	0.406	0.340	0.203	0.081	0.998	0.990	0.871	0.314	1.000	1.000	0.998	0.610
1000	0.697	0.596	0.355	0.115	1.000	1.000	0.993	0.547	1.000	1.000	1.000	0.890

hazards 10000 times, each time fitting a Cox PHM, a parametric Weibull PHM, and an RP PHM, and calculating $\hat{\beta}$ with corresponding 95% confidence intervals and standard errors. Let x be an indicator for group, with $x = 0$ for group 1, and $x = 1$ for group 2. The Cox PHM was given by

$$h(t|x) = h_0(t)\exp(\beta x).$$

The Weibull PHM was given by

$$h(t|x) = \alpha\lambda t^{\alpha-1}\exp(\beta x).$$

The Royson and Parmar PHM was given by

$$H(t|x) = H_0(t)\exp(\beta x) = \exp\{s^*(\log t, \boldsymbol{\gamma})\}\exp(\beta x),$$

with $s^*(\log t, \boldsymbol{\gamma})$ approximated using restricted cubic splines, as described in Section 3.6.2. In the RP models, three knots, i.e. one internal knot and two boundary knots, were used in the restricted cubic splines. In each model, the internal knot was placed at the median uncensored failure time, and boundary knots were placed at the first and last uncensored failure time, respectively, as described in Section 3.6.2. In all three models, the hazard ratio was given by $\exp(\beta)$.

The time-to-event distributions were Weibull with parameters $\alpha_1 = \alpha_2 = 0.7$, $\lambda_1 = 2$, and $\lambda_2 = 1$. Survival and hazard functions are shown in Figure 18. The true hazard ratio was given by

$$\text{HR} = \frac{\alpha_2\lambda_2 t^{\alpha_2-1}}{\alpha_1\lambda_1 t^{\alpha_1-1}} = \{\alpha_1 = \alpha_2 = 0.7, \lambda_1 = 1, \lambda_2 = 2\} = \frac{0.7\cdot 2\cdot t^{0.7-1}}{0.7\cdot 1\cdot t^{0.7-1}} = 2,$$

which gives $\beta = \log(\text{HR}) = \log(2) \approx 0.6931$. The censoring distribution was $U(0,\theta), \theta = 2.21, 0.52, 0.05$, with θ determined by numerically solving Eq. (34) in Section 4.1 for censoring probabilities $p_c = 0.3, 0.6, 0.9$, respectively. A scenario with no censoring was also evaluated. The different sample sizes in each group were $n = 25, 50, 100, 250, 500$.

Average estimates, standard errors, CI widths, and CI coverage proportions are presented in Table 7. Average standard errors and average values of $\hat{\beta} - \beta$ are shown in Figure 19. The coverage proportions were around 95% for most censoring proportions and sample sizes. For small sample sizes with 90% censoring, the coverage proportions were around 98-99%, which is most likely due to the confidence intervals being very wide. Standard errors were quite similar between the three methods, but often slightly lower for the Weibull PHM, which was expected since the true distributions were Weibull. The RP PHM is based on the Weibull distribution, even though it incorporates restricted cubic splines in the baseline log cumulative hazard, which is probably why the standard errors were comparable to the Weibull PHM. Standard errors and CI widths decreased as n increased, as expected. For reasonable sample sizes and censoring proportions, $\hat{\beta}$s were close to the true β of approximately 0.69. Histograms of the β-estimates from the 10000 simulations are shown in Figures B1-B4 in Appendix B.

The function `coxph` in `R` package `survival` (Therneau, 2015) was used for fitting Cox PHMs. The function `survreg` in `R` package `survival` was used for fitting Weibull accelerated failure time models. The function `ConvertWeibull` in `R` package `SurvRegCensCov` (Hubeaux & Rufibach,

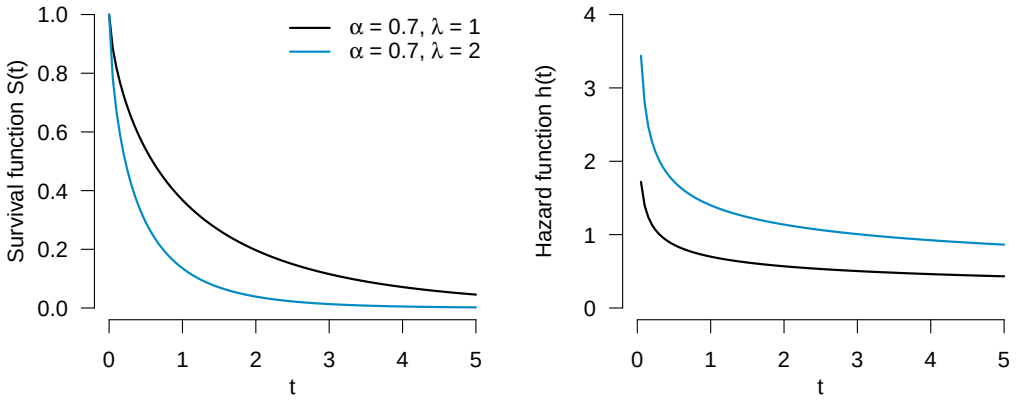

Figure 18: Survival functions and hazard functions of Weibull distributions used in simulations for comparison between PH methods.

Table 7: Results of 10000 simulations for comparison between PH methods for different censoring probabilities and sample sizes n in each group (total sample size given by $2n$). Average coefficient estimates $\hat{\beta}$, average standard error estimates $\widehat{\text{SE}}(\hat{\beta})$, average 95% CI width (CIW), and 95% CI coverage proportions (CP) are shown. The true $\beta = \log(2) \approx 0.6931$.

No censoring

	Cox PHM				Weibull PHM				RP PHM			
n	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	CIW	CP	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	CIW	CP	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	CIW	CP
25	0.7162	0.3081	1.2075	0.95	0.7254	0.2965	1.1621	0.94	0.7281	0.3023	1.1848	0.95
50	0.7053	0.2144	0.8403	0.95	0.7078	0.2083	0.8164	0.95	0.7104	0.2120	0.8311	0.95
100	0.6977	0.1503	0.5891	0.95	0.7000	0.1469	0.5758	0.95	0.7004	0.1494	0.5856	0.95
250	0.6952	0.0946	0.3707	0.95	0.6957	0.0928	0.3636	0.95	0.6961	0.0946	0.3709	0.95
500	0.6944	0.0668	0.2617	0.95	0.6947	0.0656	0.2570	0.95	0.6948	0.0673	0.2640	0.95

30% censoring

	Cox PHM				Weibull PHM				RP PHM			
n	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	CIW	CP	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	CIW	CP	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	CIW	CP
25	0.7091	0.3571	1.3997	0.95	0.7225	0.3497	1.3710	0.94	0.7214	0.3521	1.3802	0.95
50	0.7035	0.2487	0.9750	0.95	0.7092	0.2455	0.9625	0.95	0.7093	0.2468	0.9673	0.95
100	0.6982	0.1745	0.6839	0.95	0.7013	0.1729	0.6778	0.95	0.7010	0.1736	0.6806	0.95
250	0.6940	0.1099	0.4307	0.95	0.6953	0.1092	0.4280	0.95	0.6953	0.1096	0.4297	0.95
500	0.6938	0.0776	0.3041	0.95	0.6945	0.0771	0.3024	0.95	0.6944	0.0775	0.3036	0.95

60% censoring

	Cox PHM				Weibull PHM				RP PHM			
n	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	CIW	CP	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	CIW	CP	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	CIW	CP
25	0.7270	0.4846	1.8997	0.96	0.7399	0.4804	1.8833	0.95	0.7334	0.4800	1.8817	0.95
50	0.7098	0.3344	1.3109	0.95	0.7161	0.3329	1.3050	0.95	0.7150	0.3333	1.3064	0.95
100	0.6983	0.2337	0.9159	0.95	0.7016	0.2331	0.9135	0.95	0.7008	0.2332	0.9143	0.95
250	0.6932	0.1468	0.5753	0.95	0.6945	0.1465	0.5745	0.95	0.6943	0.1466	0.5748	0.95
500	0.6942	0.1036	0.4060	0.95	0.6948	0.1035	0.4056	0.95	0.6946	0.1035	0.4058	0.95

90% censoring

	Cox PHM				Weibull PHM				RP PHM			
n	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	CIW	CP	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	CIW	CP	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	CIW	CP
25	0.5307	0.9987	3.9147	0.99	0.5441	0.9947	3.8990	0.98	0.4550	1.0019	3.9274	0.99
50	0.7363	0.7369	2.8885	0.98	0.7435	0.7352	2.8819	0.98	0.6737	0.7142	2.7998	0.98
100	0.7376	0.5001	1.9602	0.96	0.7397	0.4994	1.9574	0.96	0.7245	0.4962	1.9451	0.96
250	0.7049	0.3037	1.1905	0.95	0.7061	0.3036	1.1899	0.95	0.7059	0.3036	1.1901	0.95
500	0.6950	0.2125	0.8331	0.95	0.6956	0.2125	0.8329	0.95	0.6954	0.2125	0.8330	0.95

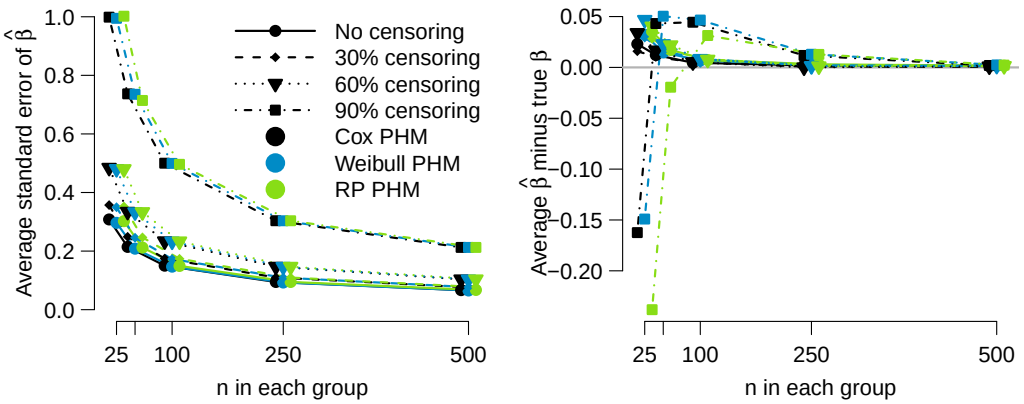

Figure 19: Average standard errors of $\hat{\beta}$, and average $\hat{\beta} - \beta$ in 10000 simulations for comparison between PH methods for different censoring probabilities and sample sizes. Points are slightly separated horizontally for visualization purposes.

2015) was used for conversion of Weibull estimates and standard errors to the PH parametrization, using the delta method to estimate standard errors. The function `flexsurvspline` in `R` package `flexsurv` (Jackson, 2016) was used for fitting Royston and Parmar PHMs. This function almost always succeeds at finding a maximum likelihood estimate of β, but sometimes fails when estimating the standard error, due to the Hessian matrix not being positive definite. This problem seems more common with an increasing number of knots. Also, `flexsurvspline` fails at fitting a model relatively often for small sample sizes and large censoring proportions, due to one or more of the initial values in the `optim` function being infinite. This can sometimes be solved by manually trying other initial values. This was however not applied in this simulation, due to time restrictions. With a high proportion censored observations and a small sample size, all observations in one or both groups can sometimes be censored, resulting in non-convergence of all three methods. Non-convergence can also occur for other reasons than this. A summary of all failures and non-convergences is presented in Table B1 in Appendix B.

4.3 Estimates, standard errors and confidence interval coverage of quantile regression

In order to evaluate the quantile regression estimates, standard errors and confidence interval coverage for comparison between two groups, for different probabilities τ, sample sizes, censoring proportions, and Weibull distributions, 5000 Monte Carlo replications were performed, each time fitting a quantile regression model, given by

$$Q(\tau|x) = \beta_{0,\tau} + \beta_{1,\tau}x,$$

where $x = 0$ for group 1 and $x = 1$ for group 2.

The Weibull distribution has a closed form expression for the τ-quantile, derived in Section 3.7.2, given by

$$Q(\tau) = [-\lambda^{-1}\log(1-\tau)]^{1/\alpha}.$$

Hence, the difference in the τ-quantiles between two groups with $T_2 \sim \text{Weib}(\alpha_2, \lambda_2)$ and $T_1 \sim$

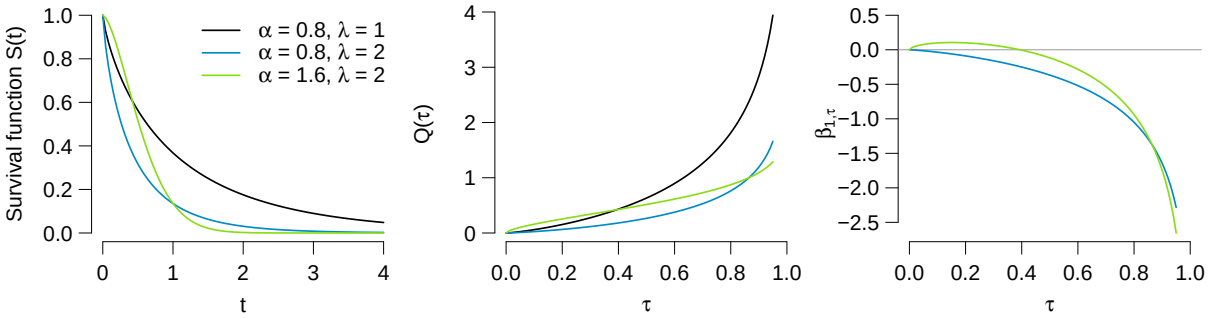

Figure 20: Survival functions, quantile functions and $\beta_{1,\tau}$s (in relation to the black line) of Weibull distributions used in simulations of quantile regression models.

Weib(α_1, λ_1) is given by

$$Q_2(\tau) - Q_1(\tau) = [-\lambda_2^{-1}\log(1-\tau)]^{1/\alpha_2} - [-\lambda_1^{-1}\log(1-\tau)]^{1/\alpha_1}.$$

The survival function, quantile function, and true $\beta_{1,\tau}$s are shown in Figure 20. The distribution represented by the black line served as a reference, and hence the other distributions were compared to it. The chosen censoring proportions were $p_c = 0, 0.3$. Higher censoring proportions were excluded, since quantiles of high probabilities cannot be estimated with too much censoring. Censoring times were simulated from a uniform distribution, $U(0, \theta)$, with θ determined in each scenario by solving Eq (34) for $p_c = 0.3$, resulting in $\theta = 2.26, 2.52$ for the two different scenarios. The sample size in each group was given by $n = 25, 50, 100, 250, 500$. The τ-values used were $0.05, 0.1, 0.15, 0.2, ..., 0.6$.

Simulation results are presented in Figure 21. Even for small sample sizes, the average $\hat{\beta}_{1,\tau}$s are close the true $\beta_{1,\tau}$s, and are equivalent for 0% and 30% censoring. The standard error estimates of the $\hat{\beta}_{1,\tau}$s increase in τ, which is due to the fact that fewer individuals are still in the risk set at higher τ-values. Comparing 0% and 30% censoring, the standard errors are higher in the latter case, especially for higher τ-values, but the difference decreases as n increases. In Appendix C, Table C1 shows all the numbers of Figure 21.

The coverage of 95% confidence intervals of $\beta_{1,\tau}$ are shown in Figure 22, calculated for each τ as the proportion of confidence intervals containing the true $\beta_{1,\tau}$. The coverage only reaches 95% for the largest sample size, and is around or slightly above 90% for $n = 100$.

The function `ctqr` in R package `ctqr` (Frumento, 2016a) was used in the simulations. In some cases, a quantile regression model was claimed to have converged, while the standard error estimates were huge. Standard errors above the value of 10, and corresponding $\beta_{1,\tau}$-estimates, were treated as non-convergences and were excluded. This was more common for small sample sizes and high τ-values. The distribution of the standard error estimates for the smallest sample size was highly right skewed, and the median was less than half of the mean, even after excluding standard error estimates above 10.

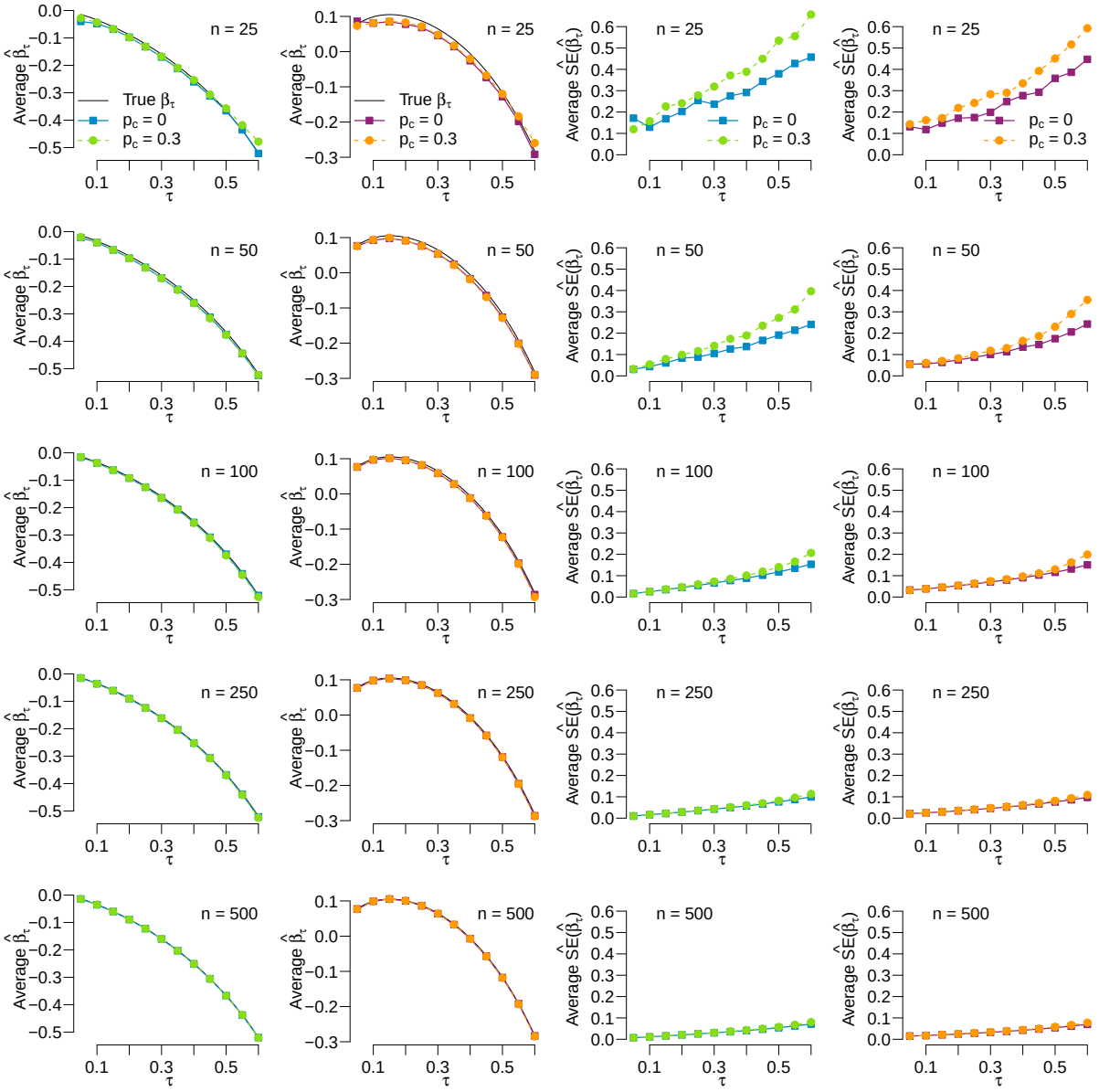

Figure 21: Average $\hat{\beta}_{1,\tau}$ and $\widehat{\text{SE}}(\hat{\beta}_{1,\tau})$ from 5000 simulations of quantile regression models, for different values of τ, sample sizes n in each group, and censoring proportions p_c. The first and third columns correspond to a Weibull distribution with $\alpha = 0.8, \lambda = 2$, compared to the reference distribution given by Weib($\alpha = 0.8, \lambda = 1$). The second and forth columns correspond to a Weibull distribution with $\alpha = 1.6, \lambda = 2$, compared to the reference distribution.

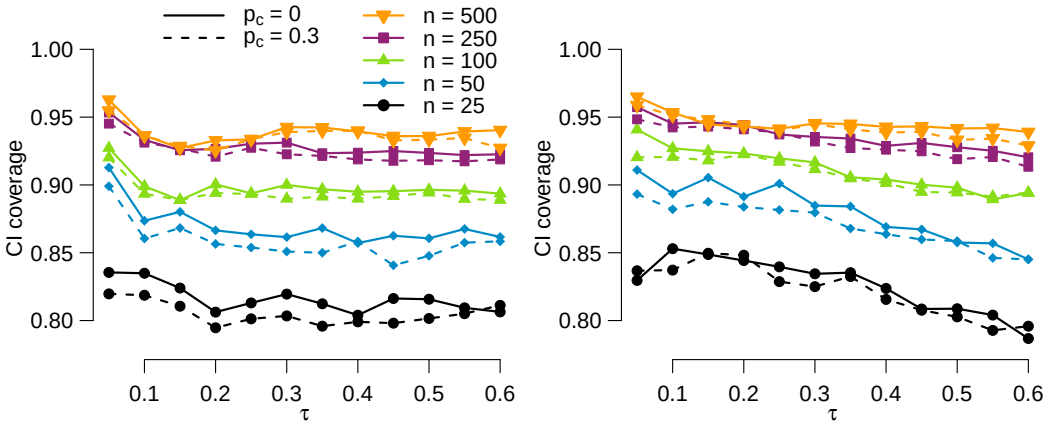

Figure 22: CI coverage of true $\beta_{1,\tau}$ from 5000 simulations of quantile regression models, for different values of τ, sample sizes n in each group, and censoring proportions p_c. The left plot corresponds to a Weibull distribution with $\alpha = 0.8, \lambda = 2$, compared to the reference distribution given by Weib($\alpha = 0.8, \lambda = 1$). The right plot corresponds to a Weibull distribution with $\alpha = 1.6, \lambda = 2$, compared to the reference distribution.

5 Application to acute lymphoblastic leukemia data

In this section, the methods described in Section 3 are applied to the acute lymphoblastic leukemia (ALL) data. The disease and treatment are briefly described in Section 2.3.1.

5.1 Data description and patient characteristics

The ALL patient dataset comes from the NOPHO ALL registry, and includes $n = 2024$ children diagnosed with ALL between July 2008, and December 2018, treated with the NOPHO ALL2008 protocol described in Section 2.3.1.

In Figure 23, a flowchart of patient exclusions is shown. A covariate of interest is the risk group, defined around the 29th day of treatment. The risk group is based on a number of factors, for example the amount of remaining cancer cells after the intensive first few weeks of treatment, and is the basis for decisions on the rest of the treatment for each patient. To be able to include risk group as a covariate known from start, the start time was defined as day 29 of treatment. Children with an event or censoring before or on day 29, or with an undefined risk group, were excluded. Children with unknown central nervous system (CNS) status at ALL diagnosis were also excluded. The CNS status is an indicator of cancer cells being present in the cerebrospinal fluid (CSF). Children with cancer cells in the CSF get partly different treatment. Another covariate of interest is the ALL type, denoting what kind of immune cells have developed into cancer cells – B-cell precursor cells (BCP) or T-cells. Children with T-cell ALL often have worse outcomes compared to children with BCP ALL, and therefore often end up in more severe risk groups and get partially differing treatment. Characteristics of the included children are presented in Table 8. A histogram of age at diagnosis is presented in Figure 24, showing a right-skewed distribution with a peak between ages 2-4 years.

As mentioned in Section 2.3, so called event-free survival (EFS) is an important outcome in many cancer studies, and it was used here as the main outcome in analyses of the childhood ALL data. EFS was defined as time (years) from start (day 29 of treatment) to an event or censoring. An event was defined as death, relapse of cancer, or a second malignancy. Right-censoring is heavily present in the data, while truncation or left-censoring is not.

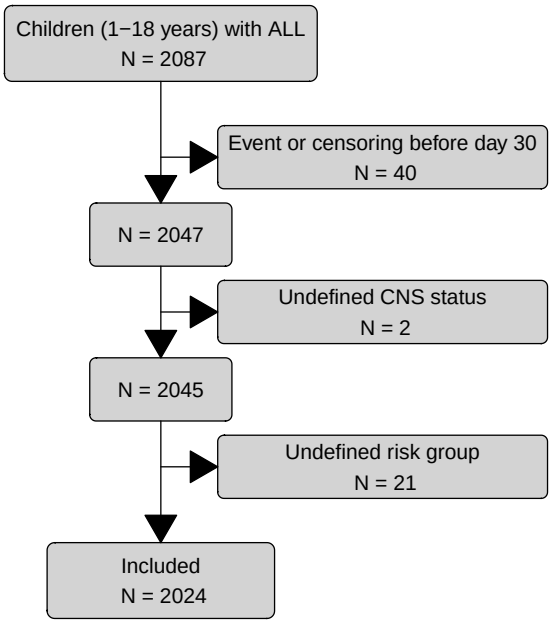

Figure 23: Flowchart of included patients in the NOPHO ALL data

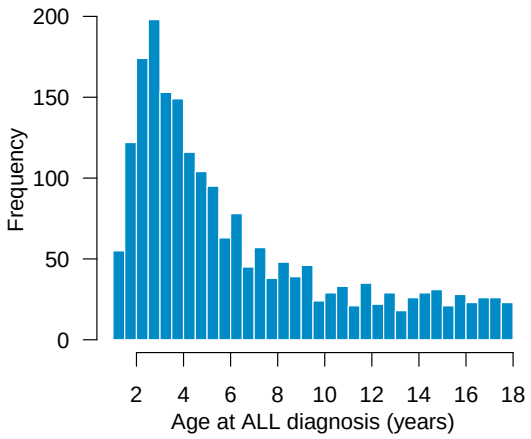

Figure 24: Distribution of age (years) at diagnosis in children with ALL

5.2 Non-parametric event-free survival and median follow-up time

The Kaplan-Meier estimate of the five-year EFS was 87.0% (95% CI 85.2% - 88.6%). The median follow-up time was 4.80 years, estimated using the reverse Kaplan-Meier method, which means letting censoring be the event and the patients with events be censored at their respective event times.

Kaplan-Meier survival curves of EFS for the full cohort, and for different subgroups are shown in Figure 25. None of the survival curves cross, but non-proportional hazards can still be suspected for the different ALL types, and for the high risk group compared to standard and intermediate risk groups. The survival curves for males and females are very similar and almost impossible to distinguish. The confidence interval of Eq. (6) was used, calculated by specifying the option `conf.type="log-log"` in the `survfit` function in `R` package `survival` (Therneau, 2015).

5.3 Cox proportional hazards models

In this section, results from Cox proportional hazards models (PHM) applied to the ALL data are presented. First, univariable models are presented in Section 5.3.1, after which the proportional hazards assumption is tested in Section 5.3.2. The functional form of the continuous covariate age at diagnosis is assessed in Section 5.3.3. In Section 5.3.4 the fit of a multivariable Cox model is evaluated.

5.3.1 Univariable Cox proportional hazards models

Table 9 shows estimates, standard errors, 95% confidence intervals, and p-values from univariable Cox models containing the covariates gender, ALL type, CNS involvement, risk group and age at diagnosis. As could be suspected from the Kaplan-Meier curves in Figure 25, the coefficient of gender was not statistically significantly different from zero at any standard significance

Table 8: Patient characteristics of children with ALL

	n (%)
Total n	2024 (100.0)
Gender:	
Male	1134 (56.0)
Female	890 (44.0)
Country:	
Nordic states	1811 (89.5)
Baltic states	212 (10.5)
ALL type:	
BCP	1749 (86.4)
T-cell	275 (13.6)
CNS Leukemia:	
No	1747 (86.3)
Yes	277 (13.7)
Risk group:	
Standard risk	1005 (49.7)
Intermediate risk	712 (35.2)
High risk	307 (15.2)
Outcome:	
Censored	1813 (89.6)
Death w/o relapse	54 (2.7)
Death after relapse	59 (2.9)
Death after second malignancy	4 (0.2)
Relapse w/o death	85 (4.2)
Second malignancy w/o death	9 (0.4)
EFS status:	
Event	211 (10.4)
No event	1813 (89.6)
	Median (IQR, min - max)
Age (years)	4.7 (2.9 - 8.8; 1.0 - 18.0)

level. The other covariates, however, were, with T-cell ALL, CNS involvement, high risk and intermediate risk groups, and older age at diagnosis, associated with a higher hazard compared to the reference levels.

5.3.2 Testing the proportional hazards assumption

For the β-estimates in Table 9, and, equivalently, the hazard ratios, to be valid, the proportional hazards assumption should be fulfilled. Univariable Cox PH models were fitted, with gender, ALL type, CNS involvement at diagnosis, age (years) at diagnosis, and risk group, as covariates. The proportional hazards assumption (PHA) was evaluated by plotting the scaled Schoenfeld residuals against the rank-ordered event times, using the function `cox.zph` in `R` package `survival` (Therneau, 2015). A smooth curve was automatically added by the function

Table 9: Coefficient estimates $\hat{\beta}$, standard errors $\widehat{\mathrm{SE}}(\hat{\beta})$, 95% confidence intervals (CIs), and p-values from univariable Cox proportional hazards models applied to the ALL data

Covariate	$\hat{\beta}$	$\widehat{\mathrm{SE}}(\hat{\beta})$	95% CI for β	p-value
Gender: Female vs Male	0.023	0.138	-0.248 - 0.294	0.87
ALL Type: T-cell vs BCP	0.768	0.164	0.446 - 1.09	2.9e-06
CNS involvement: Yes vs No	0.522	0.174	0.181 - 0.863	0.0027
Risk group: IR vs SR	0.533	0.175	0.19 - 0.875	0.0023
Risk group: HR vs SR	1.645	0.171	1.309 - 1.981	8e-22
Age at diagnosis (years)	0.072	0.014	0.045 - 0.098	1.4e-07

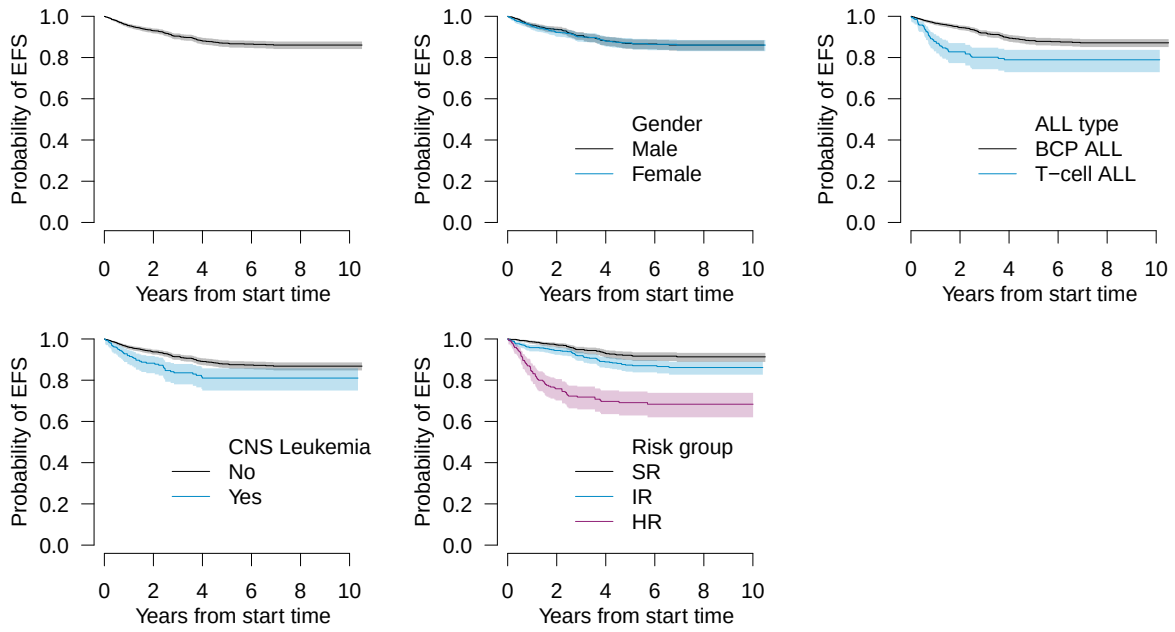

Figure 25: Kaplan-Meier curves with 95% confidence intervals of different groups of ALL data. SR=Standard risk, IR=Intermediate risk, HR=High risk.

`plot.cox.zph` to illustrate the relationship between time and the residuals. At a 5% significance level, ALL type and risk group deviated from the PHA. This means that a time invariant hazard ratio is not suitable as a description of the corresponding covariate effects.

5.3.3 The functional form of the age covariate

In many studies of childhood ALL patients, see e.g. Toft et al. (2018), the age covariate is dichotomized around the age of ten, serving as a proxy for puberty. Dichotomizing a continuous covariate adds the assumption that there are no differences in the hazard rate between two values that are below (above) the cutoff value. This assumption is very simplified, and seldom correct. A lot could be gained in power and model fit by using the continuous covariate rather than its dichotomized version. However, it is important to check that the linearity assumption of a continuous covariate holds. As described in Section 3.4.3, the functional form of a continuous covariate can be assessed by plotting the Martingale and Deviance residuals, defined in Eq. (19) and Eq. (20), of a Cox PHM not containing the continuous covariate, against the covariate. In Figure 27, the Martingale and Deviance residuals of a Cox PHM containing the covariates gender, ALL type, CNS status and risk group, are plotted against age at diagnosis. Since there is a high proportion censored observations in the ALL data, there is a large proportion of residuals close to zero. Looking at only the uncensored residuals, there is no clear relationship with age at diagnosis, even though adding the age covariate to the Cox PHM yields a p-value of 0.00089 and a hazard ratio of 1.05 (95% CI 1.02 - 1.08).

Another way to assess the functional form of a covariate is to fit a Cox PHM using a spline basis for the covariate, and compare it to a model containing a linear term for the covariate, using a likelihood ratio test. For the ALL data, this was done by comparing a model containing age at diagnosis, gender, ALL type, CNS status, and risk group to models also incorporating restricted cubic spline covariates for age, using 3, 4, and 5 knots, respectively, with likelihood

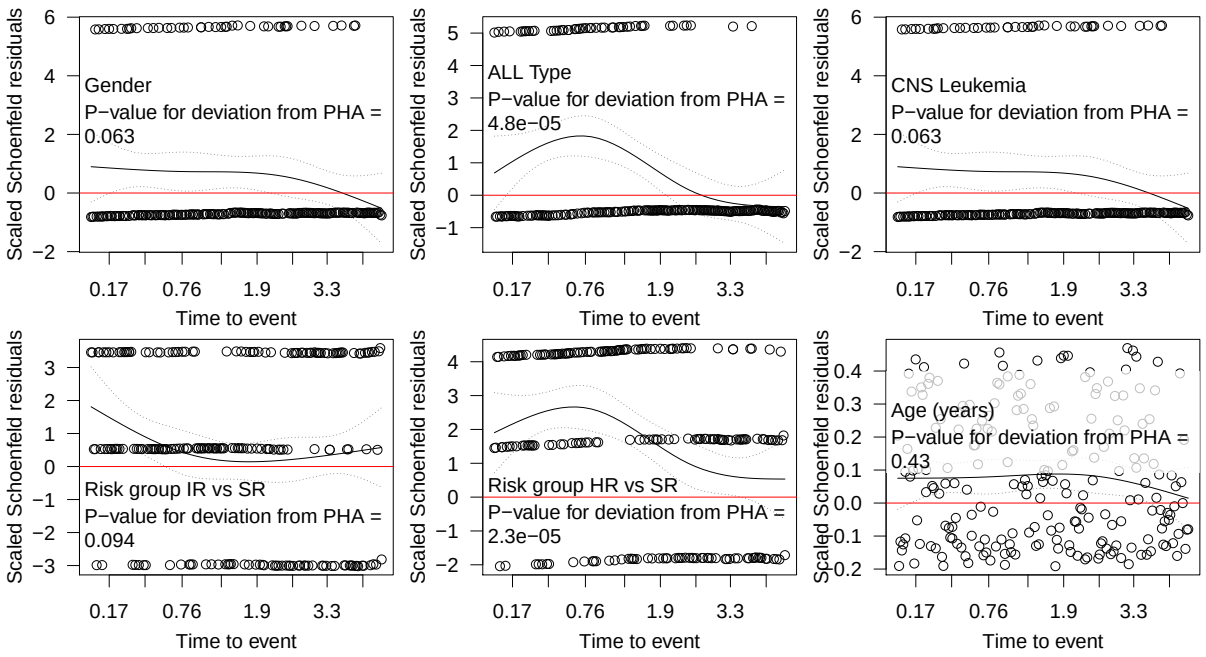

Figure 26: Scaled Schoenfeld residuals from three univariable Cox PH models containing gender, ALL type, CNS, age at diagnosis, and risk group, respectively, plotted against rank-ordered event times. A smooth curve for the association between (rank-ordered) time and the Scaled Schoenfeld residuals was added for illustrative purposes. SR=Standard risk, IR=Intermediate risk, HR=High risk.

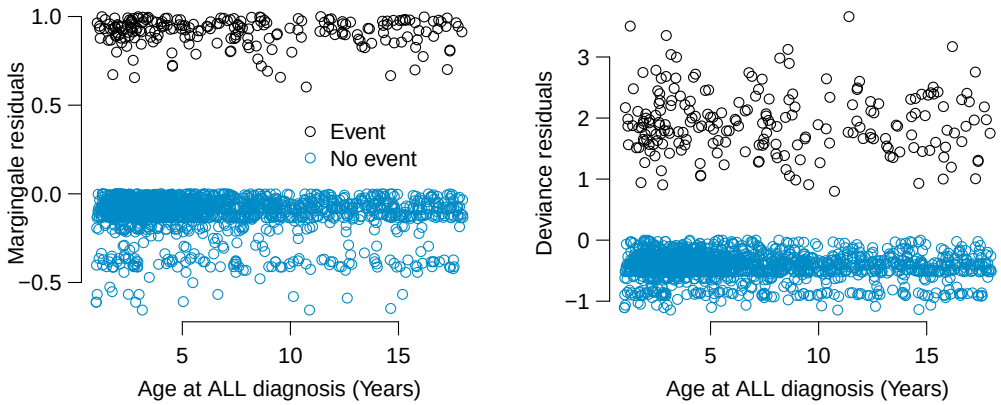

Figure 27: Martingale and deviance residuals from a Cox PHM including gender, ALL type, CNS status, and risk group, plotted against age at ALL diagnosis.

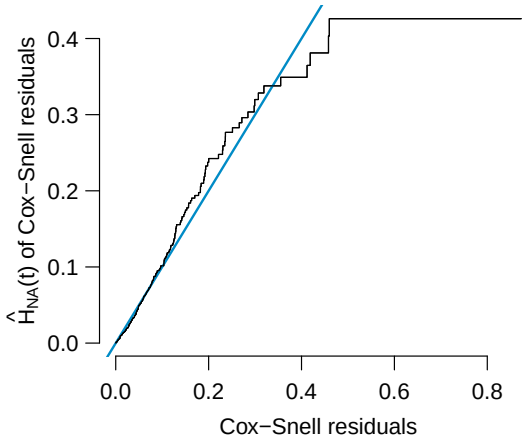

Figure 28: Nelson-Aalen estimates of Cox-Snell residuals plotted against Cox-Snell residuals of a Cox PHM including the covariates Gender, ALL type, CNS status, risk group and age at diagnosis. A line with slope 1 through the origin is shown in blue.

Table 10: Coefficient estimates $\hat{\beta}$, standard errors $\widehat{\text{SE}}(\hat{\beta})$, 95% confidence intervals (CIs), and p-values from a multivariable Cox proportional hazards model applied to the ALL data

Covariate	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	95% CI for β	p-value
Gender: Female vs Male	0.098	0.14	-0.177 - 0.373	0.49
ALL Type: T-cell vs BCP	-0.022	0.184	-0.383 - 0.338	0.9
CNS involvement: Yes vs No	0.457	0.179	0.107 - 0.807	0.01
Risk group: IR vs SR	0.411	0.181	0.056 - 0.766	0.023
Risk group: HR vs SR	1.489	0.19	1.117 - 1.861	4.2e-15
Age at diagnosis (years)	0.047	0.014	0.019 - 0.075	0.00089

ratio tests. Knot placement was done according to the recommendation of Royston and Parmar (2002), described in Section 3.6.2. The p-values of the likelihood ratio tests were 0.99, 0.24, 0.49, for 3, 4, and 5 knots, respectively. This means that there was no statistically significant deviation from the linearity assumption of the age covariate at any standard significance level.

5.3.4 Fit of a multivariable Cox proportional hazards model

As stated in Section 3.4.3, the fit of a Cox PHM can be assessed by plotting the Cox-Snell residuals, defined in Eq. (18) against the Nelson-Aalen estimates using the residuals as the time variable. Such a plot can be found in Figure 28, with Cox-Snell residuals from a Cox PHM including gender, ALL type, CNS status, risk group and age at diagnosis. The plot shows that the Cox model has a decent but not perfect fit. The estimates of the Cox model are presented in Table 10. Comparing the estimates and p-values from the multivariable model to the univariable models, presented in Table 9, it can be seen that ALL type loses statistical significance and that the corresponding coefficient is close to zero. This can be explained by the fact that the covariates ALL type and risk group are highly correlated – children with T-cell ALL are divided fairly equally between the intermediate and high risk groups, while no T-cell ALL children are present in the standard risk group.

5.4 Parametric and flexible parametric fit

As stated in Section 3.5, a lot could be gained by fitting parametric models rather than semi-parametric models, provided that the distribution assumptions are fulfilled. Figure 29 shows

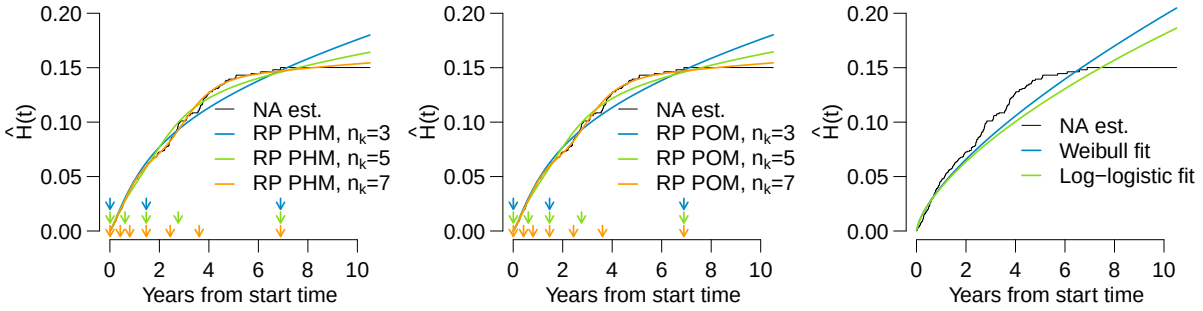

Figure 29: Cumulative hazard estimated with Royston and Parmar proportional hazards (PH) (left plot) and proportional odds (PO) (middle plot) models with 3 knots (blue), 5 knots (green), and 7 knots (orange), respectively, and parametric Weibull (blue) and log-logistic (green) cumulative hazards (right plot). Knot locations are indicated with arrows in the respective color of the lines. Nelson-Aalen estimates of the cumulative hazard shown in black.

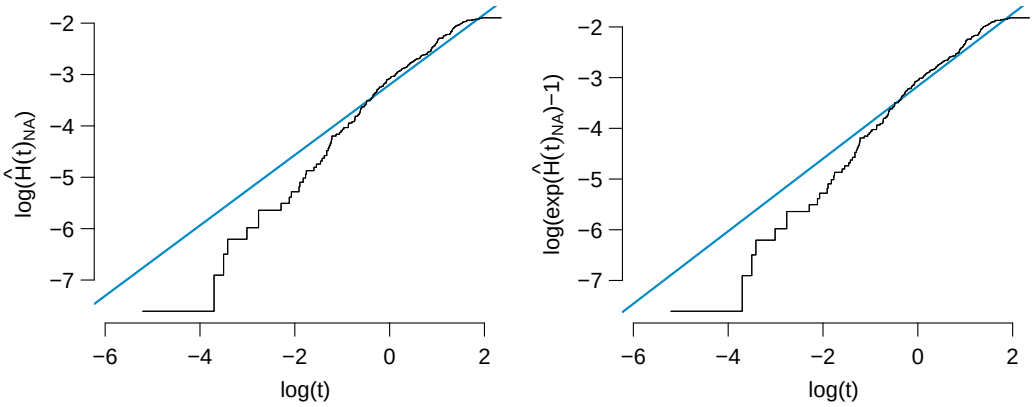

Figure 30: Transformed Nelson-Aalen estimates of $H(t)$ plotted against $\log(t)$ for NOPHO ALL survival data. Blue lines show $\log(\hat{\lambda}) + \hat{\alpha}\log(t)$ with $\hat{\lambda}$ and $\hat{\alpha}$ from a Weibull model (left plot) and log-logistic model (right plot), respectively.

the cumulative hazard estimated using Royston and Parmar PHMs and POMs with a varying number of knots, and parametric Weibull and log-logistic fits, along with the non-parametric Nelson-Aalen estimates. Knots were placed at the quantiles described in Section 3.6.2. It can be seen that the more knots placed, the more flexible the curve. However, using a large number of knots, there is a risk of over-fitting the model to the data. As seen from the right plot, the fully parametric models fit the data poorly.

Another way to assess the fit of parametric models is to plot transformations of $\hat{H}_{NA}(t)$ against $\log(t)$, as described in Section 3.5.6. The fit of a Weibull and log-logistic distribution, respectively, to the ALL data was checked by fitting Weibull and log-logistic AFT models to the data, only containing an intercept term, estimating λ and α as described in Section 3.5.2 by $\hat{\lambda} = \exp(-\hat{\mu}/\hat{\sigma})$ and $\hat{\alpha} = 1/\hat{\sigma}$. The NA estimates of the cumulative hazard function $H(t)$ were calculated, and $\log[\hat{H}_{NA}(t)]$ and $\log[\exp\{\hat{H}_{NA}(t)\} - 1]$ were plotted against $\log(t)$ in Figure 30, with straight lines added showing the $\log[\hat{H}(t)] = \log(\hat{\lambda}) + \hat{\alpha}\log(t)$ and $\log[\exp\{\hat{H}(t)\} - 1] = \log(\hat{\lambda}) + \hat{\alpha}\log(t)$ estimated form the Weibull and log-logistic AFT models, respectively. As seen from the figure, neither the Weibull nor the log-logistic distribution seemed to fit the data well.

5.5 Royston and Parmar models

As seen from Figure 26, there were deviations from the PHA for the covariates ALL type and risk group. Model fit could be improved by relaxing the proportionality assumptions. Deviations from the proportional hazards or odds assumption can be modeled by adding terms in the spline modeling of $s^*(\log(t), \boldsymbol{\gamma})$, see Eq.s (30) and (31). To assess the proportionality assumptions of the hazards and odds, univariable proportional and non-proportional models were fitted, comparing their respective AIC and BIC. In these models, 3 knots were used. For any covariate x, the proportional hazards and odds models were defined as

$$\log[H(t|x)] = \gamma_0 + \gamma_1 \log(t) + \gamma_2 v_1[\log(t)] + \beta x,$$

and

$$\log[O(t|x)] = \gamma_0 + \gamma_1 \log(t) + \gamma_2 v_1[\log(t)] + \beta x,$$

respectively. The non-proportional hazards and odds models were defined as

$$\log[H(t|x)] = \gamma_0 + (\gamma_{1,0} + \gamma_{1,1} x) \log(t) + (\gamma_{2,0} + \gamma_{2,1} x) v_1[\log(t)] + \beta x,$$

and

$$\log[O(t|x)] = \gamma_0 + (\gamma_{1,0} + \gamma_{1,1} x) \log(t) + (\gamma_{2,0} + \gamma_{2,1} x) v_1[\log(t)] + \beta x,$$

respectively. When modeling the age covariate in the non-proportional models, however, fitting was not possible due to the Hessian matrix not being positive definite, and thus the following non-proportional models were fitted

$$\log[H(t|x)] = \gamma_0 + (\gamma_{1,0} + \gamma_{1,1} x) \log(t) + \gamma_2 v_1[\log(t)] + \beta x,$$

and

$$\log[O(t|x)] = \gamma_0 + (\gamma_{1,0} + \gamma_{1,1} x) \log(t) + \gamma_2 v_1[\log(t)] + \beta x,$$

respectively.

In Table 11, the AIC and BIC of univariable Royston and Parmar models are presented, for proportional and non-proportional models. For gender, CNS involvement, and age, the AIC and BIC of the proportional models were lower than of the non-proportional models. This seems logical, considering the tests of the PHA in Figure 26 of these covariates were all non-significant at the 5% level. For ALL type, the AIC and BIC of the non-proportional models were lowest. The Kaplan-Meier curve of ALL type in Figure 25 showed clearly non-proportional hazards. The AIC showed an advantage of modeling a non-proportional effect of the risk group covariate, while the BIC did not, perhaps because only one of the groups had non-proportional hazards compared to the standard risk group. Survival estimates from proportional and non-proportional hazards Royston and Parmar univariable models are shown in Figure 31. The non-proportional hazards are captured well for ALL type and risk group.

Table 11: AIC and BIC of univariable Royston and Parmar models. PHM=Proportional hazards model, NPHM=Non-proportional hazards model, POM=Proportional odds model, NPOM=Non-proportional odds model.

	Royston and Parmar HM				Royston and Parmar OM			
	AIC		BIC		AIC		BIC	
Covariate	PHM	NPHM	PHM	NPHM	POM	NPOM	POM	NPOM
Gender	1936.6	1938.0	1959.0	1971.6	1936.1	1937.6	1958.6	1971.2
ALL type	1918.0	1902.7	1940.5	1936.4	1915.7	1902.9	1938.1	1936.6
CNS involvement	1928.6	1928.8	1951.0	1962.5	1927.6	1928.5	1950.1	1962.1
Risk group	1849.4	1827.1	1877.4	1877.6	1844.4	1827.4	1872.4	1877.9
Age	1910.7	1912.6	1933.2	1940.6	1909.9	1911.9	1932.4	1940.0

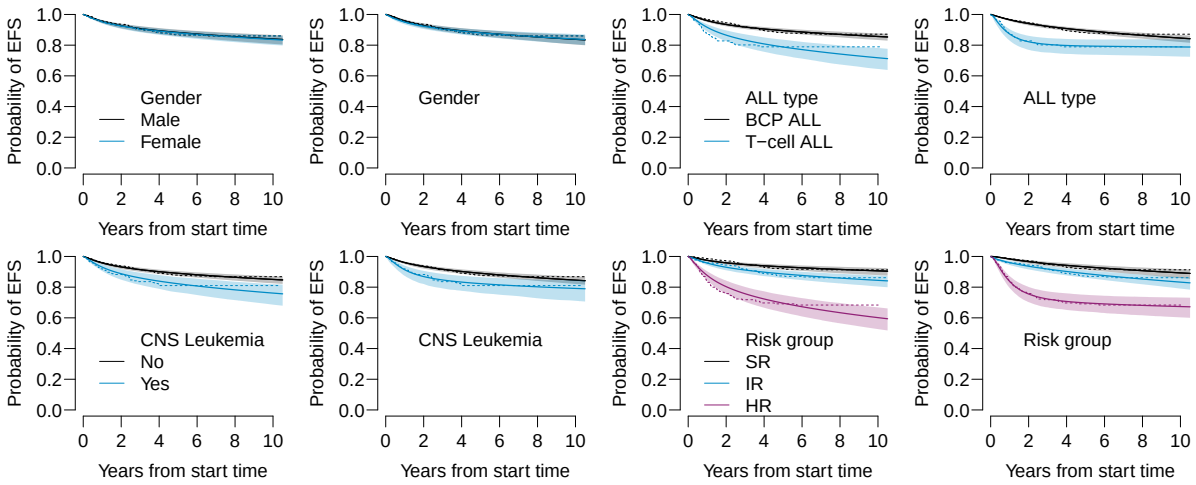

Figure 31: Survival curves estimated from univariable Royston and Parmar proportional hazards models (first and third column) and non-proportional hazards models (second and fourth column). Dashed lines show Kaplan-Meier estimates. SR=Standard risk, IR=Intermediate risk, HR=High risk.

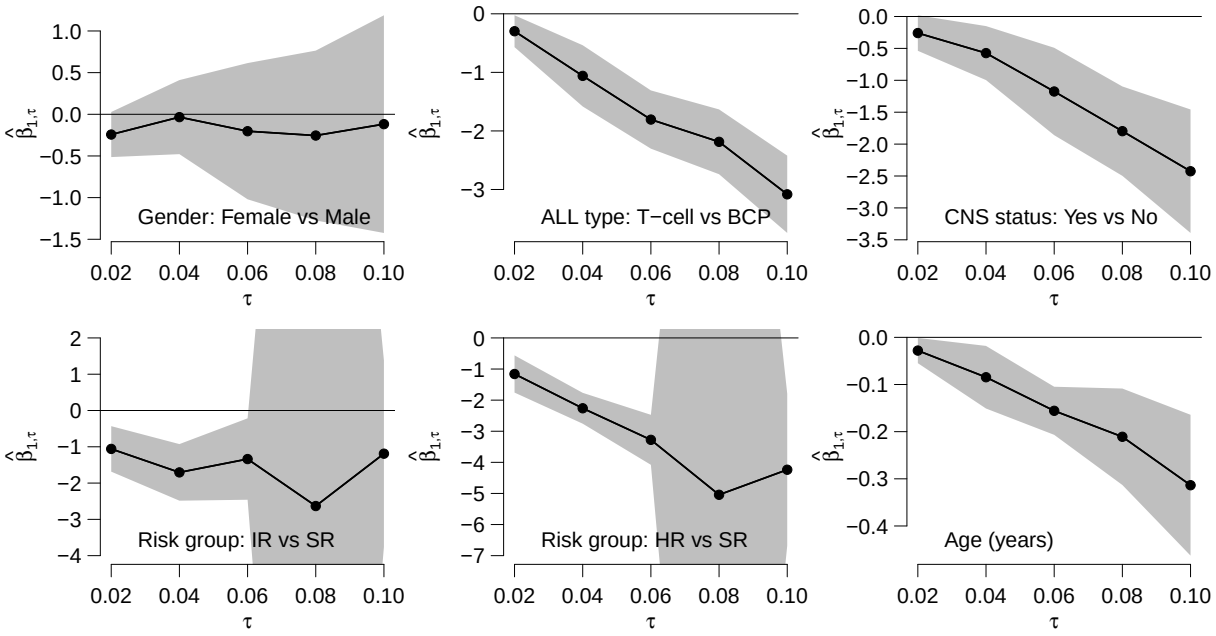

Figure 32: Plots of estimates of $\beta_{1,\tau}$ from univariable quantile regression models, for different probabilities τ. Shaded areas mark 95% confidence intervals.

5.6 Quantile regression

Univariable quantile regression models, defined in Eq. (32), were fitted for the covariates gender, ALL type, CNS involvement, risk group and age, for probabilities $\tau = 0.02, 0.04, 0.06, 0.08, 0.1$. Higher τ-values than 0.1 were not included, since the estimates became unstable and the standard errors very large for higher probabilities for the heavily right-censored ALL dataset. Estimates and 95% confidence intervals of $\beta_{1,\tau}$ from univariable quantile regression models for different τ-values are shown in Figure 32. In the univariable quantile regression for risk group, the confidence intervals of the coefficient estimate for $\tau = 0.08$ was much wider than for other τ-values, going from approximately -20 to 20. The y-axis was cut to visualize confidence intervals of estimates for other τ-values more clearly. Each quantile is modeled separately, and standard errors may vary a lot for some τ-values due to numeric instability, especially for τ-values approaching the highest observed ones. A multivariable model including the covariates listed above was also fitted. Figure 33 shows the multivariable estimates and 95% confidence intervals of $\beta_{1,\tau}$. Similarly to the univariable and multivariable Cox regression models, the ALL type covariate lost statistical significance in the presence of the other covariates.

5.7 Estimating event-free survival by ALL type with different methods

The proportional hazards assumption (PHA) of the covariate ALL type was shown in Figure 26 to be invalid. The EFS for children with BCP ALL and T-cell ALL estimated using nine different methods is shown in Figure 34. The confidence intervals for the piecewise constant hazards (PCH) model, described in Section 3.7.1, were calculated using the 2.5% and 97.5% quantiles of PCH estimated survival from 10000 bootstrap samples with random exponentially distributed weights with mean 1, using the 22 breaks from the PCH model of the unweighted childhood ALL data. Methods assuming proportional hazards or proportional odds, i.e. the parametric Weibull and log-logistic models, the Cox PHM, the Royston and Parmar (RP) PHM

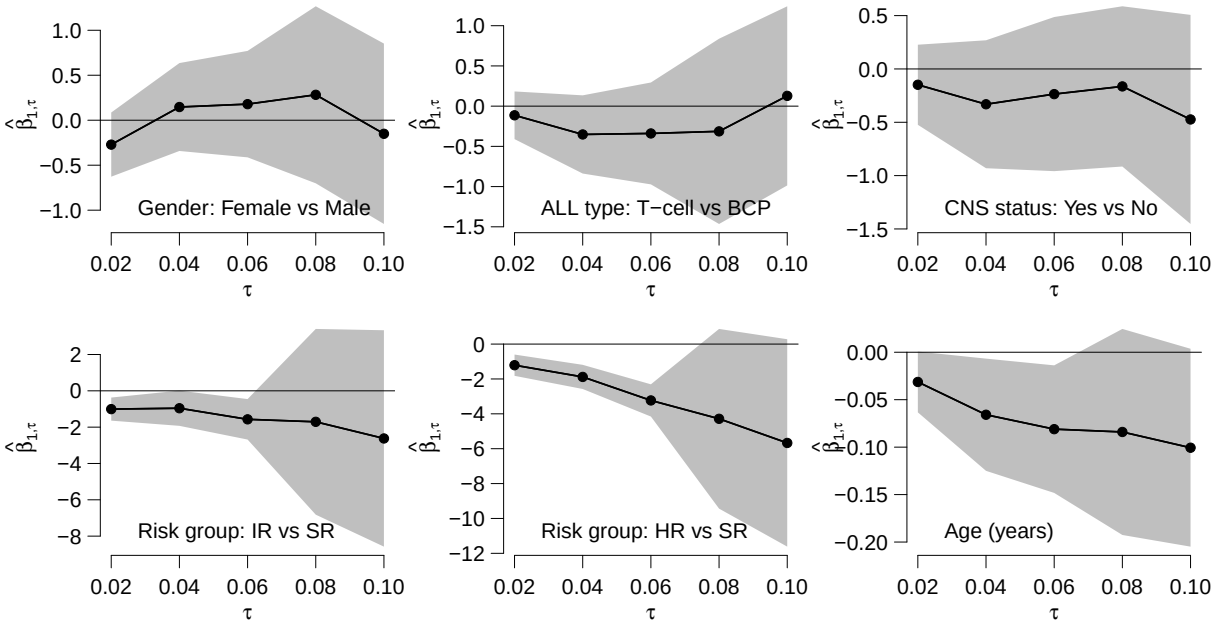

Figure 33: Plots of estimates of $\beta_{1,\tau}$ from a multivariable quantile regression model containing gender, ALL type, CNS involvement, Risk group and age, for different probabilities τ. Shaded areas mark 95% confidence intervals.

Table 12: Coefficient estimates $\hat{\beta}$, standard errors $\widehat{\text{SE}}(\hat{\beta})$, and p-values from univariable proportional hazards models (PHMs) (Cox, Weibull, Royston and Parmar) applied to the ALL data

	Cox PHM			Weibull PHM			RP PHM		
Covariate	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	p-value	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	p-value	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	p-value
Gender: F/M	0.023	0.138	0.87	0.020	0.138	0.88	0.026	0.138	0.85
ALL Type: T-cell/BCP	0.768	0.164	2.9e-06	0.762	0.164	3.5e-06	0.767	0.164	3e-06
CNS: Yes/No	0.522	0.174	0.0027	0.521	0.174	0.0028	0.522	0.174	0.0027
Risk group: IR/SR	0.533	0.175	0.0023	0.545	0.175	0.0018	0.541	0.175	0.002
Risk group: HR/SR	1.645	0.171	8e-22	1.646	0.171	7.7e-22	1.647	0.171	7.1e-22
Age (years)	0.072	0.014	1.4e-07	0.072	0.014	1.1e-07	0.072	0.014	1.2e-07

and POM, could not capture the difference in the shape of the survival curve between the two ALL subtypes seen in the Kaplan-Meier, PCH, and time-dependent RP model estimates. The RP proportional hazards and odds models are, however, better than the fully parametric models in capturing the shape of the survival function.

5.8 Comparing proportional hazards methods

In Table 12 results from univariable Cox, Weibull, and Royston and Parmar proportional hazards models are presented. Even though Figures 29 and 30 showed that a Weibull parametric distribution fitted the data poorly, the β-estimates of the Weibull PHM were not that different from the Cox and Royston and Parmar models. The standard errors were also similar between the three methods.

5.9 Comparing proportional odds methods

Results from univariable log-logistic and Royston and Parmar proportional odds models (POMs) are shown in Table 13. Also here, the estimates of the parametric log-logistic model were quite

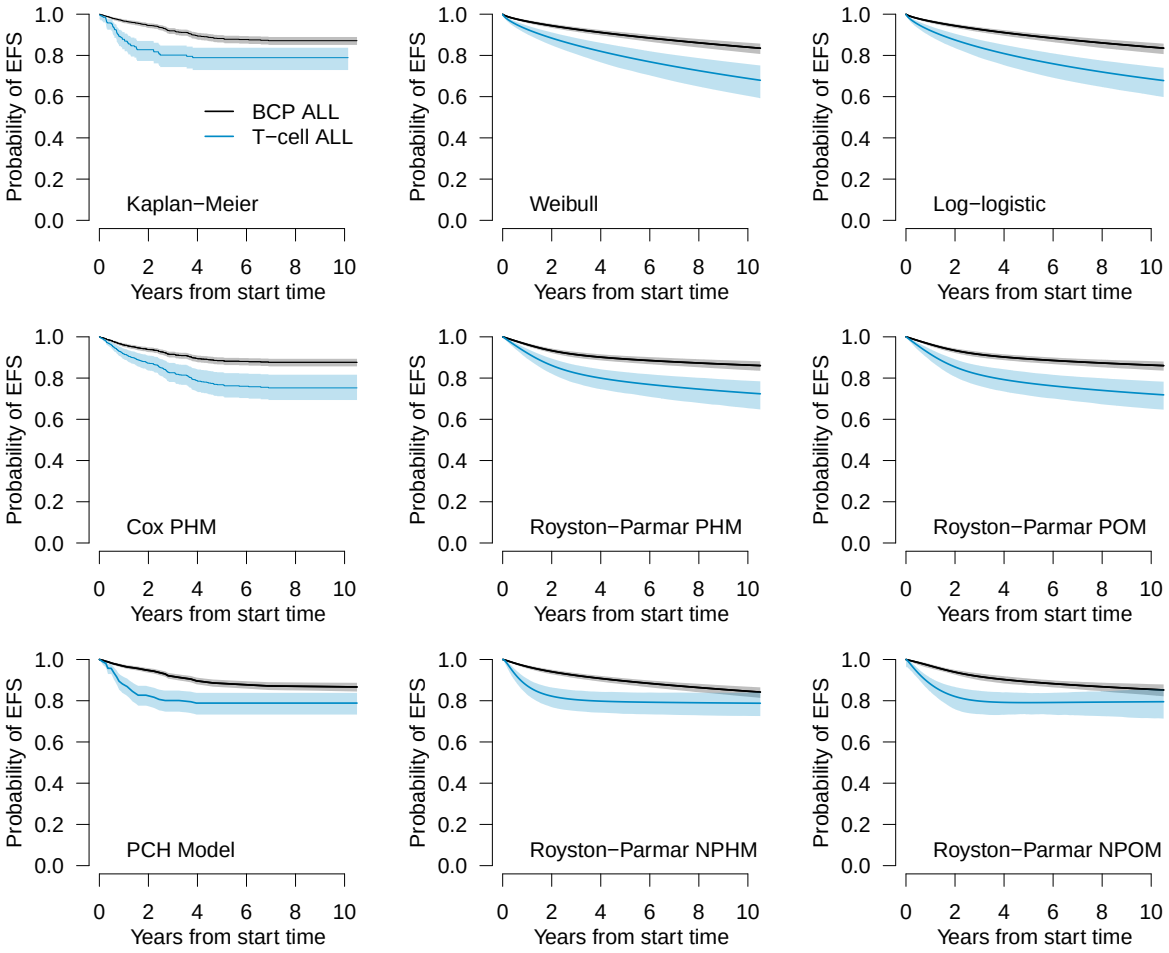

Figure 34: Event-free survival (EFS) and 95% confidence intervals for children with BCP ALL (black) and T-cell ALL (blue), estimated with nine methods. In the Royston and Parmar proportional hazards and odds models, 4 knots were used. In the Royston and Parmar non-proportional hazards and odds models (NPHM and NPOM), 3 knots were used. In the PCH model, 22 breaks were used.

Table 13: Coefficient estimates $\hat{\beta}$, standard errors $\widehat{\text{SE}}(\hat{\beta})$, and p-values from univariable proportional odds models (POM) (log-logistic, Royston and Parmar) applied to the ALL data

	Log-logistic POM			RP POM		
Covariate	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	p-value	$\hat{\beta}$	$\widehat{\text{SE}}(\hat{\beta})$	p-value
Gender: F/M	0.028	0.148	0.85	0.034	0.147	0.82
ALL Type: T-cell/BCP	0.879	0.183	1.5e-06	0.881	0.182	1.3e-06
CNS: Yes/No	0.585	0.191	0.0021	0.582	0.190	0.0022
Risk group: IR/SR	0.579	0.184	0.0016	0.572	0.183	0.0018
Risk group: HR/SR	1.862	0.190	1.1e-22	1.849	0.189	1.6e-22
Age (years)	0.079	0.015	1.1e-07	0.078	0.015	1.5e-07

similar to the estimates and standard errors of the flexible parametric Royston and Parmar model.

6 Discussion

In this thesis the reader was given an overview of a number of methods from the field of survival analysis; the non-parametric methods of Nelson-Aalen and Kaplan-Meier, the semi-parametric Cox proportional hazards model, the fully parametric accelerated failure time model, Weibull proportional hazards model, and log-logistic proportional odds model, the flexible parametric Royston and Parmar proportional hazards and odds models, and the distribution-free quantile regression. Power to detect deviations from the proportional hazards assumption, given that it was not fulfilled, and type I error rates, given that it was fulfilled, was evaluated using Monte-Carlo simulations. Given that the proportional hazards assumption was fulfilled, the three methods assuming proportional hazards were compared in terms of estimates, standard errors and confidence interval coverage, also using Monte-Carlo simulations. Coefficient estimates, standard errors and confidence interval coverage for quantile regression for a range of different probabilities was evaluated using Monte-Carlo simulations. Finally, the methods were applied to childhood acute lymphoblastic leukemia data (ALL).

The non-parametric Nelson-Aalen and Kaplan-Meier estimators are great tools for visualization and for evaluating the fit of regression models. Presenting Nelson-Aalen or Kaplan-Meier curves is an honest way of showing survival data, since no distribution or proportional hazards assumptions are made, and since the curves give an indication of the sample size – curves with large steps originate from smaller samples. The log-rank test is a good option for comparing two groups, although not suitable for crossing survival curves, and not giving the possibility to adjust for other covariates or to handle continuous covariates. Unless the sample size is too small, regression methods can preferably be applied to the data.

A very popular choice of regression model for survival data is the Cox proportional hazards model, which has the attractive property of not demanding any distribution assumptions while still being able to handle several covariates, both categorical and continuous, and giving estimates of the easily interpretable hazard ratios. Everything comes at a prize, however – in order for the hazard ratios to be a valid description of covariate effects, the proportional hazards assumption should be fulfilled. From the Monte Carlo simulations in Section 4.1 it could be seen that small deviations from this assumption are very hard to detect, especially for small sample sizes and large censoring proportions, but even with quite large sample sizes. However, for small deviations, the hazard ratio may be a "good enough" description of covariate effects. Note that, even though the proportional hazards assumption is fulfilled, the hazard ratio should not be extrapolated to describe covariate effects beyond the observed time frame. The proportional hazards assumption could very well be violated at later time points. The cumulative hazard and survival functions can be estimated and, thus, visualized from a Cox regression model, but extrapolations cannot be made, and the shape of the curves are similar to Nelson-Aalen and Kaplan-Meier curves with the difference that proportional hazards are imposed on the curves. Therefore the curves are heavily data dependent, and not smooth.

Some parametric modeling options are presented in Section 3.5. Smooth hazard, cumulative hazard, and survival curves can easily be estimated from these models. Different interpretations of the models were shown to be possible. These include the accelerated failure time interpretation, and the proportional hazards, using a Weibull distribution, and proportional odds, using a log-logistic distribution. These models are very useful when the distribution assumption holds, but extrapolations should always be made with caution. It could be that a parametric dis-

tribution is a very good description of the data in the observed time frame, but the shape of the underlying hazard function could change with time. For example in childhood cancer, the hazard function has a certain shape the first years after diagnosis, but much later on, when the cancer survivors die from other causes, the shape of the hazard is probably different.

Without making any distribution assumptions, smooth curves of the hazard, cumulative hazard, and survival functions can be made by fitting flexible parametric models to the data. In Section 3.6, the Royston and Parmar proportional hazards and proportional odds models are presented. They use the Weibull and log-logistic distributions for modeling the underlying cumulative hazard and odds, respectively, but use restricted cubic splines to come around the strictness of a fully parametric distribution assumption. Deviations from the proportional hazards and odds assumptions can easily be incorporated in these models to better represent the shapes of the hazard, cumulative hazard, and survival curves, and the coefficient estimates. However, the simple hazard ratio will no longer be time-independent, and is furthermore modeled as a function of time using splines, and is thus very hard to interpret. Time-varying coefficients in a Cox model are easier to interpret, but assumptions about the functional form of the coefficient depending on time has to be made.

Monte Carlo simulations were made in order to compare models assuming proportional hazards, given that the assumption was fulfilled and given underlying Weibull distributions. The three methods – Cox, Weibull, and Royston and Parmar proportional hazards models – proved to be quite similar in average coefficient estimates, standard errors and confidence interval coverage. However, the Royston and Parmar model sometimes failed when fitting the models, especially for high censoring proportions. Given that the proportional hazards assumption holds, and in addition that a Weibull distribution seems to fit, there is no reason not to use a Weibull proportional hazards model. An interesting extension of the simulations would be to use another distribution than Weibull and see how well the Weibull model performed in comparison to the other models. The number and placement of knots in Royston and Parmar models could also have been evaluated. In the simulations presented in this thesis, only 3 knots were used, since fitting more knots often failed, probably due to the underlying distribution being Weibull, hence, no further knots were needed.

Quantile regression for survival analysis, with the method of Frumento and Bottai (2017) presented in Section 3.7, provides another perspective to survival data modeling. Instead of effects on the hazard function, the association between covariates and the distribution quantiles are of interest. The interpretation of the coefficients are in terms of longer or shorter survival time, and is therefore direct and could very well be used in a clinical practice. The Monte Carlo simulations in Section 4.3 showed that average coefficient estimates were accurate even for small sample sizes, while confidence interval coverage was lacking for most of the investigated sample sizes. This is in line with the coverage probabilities in the simulations made by Frumento and Bottai (2017). Extrapolation cannot be made using quantile regression, since only observed quantiles can be modeled. This is not necessarily a disadvantage, since, as noted above, extrapolations should be made with much caution.

The Monte Carlo simulations presented in this thesis should not be interpreted as universally true for all situations. Only Weibull distributions were used for simulating event times, and only uniform censoring times were used. Uniform censoring times, using $U(0, \theta)$, makes sure no events are observed after $t = \theta$. This is reasonable, since most medical studies have a time frame and follow-up time limited to a certain number of years. However, with another censoring

distribution enabling later events, the simulation results would probably have been different.

In Section 5, results from applications of the above methods to the childhood ALL data are presented. Testing the proportional hazards assumption of univariable Cox models showed deviations from the assumption for the covariates ALL type and risk group. Modeling these covariates with Royston and Parmar models gave better fits and survival curves that captured the non-proportionality.

Comparing results from univariable Cox, Weibull, and Royston and Parmar proportional hazards models showed quite similar estimates and standard errors, even though a Weibull distribution did not seem to fit the data well. Univariable log-logistic and Royston and Parmar proportional odds models were also compared, and results were fairly similar, despite the log-logistic distribution not providing an accurate fit to the data. The coefficients being similar between semi-parametric, parametric and flexible parametric models is in line with a statement from Royston and Parmar (2002). This gives an indication that the coefficient estimates of fully parametric models are robust to violations of distribution assumptions. When regression coefficients are of primary interest, robustness to violations of distribution assumptions suggests that any of the models could be used. This could be further evaluated using for example Monte Carlo simulations, and probably depends on the magnitude of the violation. However, using a parametric model with poor fit of the underlying hazard distribution may be an inaccurate choice when estimating survival curves for specific combinations of covariates values. In those cases, a semi-parametric or flexible parametric model is preferred.

The age covariate is, as mentioned, often dichotomized when analyzing childhood ALL data. This implies a highly simplified and often inaccurate assumption about the association between age and survival time. Applications of the Cox and Royston and Parmar proportional hazards and odds models showed that age can very well be included as a continuous covariate with a linear assumption. This probably increases power and is a more accurate description of the association.

Event-free survival, defined as time to relapse of cancer, a second malignancy, death or censoring, whichever came first, was used as the time-to-event outcome variable in the modeling of the ALL data. A seen from the patient characteristics in Table 8, some children have a relapse of cancer and recover, while some have a relapse and then die. Event-free survival does not distinguish between the different types of events, and cannot incorporate several different types of events occurring consecutively for a single individual. One option when modeling these data could be to use multi-state models, letting individuals move between the states Alive, Relapsed, Second malignancy, and Death (being the absorbing state), see e.g. Aalen et al. (2008, Ch. 3) for an example.

The cause of death of a child suffering from ALL can be the disease itself or treatment-related mortality, since the treatment is intensive and can lead to a number of different toxicities. In this thesis no emphasis was put on the cause of death, but an option could have been to use competing risk methods, seeing death from treatment as a competing risk, preventing the observation of death from disease, and vice versa. See Aalen et al. (2008, Ch. 3) for an introduction to competing risk analysis.

In some survival datasets, there are covariates that vary with time, for example a biomarker that is measured throughout the treatment of a disease, or treatment given some time after baseline to a subgroup of patients. Such covariates can be incorporated in many of the methods covered in this thesis. This was, however, not in the scope of this thesis.

7 Conclusions

In medical research, a lot can be gained from using other methods than the standard non-parametric and semi-parametric Kaplan-Meier curves and Cox regression, even though these methods are easy to apply and interpret. In all survival regression models, model assumptions, such as the proportional hazards assumption and parametric distribution assumptions, should always be assessed and methods should be selected accordingly. The Royston and Parmar flexible parametric models provide a useful way of smoothly modeling survival data, while not relying on distribution assumptions, although coefficients modeled as functions of time are hard to interpret. Given an underlying Weibull distribution and a fulfilled proportional hazards assumption, the Cox, Weibull, and Royston and Parmar proportional hazards models all have accurate confidence interval coverage and coefficient estimates. Quantile regression offers an alternative perspective to survival data that could be practical in the communication with patients in a clinic. Coefficient estimates are accurate even for small sample sizes, but the 95% confidence interval coverage probability is too low for small and moderate sample sizes. The models presented in this thesis can be successfully used in heavily right-censored data, as shown by the application to childhood acute lymphoblastic leukemia data.

Bibliography

Aalen, O. (1975). *Statistical inference for a family of counting processes* (Unpublished doctoral dissertation). University of California, Berkeley. 3.2.1

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, *6*(4), 701–726. 3.2.1

Aalen, O., Borgan, Ø., & Gjessing, H. K. (2008). *Survival and event history analysis: a process point of view*. New York: Springer. 2.1.2, 3.1, 3.2, 3.2.3, 6

Andersen, P. K., & Gill, R. D. (1982, 12). Cox's regression model for counting processes: A large sample study. *Ann. Statist.*, *10*(4), 1100–1120. doi: 10.1214/aos/1176345976 3.4.2

Bergling, M., Billgren, C., Delaryd, C., Färlin, M., Hydén, J., Måwe, I., ... Sandberg., K. (2019). *Barncancerrapporten 2019*. Retrieved from https://www.barncancerfonden.se/informationsmaterial-och-bocker/informationsmaterial/barncancerrapporten-2019/ 2.3

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187–220. 3.4, 3.4.2

Durrleman, S., & Simon, R. (1989, May). Flexible regression models with cubic splines. *Statistics in Medicine*, *8*(5), 551–561. 3.6.1

Frumento, P. (2016a). ctqr: Censored and truncated quantile regression [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=ctqr (R package version 1.0) 3.7.2, 4.3

Frumento, P. (2016b). pch: Piecewise constant hazards models for censored and truncated data [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=pch (R package version 1.3) 3.7.1

Frumento, P., & Bottai, M. (2017). An estimating equation for censored and truncated quantile regression. *Computational Statistics and Data Analysis*, *113*, 53 - 63. doi: https://doi.org/10.1016/j.csda.2016.08.015 3.7, 3.7.1, 3.7.3, 3.7.3, 6

Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, *81*(3), 515–526. 3.4.3

Gut, A. (2009). *An intermediate course in probability* (2. ed. ed.). Dordrecht: Springer. A

Hardin, J. W. (2002). The robust variance estimator for two-stage models. *The Stata Journal*, *2*(3), 253-266. doi: 10.1177/1536867X0200200302 3.7.3

Harrell, F. E., Jr. (2015). *Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed. 2015. ed.). Cham: Springer International Publishing. 3.6.1

Harrell Jr, F. E. (2019). rms: Regression modeling strategies [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=rms (R package version 5.1-3.1) 3.6.1

Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: regression modeling of time-to-event data* (2nd ed. ed.). Hoboken, N.J.: Wiley-Interscience. 2.1.1, 3.2, 3.2.3, 3.3, 3.4

Hubeaux, S., & Rufibach, K. (2015). SurvRegCensCov: Weibull regression for a right-censored endpoint with interval-censored covariate [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=SurvRegCensCov (R package version 1.4) 4.2

Hunger, S. P., & Mullighan, C. G. (2015). Acute lymphoblastic leukemia in children. *New England Journal of Medicine*, *373*(16), 1541-1552. (PMID: 26465987) doi: 10.1056/NEJMra1400972 2.3.1

Inaba, H., Greaves, M., & Mullighan, C. G. (2013). Acute lymphoblastic leukaemia. *The Lancet*, *381*(9881), 1943 - 1955. doi: https://doi.org/10.1016/S0140-6736(12)62187-4 2.3.1

Jackson, C. (2016). flexsurv: A platform for parametric survival modeling in R. *Journal of Statistical Software*, *70*(8), 1–33. doi: 10.18637/jss.v070.i08 4.2

Kaatsch, P. (2010). Epidemiology of childhood cancer. *Cancer Treatment Reviews*, *36*(4), 277 - 285. (Cancer in Childhood) doi: https://doi.org/10.1016/j.ctrv.2010.02.003 2.3, 2.3.1

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457–481. 3.2.2

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data* (2. ed. ed.). New York: Springer. 2.1.1, 3.4.2, 3.4.3, 3.5.1, 3.5.1, 3.5.2, 3.5.2, 3.5.4, 3.5.5, 3.5.6, A

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50. 3.7

Li, H., Han, D., Hou, Y., Chen, H., & Chen, Z. (2015). Statistical inference methods for two crossing survival curves: A comparison of methods. *PLOS ONE*, *10*(1), 1-18. doi: 10.1371/journal.pone.0116774 3.2.3

Martinussen, T., & Peng, L. (2014). Alternatives to the Cox Model. In J. P. Klein, H. C. van Houwelingen, J. G. Ibrahim, & T. H. Scheike (Eds.), *Handbook of survival analysis* (p. 49-76). Boca Raton: CRC Press. 3.7.2, 3.7.3

Nelson, W. (1969). Hazard plotting for incomplete failure data. *Journal of Quality Technology*, *1*(1), 27-52. doi: 10.1080/00224065.1969.11980344 3.2.1

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, *14*(4), 945–966. 3.2.1

Nelson, W. E., & Behrman, R. E. (1996). *Nelson textbook of pediatrics* (15. ed. ed.). Philadelphia: Saunders. 2.3.1

Peng, L., & Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, *103*(482), 637-649. doi: 10.1198/016214508000000355 3.7

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, *98*(464), 1001-1012. doi: 10.1198/016214503000000954 3.7

R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/` 3.6.1

Reid, N. (1994). A conversation with sir David Cox. *Statistical Science*, *9*(3), 439–455. 3.5

Royston, P., & Parmar, M. K. B. (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. doi: 10.1002/sim.1203 3.6, 3.6.1, 3.6.2, 5.3.3, 6

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, *69*(1), 239–241. 3.4.3

Therneau, T. M. (2015). A package for survival analysis in S [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=survival` (version 2.38) 4.1, 4.2, 5.2, 5.3.2

Tobacman, J. (2011, 08). Paul Meier, 1924-2011. *Chicago Tribune*. `https://www.chicagotribune.com/news/ct-xpm-2011-08-18-ct-met-meier-obit-20110818-story.html`. (Published online August 8, 2011; accessed March 4, 2019) 3.2.2

Toft, N., Birgens, H., Abrahamsson, J., Griškevičius, L., Hallböök, H., Heyman, M., ... Schmiegelow, K. (2018, 03). Results of NOPHO ALL2008 treatment for patients aged 1-45 years with acute lymphoblastic leukemia. *Leukemia*, *32*(3), 606–615. 2.3.1, 5.3.3

Wang, H. J., & Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association*, *104*(487), 1117-1128. doi: 10.1198/jasa.2009.tm08230 3.7

Appendices

A Additional parametric distributions

The density, survival and hazard functions of the Gamma distribution are presented in Table A1 and shown in Figure A1. The survival and hazard functions are not possible to derive without using the lower incomplete Gamma function, not stated explicitly here. As seen in the figure, the hazard function $h(t)$ approaches λ as t tends to infinity (Klein & Moeschberger, 2003, Ch. 2.5, p. 42). Like the Weibull distribution, $h(t)$ is strictly decreasing for $\beta < 1$, strictly increasing for $\beta > 1$, and constant for $\beta = 1$. If n is a positive integer, then $\Gamma(n, \lambda)$ is the sum of n random variables from an exponential distribution with parameter λ (Gut, 2009, Ch. 3, p. 67-68).

The density, survival and hazard functions of the Gompertz distribution are presented in Table A1 and shown in Figure A2. The hazard function of the Gompertz distribution is exponential, and thus tends to infinity as t tends to infinity. The parameter θ determines the hazard at time $t = 0$. In biology, an exponential hazard function has been deemed appropriate for some applications (Klein & Moeschberger, 2003, Ch. 2.8, p. 58).

Table A1: Additional parametric distributions used in survival analysis

Distribution	**Parameters**	$\boldsymbol{f(t)}$	$\boldsymbol{S(t)}$	$\boldsymbol{h(t)}$
Gamma $T \sim \Gamma(\beta, \lambda)$	$\beta, \lambda > 0$	$\frac{\lambda^\beta}{\Gamma(\beta)} t^{\beta-1} e^{-\lambda t}$(*)	$1 - \int_0^t f(s)ds$	$\frac{f(t)}{S(t)}$
Gompertz $T \sim \text{Go}(\alpha, \theta)$	$\alpha, \theta > 0$	$\theta e^{\alpha t} \exp\left\{\frac{\theta}{\alpha}(1 - e^{\alpha t})\right\}$	$\exp\left\{\frac{\theta}{\alpha}(1 - e^{\alpha t})\right\}$	$\theta e^{\alpha t}$

*) $\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x)dx$. If $n \in \mathbb{N}, \Gamma(n) = (n-1)!$

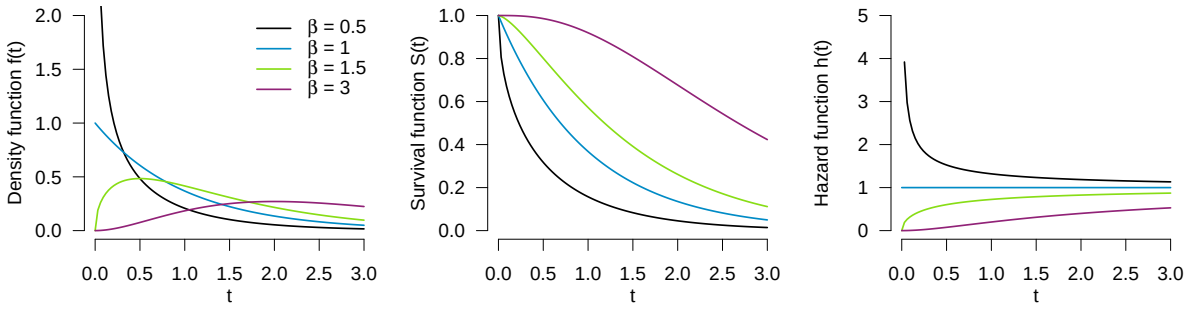

Figure A1: The Gamma distribution; density function, survival function and hazard function, for varying shape parameters β. The scale parameter $\lambda = 1$ for all β.

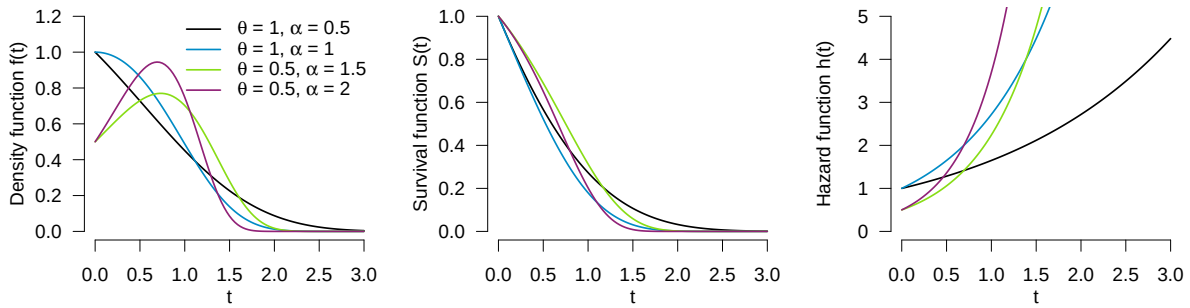

Figure A2: The Gompertz distribution; density function, survival function and hazard function, for varying shape and scale parameters, α and θ.

B Supplementary simulation results for comparison of proportional hazards methods

Table B1: Number of failures in fitting, and number of non-convergences, out of 10000 simulations, presented in Section 4.2, for comparison between PH methods for different censoring probabilities and sample sizes n in each group (total sample size given by $2n$). NEOG=no events in at least one of the groups, resulting in non-convergence with all three methods. NC denotes the occurrence of non-convergence even though there are observed events in both groups. Failures refer to failure in fitting for other reasons than convergence.

No censoring

		Cox PHM		Weibull PHM		RP PHM	
n	NEOG	No. of NCs	No. of fail.	No. of NCs	No. of fail.	No. of NCs	No. of fail.
25	0	0	0	0	0	0	3
50	0	0	0	2	0	0	3
100	0	0	0	18	0	0	10
250	0	0	0	106	0	0	22
500	0	0	0	252	0	0	54

30% censoring

		Cox PHM		Weibull PHM		RP PHM	
n	NEOG	No. of NCs	No. of fail.	No. of NCs	No. of fail.	No. of NCs	No. of fail.
25	0	0	0	0	0	0	9
50	0	0	0	0	0	0	3
100	0	0	0	0	0	0	10
250	0	0	0	0	0	0	22
500	0	0	0	0	0	0	54

60% censoring

		Cox PHM		Weibull PHM		RP PHM	
n	NEOG	No. of NCs	No. of fail.	No. of NCs	No. of fail.	No. of NCs	No. of fail.
25	3	0	0	0	0	0	159
50	0	0	0	0	0	0	5
100	0	0	0	0	0	0	10
250	0	0	0	0	0	0	22
500	0	0	0	0	0	0	54

90% censoring

		Cox PHM		Weibull PHM		RP PHM	
n	NEOG	No. of NCs	No. of fail.	No. of NCs	No. of fail.	No. of NCs	No. of fail.
25	1940	8	30	0	0	12	1515
50	290	1	37	0	0	1	1357
100	7	0	20	0	0	0	285
250	0	0	6	0	0	0	28
500	0	0	0	0	0	0	55

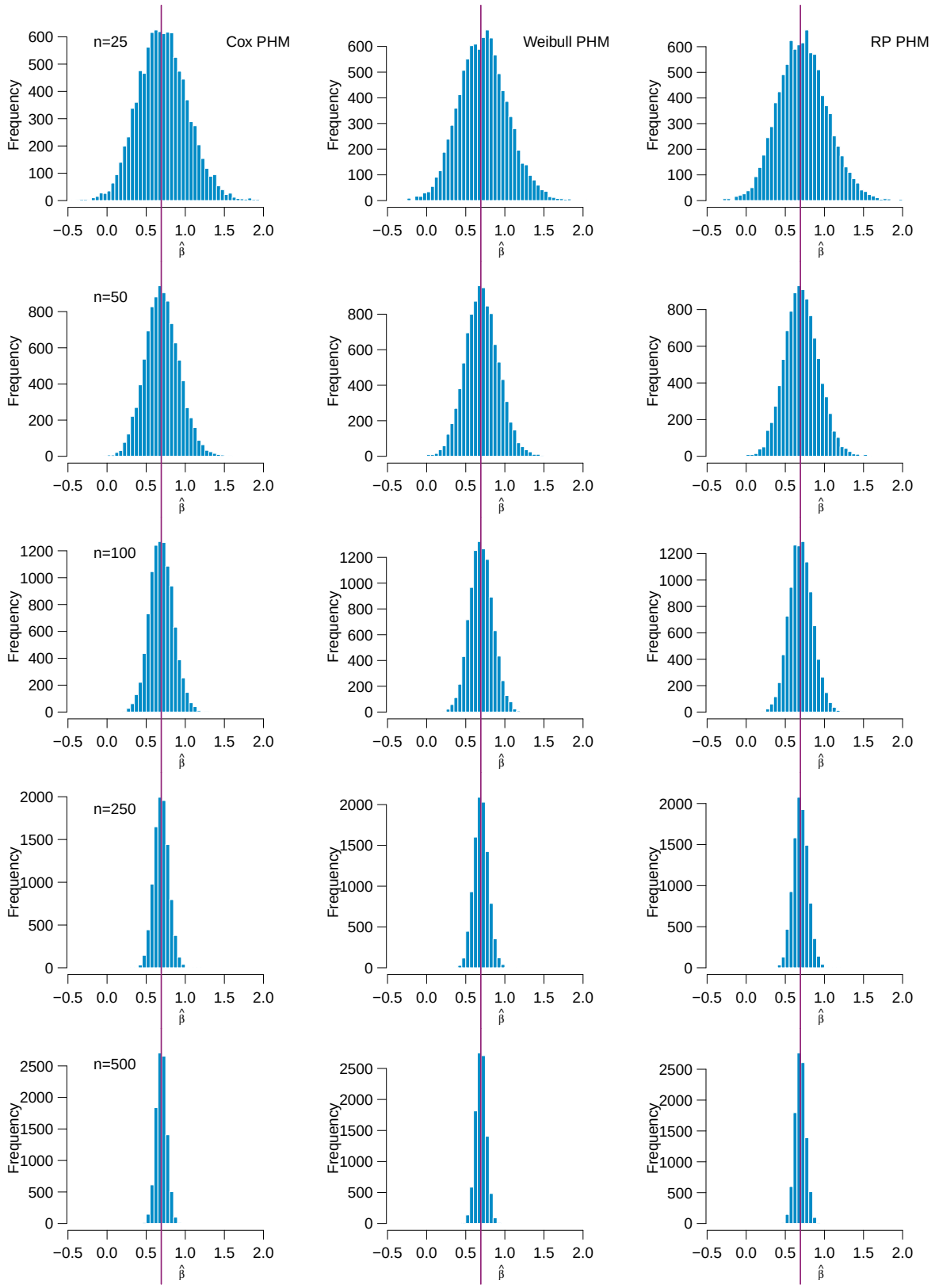

Figure B1: Histograms of $\hat{\beta}$s in 10000 simulations with no censoring, presented in Section 4.2, for comparison between PH methods for different sample sizes n in each group (total sample size given by $2n$). The left column shows estimates from Cox PHMs, the middle from Weibull PHMs, and the right from Royston and Parmar PHMs. Vertical line indicates true $\beta \approx 0.69$.

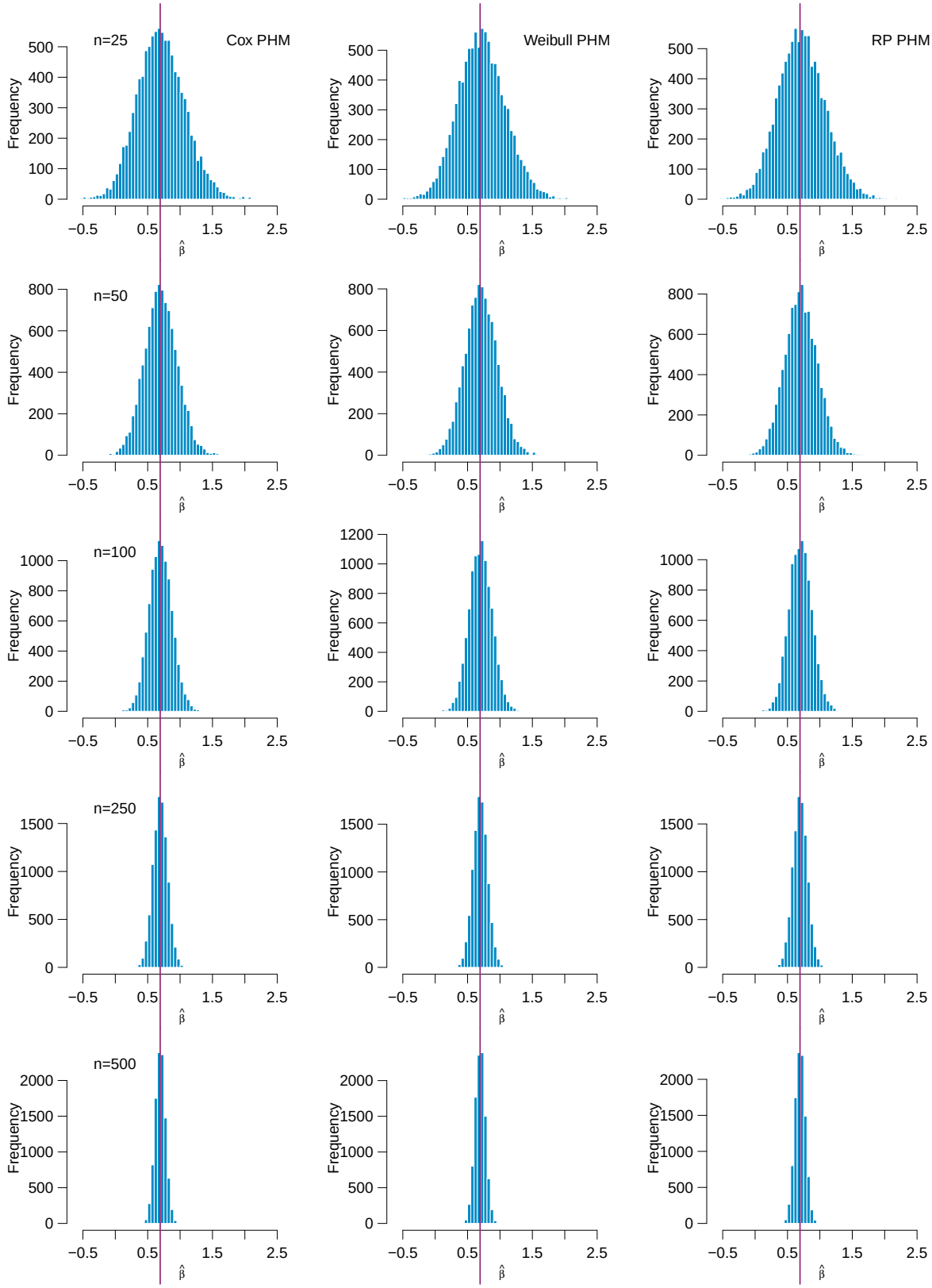

Figure B2: Histograms of $\hat{\beta}$s in 10000 simulations with 30% censoring, presented in Section 4.2, for comparison between PH methods for different sample sizes n in each group (total sample size given by $2n$). The left column shows estimates from Cox PHMs, the middle from Weibull PHMs, and the right from Royston and Parmar PHMs. Vertical line indicates true $\beta \approx 0.69$.

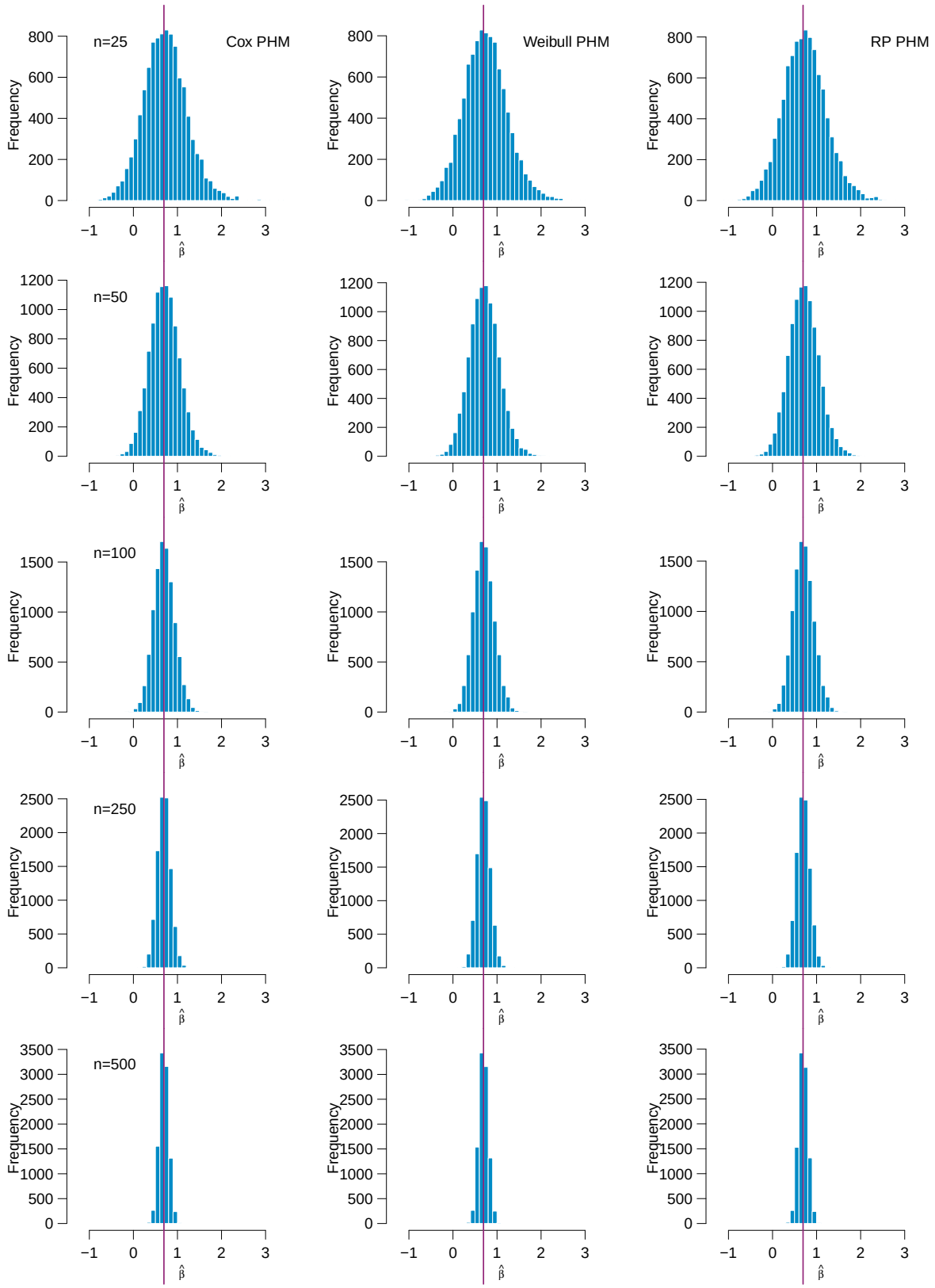

Figure B3: Histograms of $\hat{\beta}$s in 10000 simulations with 60% censoring, presented in Section 4.2, for comparison between PH methods for different sample sizes n in each group (total sample size given by $2n$). The left column shows estimates from Cox PHMs, the middle from Weibull PHMs, and the right from Royston and Parmar PHMs. Vertical line indicates true $\beta \approx 0.69$.

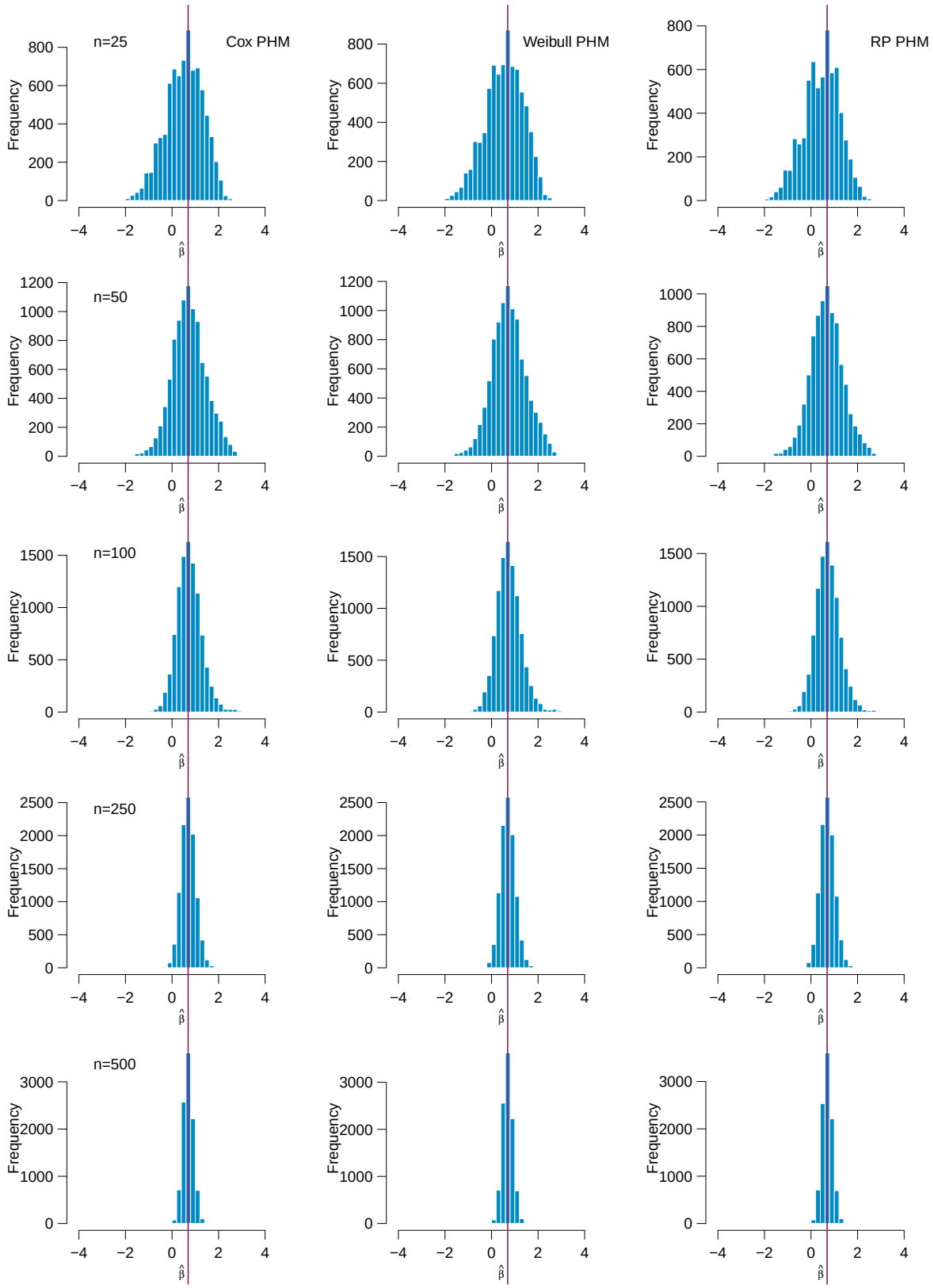

Figure B4: Histograms of $\hat{\beta}$s in 10000 simulations with 90% censoring, presented in Section 4.2, for comparison between PH methods for different sample sizes n in each group (total sample size given by $2n$). The left column shows estimates from Cox PHMs, the middle from Weibull PHMs, and the right from Royston and Parmar PHMs. Vertical line indicates true $\beta \approx 0.69$.

C Supplementary simulation results for quantile regression

Table C1: Average $\hat{\beta}_{1,\tau}$ and $\widehat{\text{SE}}(\hat{\beta}_{1,\tau})$ from 5000 simulations of quantile regression models, presented in Section 4.3, for different values of τ, sample sizes n in each group, and censoring proportions p_c. The left part of the table corresponds to a Weibull distribution with $\alpha = 0.8, \lambda = 2$, compared to the reference distribution (Weib($\alpha = 0.8, \lambda = 1$)). The right part corresponds to a Weibull distribution with $\alpha = 1.6, \lambda = 2$, compared to the reference distribution. True coefficients denoted by $\beta_{1,\tau}$ for the different scenarios and probabilities τ.

		Weib($\alpha = 0.8, \lambda = 2$)					Weib($\alpha = 1.6, \lambda = 2$)				
		$p_c = 0$			$p_c = 0.3$		$p_c = 0$			$p_c = 0.3$	
	τ	$\beta_{1,\tau}$	$\hat{\beta}_{1,\tau}$	$\widehat{\text{SE}}(\hat{\beta}_{1,\tau})$	$\hat{\beta}_{1,\tau}$	$\widehat{\text{SE}}(\hat{\beta}_{1,\tau})$	$\beta_{1,\tau}$	$\hat{\beta}_{1,\tau}$	$\widehat{\text{SE}}(\hat{\beta}_{1,\tau})$	$\hat{\beta}_{1,\tau}$	$\widehat{\text{SE}}(\hat{\beta}_{1,\tau})$
n=25	0.05	-0.0141	-0.0399	0.1713	-0.0279	0.1194	0.0769	0.0863	0.1314	0.0732	0.1433
	0.10	-0.0348	-0.0478	0.1290	-0.0435	0.1570	0.0988	0.0811	0.1182	0.0816	0.1616
	0.15	-0.0598	-0.0684	0.1685	-0.0674	0.2263	0.1051	0.0848	0.1484	0.0860	0.1718
	0.20	-0.0889	-0.0976	0.2022	-0.0982	0.2411	0.1006	0.0780	0.1709	0.0826	0.2195
	0.25	-0.1221	-0.1322	0.2540	-0.1313	0.2785	0.0869	0.0685	0.1744	0.0715	0.2428
	0.30	-0.1597	-0.1698	0.2370	-0.1677	0.3194	0.0648	0.0457	0.1990	0.0479	0.2834
	0.35	-0.2023	-0.2118	0.2756	-0.2094	0.3716	0.0341	0.0144	0.2486	0.0170	0.2906
	0.40	-0.2503	-0.2602	0.2916	-0.2539	0.3889	-0.0057	-0.0257	0.2770	-0.0208	0.3343
	0.45	-0.3047	-0.3117	0.3434	-0.3070	0.4497	-0.0556	-0.0734	0.2931	-0.0679	0.3925
	0.50	-0.3665	-0.3655	0.3794	-0.3571	0.5346	-0.1168	-0.1279	0.3579	-0.1203	0.4509
	0.55	-0.4375	-0.4354	0.4274	-0.4183	0.5552	-0.1915	-0.1983	0.3861	-0.1838	0.5165
	0.60	-0.5196	-0.5213	0.4572	-0.4784	0.6569	-0.2825	-0.2917	0.4471	-0.2598	0.5925
n=50	0.05	-0.0141	-0.0217	0.0306	-0.0203	0.0322	0.0769	0.0766	0.0552	0.0755	0.0536
	0.10	-0.0348	-0.0399	0.0437	-0.0399	0.0537	0.0988	0.0926	0.0555	0.0939	0.0609
	0.15	-0.0598	-0.0668	0.0606	-0.0666	0.0789	0.1051	0.0976	0.0633	0.0996	0.0702
	0.20	-0.0889	-0.0966	0.0831	-0.0972	0.0991	0.1006	0.0915	0.0745	0.0911	0.0828
	0.25	-0.1221	-0.1303	0.0881	-0.1316	0.1162	0.0869	0.0763	0.0869	0.0766	0.0981
	0.30	-0.1597	-0.1692	0.1052	-0.1698	0.1406	0.0648	0.0534	0.1001	0.0539	0.1180
	0.35	-0.2023	-0.2109	0.1260	-0.2126	0.1732	0.0341	0.0236	0.1131	0.0219	0.1300
	0.40	-0.2503	-0.2597	0.1372	-0.2613	0.1894	-0.0057	-0.0172	0.1348	-0.0187	0.1641
	0.45	-0.3047	-0.3126	0.1661	-0.3162	0.2350	-0.0556	-0.0648	0.1466	-0.0690	0.1863
	0.50	-0.3665	-0.3751	0.1912	-0.3768	0.2716	-0.1168	-0.1262	0.1744	-0.1285	0.2304
	0.55	-0.4375	-0.4448	0.2140	-0.4441	0.3112	-0.1915	-0.2005	0.2056	-0.2015	0.2898
	0.60	-0.5196	-0.5243	0.2412	-0.5230	0.3965	-0.2825	-0.2904	0.2429	-0.2891	0.3558
n=100	0.05	-0.0141	-0.0162	0.0160	-0.0158	0.0156	0.0769	0.0765	0.0325	0.0766	0.0318
	0.10	-0.0348	-0.0375	0.0251	-0.0376	0.0254	0.0988	0.0968	0.0383	0.0975	0.0385
	0.15	-0.0598	-0.0630	0.0345	-0.0632	0.0363	0.1051	0.1017	0.0452	0.1018	0.0462
	0.20	-0.0889	-0.0923	0.0444	-0.0922	0.0460	0.1006	0.0956	0.0530	0.0959	0.0548
	0.25	-0.1221	-0.1256	0.0543	-0.1258	0.0597	0.0869	0.0816	0.0614	0.0821	0.0639
	0.30	-0.1597	-0.1643	0.0658	-0.1649	0.0726	0.0648	0.0585	0.0706	0.0588	0.0742
	0.35	-0.2023	-0.2069	0.0772	-0.2076	0.0853	0.0341	0.0278	0.0798	0.0277	0.0837
	0.40	-0.2503	-0.2542	0.0882	-0.2560	0.1008	-0.0057	-0.0118	0.0903	-0.0121	0.0965
	0.45	-0.3047	-0.3081	0.1020	-0.3109	0.1191	-0.0556	-0.0612	0.1030	-0.0625	0.1108
	0.50	-0.3665	-0.3700	0.1175	-0.3744	0.1400	-0.1168	-0.1223	0.1156	-0.1236	0.1294
	0.55	-0.4375	-0.4413	0.1348	-0.4458	0.1661	-0.1915	-0.1970	0.1314	-0.1986	0.1623
	0.60	-0.5196	-0.5204	0.1532	-0.5262	0.2065	-0.2825	-0.2859	0.1507	-0.2929	0.1990
n=250	0.05	-0.0141	-0.0151	0.0101	-0.0150	0.0098	0.0769	0.0770	0.0211	0.0771	0.0207
	0.10	-0.0348	-0.0360	0.0160	-0.0359	0.0160	0.0988	0.0981	0.0249	0.0981	0.0250
	0.15	-0.0598	-0.0612	0.0222	-0.0613	0.0224	0.1051	0.1040	0.0293	0.1043	0.0300
	0.20	-0.0889	-0.0906	0.0284	-0.0906	0.0287	0.1006	0.0988	0.0343	0.0991	0.0355
	0.25	-0.1221	-0.1238	0.0353	-0.1241	0.0359	0.0869	0.0850	0.0397	0.0852	0.0412
	0.30	-0.1597	-0.1614	0.0423	-0.1617	0.0434	0.0648	0.0627	0.0455	0.0627	0.0474
	0.35	-0.2023	-0.2041	0.0493	-0.2047	0.0517	0.0341	0.0317	0.0518	0.0314	0.0541
	0.40	-0.2503	-0.2520	0.0571	-0.2527	0.0605	-0.0057	-0.0085	0.0589	-0.0088	0.0619
	0.45	-0.3047	-0.3062	0.0658	-0.3077	0.0700	-0.0556	-0.0579	0.0669	-0.0586	0.0706
	0.50	-0.3665	-0.3682	0.0762	-0.3698	0.0820	-0.1168	-0.1193	0.0759	-0.1203	0.0812
	0.55	-0.4375	-0.4394	0.0874	-0.4417	0.0962	-0.1915	-0.1948	0.0858	-0.1956	0.0935
	0.60	-0.5196	-0.5216	0.0993	-0.5246	0.1135	-0.2825	-0.2861	0.0972	-0.2873	0.1090
n=500	0.05	-0.0141	-0.0146	0.0071	-0.0146	0.0070	0.0769	0.0772	0.0151	0.0772	0.0150
	0.10	-0.0348	-0.0352	0.0113	-0.0352	0.0114	0.0988	0.0991	0.0176	0.0992	0.0178
	0.15	-0.0598	-0.0601	0.0158	-0.0601	0.0160	0.1051	0.1054	0.0209	0.1055	0.0214
	0.20	-0.0889	-0.0894	0.0204	-0.0894	0.0206	0.1006	0.1003	0.0245	0.1004	0.0253
	0.25	-0.1221	-0.1225	0.0253	-0.1225	0.0256	0.0869	0.0864	0.0283	0.0865	0.0295
	0.30	-0.1597	-0.1604	0.0304	-0.1603	0.0312	0.0648	0.0639	0.0326	0.0640	0.0340
	0.35	-0.2023	-0.2033	0.0354	-0.2033	0.0369	0.0341	0.0329	0.0372	0.0328	0.0388
	0.40	-0.2503	-0.2511	0.0408	-0.2513	0.0430	-0.0057	-0.0069	0.0422	-0.0072	0.0441
	0.45	-0.3047	-0.3057	0.0470	-0.3057	0.0497	-0.0556	-0.0568	0.0478	-0.0574	0.0506
	0.50	-0.3665	-0.3669	0.0542	-0.3672	0.0577	-0.1168	-0.1177	0.0541	-0.1183	0.0579
	0.55	-0.4375	-0.4379	0.0625	-0.4376	0.0681	-0.1915	-0.1919	0.0616	-0.1931	0.0671
	0.60	-0.5196	-0.5199	0.0713	-0.5206	0.0805	-0.2825	-0.2832	0.0700	-0.2844	0.0779