Statistical Models to Capture the Association Between Progression-free and Overall Survival in Oncology Trials

Enya Maria Weber

Submitted for the degree of Doctor of Philosophy

at Lancaster University

January 2020

Abstract

In oncology trials, different clinical endpoints can be measured. For the survival analysis of patients, the most traditional primary endpoint is overall survival (OS), which is defined as the time from study entry to death from any cause. Besides, progression-free related measurements such as progression-free survival (PFS) might be also considered. For assessing the performance of therapies, OS is the most reliable endpoint. However, utilizing earlier endpoints such as information from disease progression might lead to a gain in efficiency. However, the gain in efficiency might depend on the relationship between those two endpoints.

This thesis explores various statistical models for capturing the association between PFS and OS. The research is partitioned into three topics. At first, it considers methods for quantifying the association between PFS and OS in oncology trials, in terms of Kendall's τ rank correlation rather than Pearson correlation. Copula-based, non-parametric, and illness-death model–based methods are reviewed. In addition, the approach based on an underlying illness-death model is generalized to allow general parametric models. The simulations suggest that the illness-death model–based method provides good estimates of Kendall's τ across several scenarios. In some situations, copula-based methods perform well but their performance is sensitive to the choice of copula. The Clayton copula is most appropriate in scenarios which might realistically reflect an oncology trial, but the use of copula models in practice is questionable. In the second and third topic, the estimation of the group difference faces the issue of non-proportionality for treatments effects. Instead of the standard hazard ratio we use the average hazard ratio for estimating the group difference as it is able to cope with non-proportional hazards well as it considers group difference depending on time. Subsequently, it compares methods for jointly mod-

elling time-to-progression and time-to-death within a Bayesian framework. By incorporating treatment effects, we investigate an illness-death model-based approach and also copula-based approaches. According to the simulations results the Gaussian copula-based model performed the best overall, but the illness-death model-based approach showed a good performance as well. However, in contrast to the good performance of the Clayton copula-based approach in the first topic, the Clayton copula model did not perform well regarding the estimation of AHR. The third topic explores various semi-parametric multi-state model-based methods for gaining efficiency in testing for, and estimating the treatment effects in terms on, overall survival in oncology trials compared to standard methods based on directly applying Cox regression or the log-rank test. The semi-parametric multi-state model-based method fits a Cox model to (a subset of) transition intensities in an illness-death model assuming either a Markov or semi-Markov model and uses AHR to measure treatment effect. In most of the situations, the semi-parametric multi-state model-based methods perform better than the Cox-based approach. The performance of the methods in each topic is investigated by simulations and also illustrated using data from a clinical trial of treatments for advanced ovarian cancer in topic 2 and for colon cancer in topics 1 and 3.

To my dad.

Acknowledgements

First and foremost I would like to thank my supervisor Dr Andrew Titman for his great support throughout the PhD process. I am very grateful for the work with him and all the knowledge he shared with me. His guidance and patience are highly appreciated. I would like to thank Dr Simon Wandel and Dr Satrajit Roychoudhury for the joint research, which has contributed to the research in Chapter 4. I had the pleasure to work with Dr Simon Wandel during an internship at Novartis in Basel and I thank him for his encouragement and support.

I am very grateful for the European Union's Horizon2020research funding and innovation programme which has made my PhD experience possible. I have been very fortunate to be a part of the Marie Skłodowska-Curie Initial Training Network.

I also would like to acknowledge all the friends which have made my experience throughout the PhD time and my life in Lancaster so valuable. I am eternally grateful to my best friend Alena for her support and for the wonderful friendship despite the distance. All the Skype calls over the last years have been blessings.

Finally, I would like to express my deepest gratitude to my mum and dad for their unconditional love and support. Their constant encouragement and ongoing belief in me have been a motivation to complete the PhD. They have always been there for me and I am truly grateful for that. I hope they know how valuable their help has been in this process.

Declaration

I hereby declare that the work in this thesis has been done by myself and has not been submitted for a higher degree elsewhere.

The research in Chapter 3 has been published for publication as Weber, E.M., Titman, A. (2018) Quantifying the association between progression-free survival and overall survival in oncology trials using Kendall's τ. *Statistics in Medicine.*

Enya Maria Weber

Contents

Chapter

List of Tables

Table

List of Figures

Figure

CHAPTER 1

Introduction

The topic of this thesis relates to research within the area of survival analysis in oncology trials. In general, developing new therapies in cancer trials might require a comprehensive analysis of the survival history of patients. In assessing the performance of therapies in cancer trials, different clinical endpoints can be considered. The most reliable and commonly used primary endpoint is overall survival (OS), which is defined as the time from study entry to death from any cause. However, for analyzing overall survival of the patients, the occurrence of progression, different clinical intermediate outcome variables such as surrogate endpoints or auxiliary variables for the overall analysis might potentially be relevant. The occurrence of progression, defined as the growth or spread of cancer, can also be considered as a possible endpoint. Progression-free survival (PFS) is defined as the time from study entry until progression or death, depending on what occurs first.

The thesis discusses approaches to quantifying the association between PFS and OS. Furthermore, joint models of progression-free related measurements and overall survival within a Bayesian framework and the incorporation of treatment effects on the survival process are explored. Moreover, a review of methods for gaining efficiency by using information on time-to-progression to estimate the treatment effect on overall survival will be a further focus in the thesis.

The research topic in this thesis is motivated by the generally growing interest in designing more efficient trials both by reducing the number of patients required and the required follow-up time. Due to the relevance of the development of new cancer therapies, efficient and improved design in cancer trials are needed. Within this context, the use of the information on PFS has been of interest for the design and analysis of a randomized clinical trials over the last several decades. In particular, use of a surrogate marker as a replacement for the true primary endpoint have been a research-related issue in such trials. The benefit of

using PFS as the primary surrogate for analyzing survival times in a cancer trial might be potentially cost-effective and time-efficient, as long follow-up periods after progression of the disease can be avoided.

In 1989, Prentice[61] proposed requirements for the validation of surrogates in randomized clinical trials, such that the treatment comparison based on a surrogate response variable reflects the true endpoint treatment comparison well. The criterion considers the surrogate endpoint as a response variable for which a test of the null hypothesis of no association between treatment and true endpoint is valid. In other words, the surrogate should be able to capture any dependence of true response on treatment and not to provide some quantitative information on the comparison of true endpoint rates among treatments. The definition and operational criteria in the paper of Prentice[61] have resulted in further discussion and extensions, amongst others by Freenman, Graubard and Schaztkin,[31] Buyse and Molenberghs.[13] However, all those operational criteria were restricted to the validation of a surrogate endpoint of interest on the basis of one randomized clinical trial. Therefore, approaches to allow a potentially more powerful validation of a surrogate endpoint based on one more than a single randomized clinical trial were proposed by Buyse *et al*,[14] Gail *et al*,[33] Daniel *et al*[22] and Alonso *et al*.[1] Those approaches within a meta-analytic framework induced a modified validity of a surrogate endpoint depending on the association between the surrogate and true endpoint based on both the trial level and individual level.

In some cancer therapies, PFS has been accepted as a suitable surrogate endpoint to replace OS, especially in earlier phases of the drug development.[24,72] For instance, within a meta-analytic framework PFS has been approved to be a suitable surrogate endpoint for survival in patients with advanced colorectal cancer.[12] Furthermore, a meta-analysis has also determined the usefulness of progression-free survival in recurrent gastric cancer trial.[60] However, the success of establishing PFS as a surrogate for OS in oncology trials seems to depend

on the type of cancer aside the formal validation studies.[67] For instance, PFS has not been approved to be a surrogate endpoint for overall survival in metastatic breast cancer.[9,11,67] One reason might be that progression in some specific cancer is difficult to detect and identify.

However, instead of using a surrogate endpoint with the intention to replace the true outcome of interest, auxiliary variables such as the information about the time-to-progression can be used for an improved efficiency in analyzing overall survival. In general, methods including auxiliary information in order to establish more efficient ways of estimating the treatment effects on survival time in clinical trials have already been investigated by Lagakos,[46] Kosorok and Fleming,[45] Finkelstein and Schoenfeld[27] and Fleming *et al*.[29] Cook and Lawless[20] showed that under a parametric three-stage model for modelling an intermediate variable and the true endpoint of interest some gain in efficiency is possible if there is a close association between the two endpoints. Broglio and Berry[8] divided the time-to-death into two parts, progression-free survival and survival given progression and studied the advantage of considering the treatment effect on each of the endpoints separately. Faucett *et al*.[25] focused on an approach utilizing the progression times as auxiliary outcomes in order to compensate the observed censored cases in survival analysis by the use of a joint model and multiple imputation method for the survival times of the censored objects. Conlon *et al*[19] explored ways of utilizing times to progression as auxiliary information for the analysis of overall survival in order to gain efficiency. A parametric multi-state model with an incorporated cured fraction for progression was used to jointly model the time-to-progression and time-to-death. In addition, a multiple imputation method which involves imputing death times for censored subjects was applied due to the motivation to increase the potential to shorten the length of a trial and reduce sample sizes.

An efficient design in oncology trials is essential. The aim of the thesis is

to explore to what extent information on time-to-progression can be used to analyze overall survival. Therefore, statistical methods capturing the relationship of PFS and OS will be studied. In particular, the intention of the thesis is to find ways to improve the efficiency in estimation of the treatment effect on overall survival based on time-to-progression.

The remainder of the thesis is structured as follows. Chapter 2 contains a brief summary of concepts in survival analysis that are relevant for the subsequent chapters. Chapter 3 is about quantifying the association between progression-free survival and overall survival in oncology trials using Kendall's τ. This chapter considers methods for estimating the association between progression-free and overall survival in oncology trials. In Chapter 4, joint models of PFS and OS in a Bayesian framework will be investigated. Methods for joint modelling time-to-progression and death in oncology trials to impose testing for the group difference in terms of the average hazard ratio will be investigated. Chapter 5 explores various methods for gaining efficiency in testing for, and estimating the treatment effects on, overall survival in oncology trials compared to the standard methods based on Cox regression or the log-rank test. Chapter 6 concludes with a discussion and suggests potential work for the future.

CHAPTER 2

Preliminary Methods

The intention of this chapter is to provide an overview of the basic concepts within the time-to-event analysis which are relevant for the thesis. These concepts cover multi-state models, modelling bivariate event times and the evaluation of treatment effects.

2.1 Multi-state model framework

A common framework for modelling the event history of patients in a clinical trial is a multi-state model (Anderson *et al*,[3] Machado *et al*[53] and Putter *et al*[62]). The survival experience of a patient may contain information on several events observed during the trial. This individual process may include the movements between several disease states which could be described through a continuous time stochastic process. A common way to describe such a process is via a multi-state model in which the number of states are finite. The states may refer to stage of the illness and the movement between two states is termed a transition. Two cases of states can be distinguished. Transient state are states that can be revisited. States are terminal (or absorbing), if it can't be revisited. For example, death is terminal, as there no transitions can emerge from that state.

The simplest version of a multi state model is the case with only two states such as "dead" and "alive" and consequently one transition. In more complex multi-state models for survival analysis, the "alive" state may be divided into two or more intermediate transient states each of them referring to a particular stage of the illness. The most commonly used such model is the three-state illness-death model where there is a single intermediate state representing "illness". The survival process in patients with cancer can be expressed in terms of such a multi-state model where the states correspond to pre-progression (assumed to apply at the time of study entry and death), progression of the disease, and death. Time-to-progression (TTP) corresponds to the time of entry into the progression state, whereas PFS corresponds to the time of exit from the

pre-progressions state. OS corresponds to the time of entry into the death state. This particular example of the illness-death model is depicted in Figure 2.1.

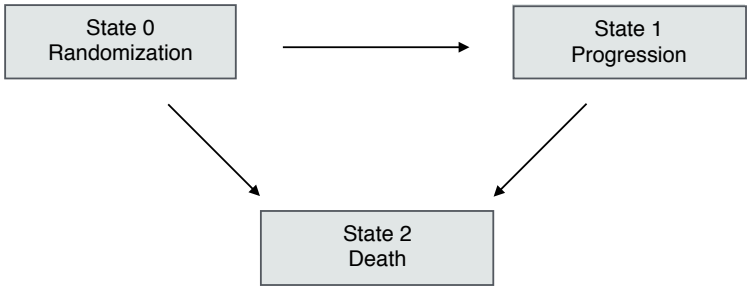

Figure 2.1: The three-state illness-death model for cancer survival

In consideration of the mathematical framework,[3,53] a multi-state model represents a stochastic process $(X(t), t \in T)$ with values in a finite space $S = \{1, \ldots, N\}$, where S represents the states. In other words, the multi-state-process $X(t) \in S$ describes the state occupied at time $t \in T$, where T is a finite time interval here. The history or filtration $\mathcal{F}_t$ refers to the observed process that captures information, e.g. on times of previous transitions and on states already occupied, up to time t. A multi-state process can be defined in terms of the transition probabilities between states:

$$P_{kl}(s, t; \mathcal{F}_s) = \mathbb{P}(X(t) = l | X(s) = k; \mathcal{F}_s),$$

where $s, t \in T$ with $s \leq t$.

Alternatively, it is common to characterize a multi-state model by the transition intensities, derivatives of the transition probabilities, from state $k \in S$ to state $l \in S$ as follows

$$\begin{aligned} \pi_{kl}(t; \mathcal{F}_t) &= \lim_{\Delta t \to 0} \frac{P_{kl}(t, t + \Delta t; \mathcal{F}_t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{\mathbb{P}(X(t + \Delta t) = l | X(t) = k; \mathcal{F}_t)}{\Delta t}. \end{aligned} \tag{2.1}$$

Transition intensities describe the instantaneous risk or hazard of a movement to state l at time t, conditionally on the present state k and the history $\mathcal{F}_t$.

When modelling a multi-state model for the survival course of patients, various assumptions can be made to specify the multi-state model regarding dependence on time. For example, the case to assume the transition intensities (2.1) to be constant over time defines a time-homogeneous multi-state model.[15] Furthermore, a Markov, semi-Markov or non-Markov multi-state models can be chosen. The Markov property, described in detail in Cox and Miller (p.76),[21] is a characteristic of a stochastic process, where the imminent future of a stochastic process does not depend on the past, only on the present. In the context of a Markov illness-death model, the hazard of death after progression only depends on the current state and not how long the patient has spent in the pre-progression state. A Markov multi-state model means mathematically in terms of the transition intensity as follows:

$$\pi_{kl}(t; \mathcal{F}_t) = \pi_{kl}(t), \tag{2.2}$$

where $k, l \in S$ and $\mathcal{F}_t$ is the history of the process up to time t.

Throughout the thesis, the most common assumption for the illness-death model is the semi-Markov assumption. The details of semi-Markov processes are explained in Cox and Miller (p.352).[21] In this context, a semi-Markov illness-death model implies a clock reset to zero, when the patient enters the state of progression yielding the following expression for the transition intensity:

$$\pi_{kl}(t; \mathcal{F}_t) = \pi_{kl}(t - t_k), \tag{2.3}$$

where t_k is defined as the time of entry into current state k. In the case of using a semi-Markov model, we assume a *homogeneous semi-Markov* model where the hazard of death after progression depends on time since progression rather than time since randomization. It may be seen as a special case of the *general semi-Markov* model[4] which allows the transition intensity to depend both on the duration in the progression state and the time since randomization. A semi-Markov assumption is more often assumed in the oncology context, but Markov

is more straightforward to deal with.

The transition probabilities can be derived from the transition intensities. For the illness-death model $P_{kl}(t_1, t_2)$ from time t_1 to t_2, where k, l describes the state with $k \in \{0, 1, 2\}$ and $l \in \{0, 1, 2\}$, the expression of the transition probabilities are given by[53]

$$P_{00}(t_1, t_2) = S_0(t_2 - t_1) = \exp(-\Pi_{01}(t_2 - t_1) - \Pi_{02}(t_2 - t_1)), \tag{2.4}$$

$$P_{11}(t_1, t_2) = S_1(t_2 - t_1) = \exp(-\Pi_{12}(t_2 - t_1)), \tag{2.5}$$

$$P_{12}(t_1, t_2) = S_1(t_2 - t_1) = \int_{t_1}^{t_2} P_{11}(t_1, u)\pi_{12}(u; \mathcal{F}_u)P_{22}(u, t_2)du, \tag{2.6}$$

where $\Pi_{kl}(t_1, t_2) = \int_{t_1}^{t_2} \pi_{kl}(t, \mathcal{F}_t)dt$ is the cumulative transition intensity between states k and l, where $k \leq l$. The transition probabilities $P_{11}(t_1, t_2)$ and $P_{12}(t_1, t_2)$ refer to a Markov illness-death model, if the transition intensity $\pi_{12}(t; \mathcal{F}_t)$ has the expression as in (2.2). In a semi-Markov illness-death model, $P_{11}(t_1, t_2)$ and $P_{12}(t_1, t_2)$ use the expression of the transition intensity $\pi_{12}(t; \mathcal{F}_t)$ defined as in (2.3).

2.1.1 Semi-competing risks framework

In this section, a variation of the competing risks framework will be briefly discussed. In clinical trials, distinct events of failures may occur when observing the patients. When the follow-up time stops after the occurrence of first event of failure apart from censoring, such data are called competing risks data. A situation in a clinical trial, where a terminal event such as death censors a non-terminal event such as progression of the disease may be often observed. Fine *et al*[26] defined that particular situation allowing multivariate events in a trial as semi-competing risk data. Such data are usually modelled by assuming a joining survival function of two event times over the positive quadrant whereas the observation is restricted to the upper wedge. Those approaches generally contain assumptions of latent failure times and imply therefore non-realistic assump-

tions on the marginal distribution of the non-terminal event. This modelling problem also occurs in the competing risks framework. In order to avoid the latent failure times in modelling semi-competing risks data, Xu and Kalbfleisch[74] proposed a restricted version of the illness-death model to model that data. In general, semi competing risks data can also be described within the multi-state model without referring to any latent failure times. According to Kalbfleisch and Xu, an illness-death model with shared frailty and some assumptions on the hazard functions imposes a model for modelling semi-competing risks data. Implicitly, modelling semi-competing risks data may be seen as a special case of multi-state model framework.

2.2 Copula models for bivariate survival data

Copulas are used for modelling dependence between variables in many applications. It is a common tool in the area of multivariate analysis, where the dependence of the variables is subject of research in order to reflect the correlation of those variables. The use of copulas is suitable for cases of two or more responses, where each follows a distribution on different parametric families. There are also situations where the responses don't have an obvious parametric distribution. Modelling their dependence based on copulas is still possible, as a non-parametric estimator such as Kaplan-Meier (in the case of right-censored data) or empirical CDF more generally, can be then applied to reflect their marginal behaviour. A comprehensive description of copula theory, the comparison of copula families and modelling different types of dependence with copulas are given in Joe.[41]

In detail, copula functions are continuous multivariate distributions where each of its variables follows a uniform marginal distribution over $(0,1)$. According to Sklar's theorem,[71] any arbitrary multivariate joint distribution can be expressed through its marginal distribution and a copula function separately. The

copula function includes all the information of the dependence between the endpoints independently from their marginal distribution.

In the thesis, we focus on specific application to bivariate survival data. Approaches for joint modelling of two events are often based on copulas, as they imply the dependence structure between those endpoints. In the subsequent chapters, we consider the bivariate copula function $C : [0,1]^2 \longrightarrow [0,1]$ of the failure endpoints (S,T), where $S :=$ PFS and $T :=$ OS, can be expressed by joint survivor function as follows

$$\mathbf{S}(s,t) = P(S \geq s, T \geq t) = C(\mathbf{S}_S(s), \mathbf{S}_T(t)), \quad s,t \geq 0, \tag{2.7}$$

where $\mathbf{S}_S$ and $\mathbf{S}_T$ are the marginal survivor functions of S and T, respectively. The exact specification of the copula function depends on the type of copula. In Chapters 3 and 4, the details of each copula used in the thesis will be described in the corresponding chapter.

2.3 Tests for and estimation of group difference

In many situations in clinical trials, time-to-event data is considered to be the endpoint of interest. In general, the analysis of time-to-event data, in order to reflect the treatment effect on the endpoint of interest, is then based on the assumption of proportional hazards. Proportional hazards correspond to a situation where the hazard ratio between the two treatments groups is constant over time. Given the assumption of proportional hazards, the standard tools to analyze the group difference are the log-rank test and the Cox model. While the log-rank test assesses whether there is a statistically significant group difference, the hazard ratio based on the Cox model determines the size of such a treatment effect. However, the log-rank test is most powerful under proportional hazards of overall survival between treatment groups. In the case of non-proportional hazards, i.e. when the hazard ratio between treatments is a function of time, a loss of power of the log-rank test is expected and the hazard ratio estimate from

the Cox model will not be strictly meaningful. However, the assumption of proportional hazards is restrictive and clinical time-to-event data seen in practice often violates a proportional hazards assumption. According to Royston and Parmar,[66] the observation of non-proportional hazards in clinical trials might arise due to more complex trials with longer follow-up periods for a better assessment of the group difference. Kancho *et al*[39] reviewed whether non-proportional hazards and non-constant hazard rates are taken into account in current practice when designing and analyzing randomized clinical trials. They concluded that proportional hazards and constant rates are mostly assumed in the design and analysis of a clinical trial. Furthermore, a preprint by Lin *et al*[49] studied alternative methods such as weighted log-rank tests, Kaplan-Meier curve-based tests and combination tests in order to analyze time-to-event endpoint under non-proportional hazards. Based on their simulation research, they showed that the performance of the considered alternative methods depends on the scenario.

Kalbfleisch and Prentice[43] proposed a generalization of the hazard ratio, appropriate even under non-proportional hazards. This approach is referred to as the average hazard ratio and is based on a flexible weighting function to incorporate the impact of time on the treatment effect. Rauch *et al*[63] investigated whether that average hazard ratio for the time-to event outcomes performs well, even under the assumption of proportional hazards. Overall, they concluded that the average hazard ratio seems to provide a meaningful interpretation of the treatment effect. Due to that benefit, they recommended the use of that alternative approach more in practice rather than the commonly applied hazard ratio in clinical trials, in particular when the proportional hazards assumption is not valid.

A natural approach to incorporating a treatment effect for an illness-death model is to assume treatment effects with respect to each individual transition. The transitions are modelled directly within that framework, hence proportional

hazards within each transition intensity can be assumed. However, since the overall survivor distribution is not directly modelled within that framework and is a complicated function of all the transition intensities, a non-proportional hazard is induced with respect to overall survival.

As OS times can be derived from the modelled times within an illness-death model, a general expression of the survival function for OS at any time t is

$$\begin{aligned} \mathbf{S}_{OS}(t) &= P_{00}(0,t) + P_{01}(0,t) \\ &= P_{00}(0,t) + \int_0^t P_{00}(0,u)\pi_{01}(u)P_{11}(u,t)du, \end{aligned} \tag{2.8}$$

where $\pi_{01}(u)$ is the transition intensity from state 0 to state 1. $P_{00}(0,t)$ corresponds to the survival probability to stay in state 0 within the time interval $[0,t]$ and $P_{11}(u,t)$ corresponds to the survival probability to stay in state 1 within the time interval $[u,t]$ (see Section 2.1, in particular formulas in (2.4) and (2.5), for more details).

According to the composition of the survival function as in (2.8), modelling data from the illness-death model framework induces a non-proportional hazards assumption for overall survival between the treatment groups.

Due to that situation, the average hazard ratio rather than the hazard ratio will be applied in the thesis in order have a more reliable and powerful comparison of the hazard rates between treatment groups.

As to the mathematical definition of AHR, let $f(t)$ be the pooled marginal density and $h(t)$ be the pooled marginal hazard. The survival function for both the treatment group and the control group as well as the hazard function for both the treatment group and the control group are donated by $S_{tr}(t)$, $S_c(t)$, $h_{tr}(t)$ and $h_c(t)$, respectively. According to the definition by Kalbfleisch and Prentice (2012),[43] the average hazard ratio is given by

$$AHR = \frac{\int_0^\infty \frac{h_{tr}(t)}{h_c(t)+h_{tr}(t)} w(t)f(t)dt}{\int_0^\infty \frac{h_c(t)}{h_c(t)+h_{tr}(t)} w(t)f(t)dt}, \tag{2.9}$$

where $w(t)$ is a weight function characterizing the influence of time on the hazard ratio. Schemper *et al*[68] showed that the average hazard ratio with the weight function $w(t) = \frac{S_c(t)f_{tr}(t)+S_{tr}(t)f_c(t)}{f_c(t)+f_{tr}(t)}$ simplifies to the concordance odds definition of the AHR as follows

$$AHR = \frac{\mathbb{P}[T_{tr} < T_c]}{1 - \mathbb{P}[T_{tr} < T_c]}, \tag{2.10}$$

where $\mathbb{P}[T_{tr} < T_c]$ is the concordance probability with randomly chosen survival times T_{tr} and T_c from the treatment group and the control group, respectively.

To summarize, Section 2.1 which is about the framework of multi-state models is of relevance for all subsequent chapters. In addition, Chapters 3 and 4 use copula models described briefly in Section 2.2. The AHR summarized in Section 2.3 will be an object of interest in Chapters 4 and 5, as the framework in those chapters will be based on illness-death models incorporating treatment groups.

CHAPTER 3

Quantifying the association between progression-free survival and overall survival in oncology trials using Kendall's τ

3.1 Introduction

Quantifying the association between progression-free survival (PFS) and overall survival (OS) in cancer trials can provide an indication of the extent to which PFS may be an effective surrogate for OS. Estimating the correlation between PFS and OS may be a potential step for validating PFS as a surrogate endpoint for OS, usually accompanied by meta-regression to also establish correlation between treatment effects across multiple studies.[12,60] Existing approaches to quantifying the correlation between PFS and OS include parametric and semi-parametric copula models[10,69] as well as non-parametric methods based on inverse probability of censoring weights.[47] Recently, there has also been a focus on illness-death model-based methods for quantifying the correlation between PFS and OS. Specifically, Fleischer *et al*[28] proposed a parametric method describing the correlation based on the strong assumption of constant hazards of progression and death before and after progression. Li and Zhang[48] extended this statistical model to allow increasing or decreasing hazards through Weibull hazard functions. However, the model is still relatively restrictive since a common shape parameter is used for the Weibull hazards for time-to-progression, time-to-death without progression and time from progression until death. In this chapter, we generalize these illness-death model-based methods to a much wider class of models and also use them to quantify the Kendall's τ correlation rather than the Pearson correlation. The performance of these methods is compared to existing methods both through simulation and by application to a colon cancer data set.

The remainder of this chapter is structured as follows. In Section 3.2, we describe the existing approaches to estimating the correlation. After illustrating the semi-parametric and parametric copula models, the non-parametric inverse of probability of censoring method is detailed. Subsequently, for the illness-death model-based approach we propose a generalized approach applicable to

any parametric model. Section 3.3 shows the results of simulation studies, where the performance of the methods, in terms of bias and efficiency are investigated and compared through several simulation scenarios. In Section 3.4, the methods are illustrated on the colon cancer data set. The chapter concludes with a discussion.

3.2 Different approaches to estimating the correlation between progression-free survival and overall survival

3.2.1 General setting

In this section, existing approaches to quantifying the association between PFS and OS are presented. Throughout it is assumed that the desired measurement of the association between PFS and OS is the Kendall's τ rank correlation. Let (S_1, T_1) and (S_2, T_2) be random variables representing the PFS and OS times for two independent patients 1 and 2. Kendall's τ is defined as

$$\tau = P\{(S_1 - S_2)(T_1 - T_2) > 0\} - P\{(S_1 - S_2)(T_1 - T_2) < 0\}, \tag{3.1}$$

where the first term refers to the probability of concordance and the second to the probability of discordance.

3.2.2 Copula models for bivariate survival data

An existing approach for the measurement of the dependence structure between failure times PFS and OS is through the use of copula models. Burzykowski *et al*[10] proposed a parametric copula method for quantifying the correlation between PFS and OS. A semi-parametric approach is also possible using the methods developed by Shih and Louis[69] for semi-parametric bivariate survival

copulas. Generally, the idea of these models is to estimate the marginal distributions for PFS and OS and to impose a particular dependence structure between these two endpoints.

The joint survival function of the failure endpoints (S, T), where $S :=$ PFS and $T :=$ OS, can be expressed by

$$\mathbf{S}(s,t) = P(S \geq s, T \geq t) = C(\mathbf{S}_S(s), \mathbf{S}_T(t)), \quad s,t \geq 0, \tag{3.2}$$

where $\mathbf{S}_S$ and $\mathbf{S}_T$ are the marginal survivor functions of S and T, respectively, and $C : [0,1]^2 \longrightarrow [0,1]$ is a bivariate copula function. There are a wide range of copula families allowing different patterns of dependency. In this chapter, we consider the Clayton, Hougaard and Frank copula functions, all of which belong to the class of Archimedean copulas. This class is often used as it allows the dependence between the two variables to be defined by a single parameter, δ. In general, a bivariate distribution in terms of the Archimedean copula family is given by

$$C(u,v) = \phi_\delta[\phi_\delta^{-1}(u) + \phi_\delta^{-1}(v)], \quad 0 \leq u,v \leq 1, \tag{3.3}$$

where ϕ_δ is some generator function satisfying $0 \leq \phi_\delta \leq 1, \phi_\delta(0) = 1, \phi'_\delta < 0$ and $\phi''_\delta > 0$. If ϕ_δ is a Laplace transform of some distribution then the corresponding Archimedean copula is equivalent to a proportional frailty model.[58] Three special cases of the proportional frailty model are of interest here. First, Clayton's model[16] can be represented as

$$C_\delta(u,v) = (u^{(1-\delta)} + v^{(1-\delta)} - 1)^{1/(1-\delta)}, \qquad \delta > 1, \tag{3.4}$$

where $\phi_\delta(x) = (1+x)^{1/(1-\delta)}$ is the Laplace transform of a gamma distribution with rate parameter 1 and shape parameter $1/(\delta - 1)$. S and T are positively associated when $\delta > 1$ and become independent as $\delta \to 1$. The second example is Hougaard's model,[38] where the function is given by

$$C_\delta(u,v) = \exp(-[-\log(u)^{(1/\delta)} - \log(v)^{(1/\delta)}])^\delta, \qquad 0 < \delta < 1, \tag{3.5}$$

where $\phi_\delta(x) = \exp(-x^\delta)$ is the Laplace transform of the positive stable distribution with density[38]

$$-\frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{\Gamma(k\delta+1)}{k!}(-x^{-\delta})^k\sin(\delta k\pi),\ \ x>0. \tag{3.6}$$

S and T are positively associated when δ is small and become independent when $\delta \to 1$.

The copula function of Frank's model[30] can be expressed as

$$C_\delta(u,v) = \log_\delta\left[1+\frac{(\delta^u-1)(\delta^v-1)}{\delta-1}\right], \qquad \delta>0, \tag{3.7}$$

where $\phi_\delta(x) = \log_\delta(1-(1-\delta)\exp(-x))$ is the Laplace transform of a logarithmic series distribution for $0<\delta<1$ and $\log_\delta$ represents the logarithm to the base δ. S and T are positively associated for the case $\delta<1$, negatively associated for the case $\delta>1$, and become independent when $\delta \to 1$.

Burzykowski *et al*[10] presented a fully parametric copula model. The marginal distributions are assumed to have a particular parametric form, for instance they may be assumed to each have separate Weibull distributions. Combining the marginal distributions for survival with a copula function, a corresponding bivariate joint survival copula function based on Weibull distributions can be constructed. As shown in the work of Burzykowski *et al*,[10] a likelihood function can be derived by taking all potential censoring cases in the datasets into account. The scale and shape parameters in the two hazard functions as well as the copula dependence parameter δ can be jointly estimated by using maximum likelihood estimation.

Shih and Louis[69] proposed a semi-parametric model in which a parametric copula is assumed for the dependence, but the marginal distributions are left unspecified. A two-stage approach is taken for estimation. The idea is to estimate the marginal survivor functions (u,v) by the non-parametric Kaplan-Meier estimator in the first stage. After deriving the likelihood function incorporating the different cases of censoring,[69] maximum likelihood can be used to estimate

the unknown copula association parameter δ conditional on the values of the survival functions of PFS and OS.

A particularly useful aspect of Archimedean copula methods is that Kendall's τ can be expressed directly as a function of $\phi_\delta^{-1}(x)$ as follows (see the work of Genest and MacKey[36]):

$$\tau = 1 + 4\int_0^1 \frac{\phi_\delta^{-1}(v)}{\partial\phi_\delta^{-1}(v)/\partial v}dv. \tag{3.8}$$

As a direct relationship exists between the Kendall rank correlation τ and δ, an estimate of τ can be derived from the maximum likelihood estimator of δ. Application of (3.8) for the respective generator functions leads to Kendall's τ for Clayton's copula of $\tau = \frac{\delta-1}{\delta+1}$, for Hougaard's copula of $\tau = 1 - \delta$ and for Frank's copula of $\tau = 1 - 4\frac{D_1(-\log(\delta))-1}{\log(\delta)}$, where D_1 represents the Debye function of order 1.[50]

While copula methods are very convenient for modelling bivariate survival data, their specific use for modelling PFS and OS is potentially questionable. By definition the PFS time must be less than or equal to the OS time. However, the copula model does not restrict the ordering of PFS and OS. Moreover, if PFS and OS are assumed to have continuous marginal distributions, the copula model assumes the values of PFS and OS coincide with probability 0, whereas in fact they will coincide whenever a patient dies before progression. As a consequence the copula model is guaranteed to be somewhat misspecified, even if the marginal distributions are correctly specified. There is therefore potential that estimates of dependency from the copula model will be biased. Dejardin *et al*[23] performed a limited simulation study to investigate the possible bias of applying a bivariate shared Gamma frailty model (equivalent to the Clayton copula model) to data on PFS and OS in which a three-state unidirectional model was assumed for the generation process. They identified a small bias in the estimate of Kendall's τ, with the magnitude of bias being greater for the scenario with lower τ.

3.2.3 Non-parametric methods based on Inverse Probability of Censoring Weight (IPCW)

Non-parametric estimation of Kendall's τ for censored data is possible through the use of inverse probability of censoring weighting (IPCW).[47] The methods are applicable to general bivariate survival data, including PFS and OS times as a special case where some of the calculations are simplified if it is assumed that a common censoring time will apply to both the PFS and OS times for a given patient. Let (S_i, T_i) for every $i = (1, \ldots, n)$ be independent replications of the failure endpoint times (S, T). A pair of two replications can be seen as the survival experience from two individuals. The concordance or discordance status of the pairs is required for the empirical calculation of Kendall's τ. The concordance status for subjects i and j given by

$$C_{ij} = \begin{cases} 1 & \text{if} \quad (S_i - S_j)(T_i - T_j) > 0 \\ -1 & \text{if} \quad (S_i - S_j)(T_i - T_j) < 0. \end{cases}$$

In the case of no censoring, the concordance status can be determined for all $\binom{n}{2}$ possible pairs. Hence, Kendall's τ can be estimated by its sample version

$$\tau = \binom{n}{2}^{-1} \sum_{i<j} C_{ij}, \tag{3.9}$$

where summing over $i < j$ avoids taking a pair into account twice.

In the presence of censored data, the concordance or discordance status can only be determined for orderable pairs. Let R_{ij} be an indicator of whether the pair (i, j) is orderable. Defining the respective censoring times C_i and C_j for subjects i and j, then R_{ij} is given by[57]

$$R_{ij} = I\left(\tilde{S}_{ij} < \tilde{C}_{ij}, \tilde{T}_{ij} < \tilde{C}_{ij}\right) = \begin{cases} 1 & \text{if pair orderable} \\ 0 & \text{otherwise,} \end{cases}$$

where $\tilde{S}_{ij} = \min(S_i, S_j)$, $\tilde{T}_{ij} = \min(T_i, T_j)$ and $\tilde{C}_{ij} = \min(C_i, C_j)$.

Oakes[57] extended the estimator in (3.9) to an estimator for τ by taking the sum over the orderable pairs only. The IPCW technique can also be applied to the Oakes estimator in order to correct the bias caused by the presence of missing data. The contribution of each orderable pair to the Kendall's τ is weighted by the inverse probability of being orderable. Then the estimator can be represented as follows.[47]

$$\tau = \frac{\sum_{i<j} R_{ij} C_{ij} W_{ij}}{\sum_{i<j} R_{ij} W_{ij}} \in [-1, 1],$$

where $W_{ij} = \frac{1}{\hat{p}_{ij}}$ are the weights defined by the inverse estimated selection probabilities for orderable pairs p_{ij} as follows:

$$\begin{aligned} p_{ij} &= \mathbb{P}(R_{ij} = 1 | \tilde{S}_{ij}, \tilde{T}_{ij}) \\ &= \mathbb{P}(\tilde{S}_{ij} < \tilde{C}_{ij}, \tilde{T}_{ij} < \tilde{C}_{ij} | \tilde{S}_{ij}, \tilde{T}_{ij}) \\ &= G\{\max(\tilde{S}_{ij}, \tilde{T}_{ij})\}^2, \end{aligned}$$

where $G(\cdot)$ is the survival function of censoring which can be estimated, under an assumption of random censoring, via a Kaplan-Meier estimate obtained by reversing the censoring indicator. While in the above we assume a common censoring time for PFS and OS, in many oncology trials the progression time is effectively censored at the last screening time. The methods in Lakhal *et al*[47] allow for different but dependent censoring times for the two event times. However, note that censoring for time-to-progression is not the same as PFS since patients who die after their last screening time will not be treated as censored. Hence the case of differential censoring times is not easily accommodated.

The IPCW approach to estimating the association between PFS and OS is potentially quite attractive as it requires no assumptions to be made about either the dependence structure or the marginal distributions of the times to PFS and OS. However, for consistency the IPCW method requires that $\mathbb{P}(R_{ij} =$

$1|\tilde{S}_{ij}, \tilde{T}_{ij}) > 0$ for all potential $\tilde{S}_{ij}, \tilde{T}_{ij}$. Effectively this means that there must be some chance of a pair of observations being orderable regardless of the times until progression or death. For this to be the case the support of the distributions of PFS and OS must be contained within the support of the censoring distribution. Such an assumption is very unlikely to be plausible in most oncology trials, where at the time of analysis the maximum follow-up time will typically be shorter than the longest possible survival time. The bias in the estimate of τ for data with limited follow-up will depend on how much of the upper tails of the PFS and OS distributions are not observable and how representative the dependence in the body of the distribution is to that of the tails. It is of interest to investigate the extent of bias in estimation of τ when follow-up is limited and how the bias depends on the type of the censored data. This behaviour could be investigated through simulation by generating data in a greater range of censoring scenarios to understand how shorter follow-up times affect the estimate of τ.

3.2.4 Model-based methods

Fleischer *et al*[28] presented a parametric multi-state model describing the Pearson correlation between survival outcomes PFS and OS under the assumption that the transition intensities between the states in the underlying multi-state model are constant. More recently, Li and Zhang[48] extended the method by using Weibull hazard functions to describe the transition intensities therefore allowing them to either be monotonically increasing or decreasing with time. Their model corresponds to a *homogeneous semi-Markov* model where the hazard of death after progression depends on time since progression rather than time since randomization (see Section 2.1 in the Preliminary Methods for more details). The model has four parameters, $\lambda_{01}, \lambda_{02}, \lambda_{12}$ and α, with the model in the work of Fleischer *et al*[28] arising as a special case where $\alpha = 1$. Li and Zhang

expressed the model in terms of the distributions of latent event times: time-to-progression, survival before progression and time from progression to death. However, expressing the model in terms of the transition intensities is more desirable since it avoids making untestable assumptions about the independence between time-to-progression and survival before progression.[2]

The three-state model is already shown in Figure 2.1 in the Preliminary Methods, but depicted in Figure 3.1 below for an improved illustration. Considering the expression of the transition intensities in the three-state model of Li and Zhang, we use a slightly different parametrization for the scale parameter throughout the thesis for reasons. The hazard functions of the Weibull distributed transitions with scale parameters λ_{01}, λ_{02}, λ_{12} and one shape parameter α are given by

$$\pi_{01}(t) = \alpha\Big(\frac{1}{\lambda_{01}}\Big)^{\alpha} t^{\alpha-1},$$
$$\pi_{02}(t) = \alpha\Big(\frac{1}{\lambda_{02}}\Big)^{\alpha} t^{\alpha-1},$$
$$\pi_{12}(s) = \alpha\Big(\frac{1}{\lambda_{12}}\Big)^{\alpha} s^{\alpha-1},$$

where t and s refer to time since randomization and to time since progression, respectively.

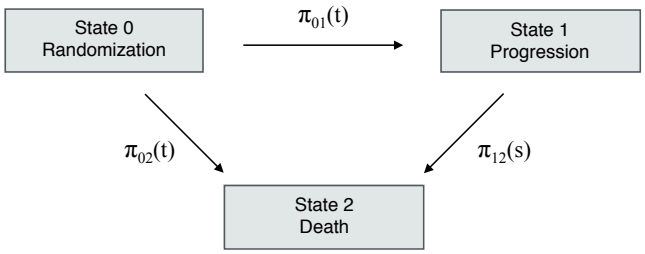

Figure 3.1: The three-state illness-death model for cancer survival

Estimation of the correlation between PFS and OS involves first estimating the parameters of the multi-state model via maximum likelihood. Li and Zhang derived a closed-form expression for the Pearson correlation between PFS and

OS for given parameters, into which the maximum likelihood estimates can be substituted.

In order to ensure the existence of a closed-form expression for the Pearson correlation, Li and Zhang assumed the same shape parameter α for the three Weibull functions. However, the necessity for analytical tractability leads to a somewhat restrictive model. For instance, in real data examples the hazard of progression may increase with time whereas the hazard of death before progression may be close to constant. Furthermore, PFS and OS as time-to-event outcomes will typically be highly positively skewed and as a consequence a non-linear dependence between PFS and OS would be expected, meaning the Pearson coefficient is unlikely to be an appropriate measure of association.

Due to these aspects, we extend the method to allow estimation of the Kendall rank correlation coefficient for general parametric illness-death models.

3.2.5 Generalized model-based methods

In this section, we generalize the illness-death model-based approach of Li and Zhang[48] to achieve more flexibility and to allow estimation of Kendall's τ rather than the Pearson correlation coefficient.

The modified approach continues to use a multi-state illness-death model, but allows any parametric formulation for the transition intensities between states. In particular, we can allow post-progression survival to depend on both time-to-progression, denoted by t_0, and time since progression, denoted by s. We assume $\pi_{01}(t), \pi_{02}(t)$ and $\pi_{12}(s; t_0)$ are parametrized by a vector of parameters $\boldsymbol{\theta}$ which can be consistently estimated from data with a finite follow-up period through maximum likelihood estimation.

From the definition of Kendall's τ in (3.1) and under an assumption that the bivariate lifetime random variables $(S_n, T_n)_{n\in\mathbb{N}}$ representing the PFS and OS times of the n patients are independent and identically distributed, for the gen-

eral illness-death model the Kendall's τ implied by the model is as follows:

$$\begin{aligned}\tau_{\text{mod}} = \quad & 4\int_0^\infty \pi_{02}(s)\exp\{-2\Pi_0(s)\}ds \\ & +4\int_0^\infty\int_0^{s_1}\int_0^\infty \pi_{01}(s_1)\pi_{01}(s_2)\exp\{-\Pi_0(s_1)-\Pi_0(s_2)\}\pi_{12}(s_3;s_1) \\ & \quad \times\exp\{-\Pi_{12}(s_3;s_1)\}[1-\exp\{-\Pi_{12}(s_1+s_3-s_2;s_2)\}]ds_3ds_2ds_1 \\ & +4\int_0^\infty\int_0^{s_1} \pi_{02}(s_1)\pi_{01}(s_2)\exp\{-\Pi_0(s_1)-\Pi_0(s_2)\} \\ & \quad \times(1-\exp\{-\Pi_{12}(s_1-s_2;s_2)\})ds_2ds_1-1,\end{aligned} \tag{3.10}$$

where $\Pi_0(t)=\int_0^t \pi_{01}(u)+\pi_{02}(u)du$ and $\Pi_{12}(s;v)=\int_0^s \pi_{12}(u;v)du$. The first term in (3.10) refers to the case where one patient dies before progression, before the other has died or progressed. The second term refers to the case where patients 1 and 2 progress at times s_1 and s_2, respectively, where $s_1 > s_2$ and subsequently patient 1 survives an additional s_3 whereas patient 2 dies within $s_1 - s_3 + s_2$ of progression. The third term refers to the case where patient 1 progresses and dies before patient 2, despite patient 2 dying without progression. A full derivation of (3.10) is given in Appendix A.1.

In the model of Fleischer *et al*,[28] where exponential distributed transitions are considered, the expression in 3.10 can be simplified. By substituting $\pi_{01}(t) = \frac{1}{\lambda_{01}}, \pi_{02}(t) = \frac{1}{\lambda_{02}}$ and $\pi_{12}(t;t_0) = \frac{1}{\lambda_{12}}$ into (3.10) and directly integrating, after some algebraic manipulation we obtain

$$\tau_{\text{mod}} = \frac{(\lambda_{02}\lambda_{12})^2+2(\lambda_{01}\lambda_{03}\lambda_{12}^2+\lambda_{01}\lambda_{02}^2\lambda_{12}+\lambda_{01}\lambda_{12}^2+\lambda_{01}^2\lambda_{02}\lambda_{12})}{(\lambda_{02}\lambda_{12}+\lambda_{01}\lambda_{12})(\lambda_{02}\lambda_{12}+\lambda_{01}\lambda_{12}+\lambda_{01}\lambda_{02})}-1.$$

Rather than the exponential rate parameter as in Fleischer *et al* we use the scale parameter, inverse of the rate parameter, denoted by λ_{01}, λ_{02} and λ_{02} in order to be consistent with the definition for the scale parameter for the Weibull hazards in the thesis.

However, the integrals in (3.10) are analytically intractable for the model of Li and Zhang. Nevertheless, the lack of a closed form expression is not a major

hindrance since τ_{mod} can be obtained quite easily and with arbitrary accuracy via numerical or Monte-Carlo methods. Moreover, making the underlying model more complex, for instance by allowing separate Weibull shape parameters for each transition intensity, has little or no bearing on the computational difficulty of calculating τ_{mod}.

Monte-Carlo methods provide a particularly convenient way of evaluating the model-based Kendall's τ. We can use the fact that for a model where S and T are continuous,

$$\begin{aligned}\tau &= 2P(S_1 > S_2, T_1 > T_2) - 2P(S_1 < S_2, T_1 > T_2) \\ &= 2P(S_1 > S_2, T_1 > T_2) - \{1 - 2P(S_1 > S_2, T_1 > T_2)\} \\ &= 4P(S_1 > S_2, T_1 > T_2) - 1. \end{aligned} \tag{3.11}$$

It is therefore only necessary to evaluate $P(S_1 > S_2, T_1 > T_2)$ which can be achieved by simulating $2M$ pairs of (S_i, T_i) and then taking

$$\hat{P}(S_1 > S_2, T_1 > T_2) = M^{-1} \sum_{i=1}^{M} I(S_i > S_{i+M}, T_i > T_{i+M}). \tag{3.12}$$

Simulation for general illness-death models can be achieved using the methods in Beyersmann *et al*.[7] The Monte-Carlo standard error associated with the approximation is at most $1/2\sqrt{M}$. Typically, $M = 1 \times 10^6$ or 1×10^7 samples can be generated using very little computation time, meaning the Monte-Carlo standard error is negligible. A point estimate for τ_{mod} can be obtained by simulating $2M$ independent pairs of PFS and OS times from the illness-death model with parameter estimates $\hat{\boldsymbol{\theta}} := (\hat{\lambda}_{01}, \hat{\lambda}_{02}, \hat{\lambda}_{12}, \hat{\alpha}_{01}, \hat{\alpha}_{02}, \hat{\alpha}_{12})$. The parameters of the parametric illness-death model are estimated as in Li and Zhang[48] via maximum likelihood (see Appendix A.4 for more details). The only difference is that we used distinct shape parameters α_{01}, α_{02} and α_{12} corresponding to each transition intensity instead of a common shape parameter α for all transitions.

A simulation-based approach may also be used to obtain confidence intervals for τ_{mod} using a variant of the *simulation delta method*.[51] This involves firstly

generating B samples

$$\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_B^* \sim N(\hat{\boldsymbol{\theta}}, I(\hat{\boldsymbol{\theta}})^{-1}),$$

where $I(\hat{\boldsymbol{\theta}})$ is the observed Fisher information of the log-likelihood. For each of the B samples, a pair of (S, T) from the illness-death process with parameters $\boldsymbol{\theta}_b^*$ are simulated $2M$ times. The next step is to estimate τ_{mod} denoted by τ_{mod}^{*b} for every $b \in [1, B]$ using (3.12). Confidence intervals can then be constructed based either upon the sample standard deviation or sample quantiles of $\tau_{\text{mod}}^{*1}, \ldots, \tau_{\text{mod}}^{*B}$. A non-parametric bootstrap variant of this algorithm is also possible where B bootstrap samples are generated by repeatedly resampling from the original data and the maximum likelihood estimates are recomputed to generate each $\boldsymbol{\theta}_b^*$.

3.3 Simulations

3.3.1 Simulation set up

In this section, the performance of the methods is studied through simulation. It is assumed that the true underlying model is an illness-death model as this plausibly reflects the underlying disease process. For the first simulation scenario A, we assume a homogeneous semi-Markov model with Weibull transition intensities and take the values of the shape and scale parameters for each intensity to be those that best fitted to an external dataset from a trial of treatments for colon cancer[55] which will be reanalyzed in Section 3.4. As it would usually be expected, there is a lower hazard of death before progression than after progression. However, in the second simulation scenario B, we design this mechanism to be the other way around and therefore refer to this as the 'unrealistic scenario'. In both scenarios, we assume a multi-state model with a homogeneous semi-Markov assumption, where the imminent future is only dependent on the time spent in the present state and not on other previous history.

For the final simulation scenario C, we seek to investigate sensitivity of the illness-death model-based method to misspecification of a homogeneous semi-Markov assumption, by generating data in which time-to-progression also affects the hazard of death given progression. Specifically we assume a general semi-Markov process where the Weibull hazard function of death after progression depends on the time of progression t_0 as well as time since progression. The sojourn time in the post-progression state depends on whether progression or not occurred before a fixed time point, e.g. 2 months. Therefore, we use different shape and scale parameters for the Weibull hazard function of death after progression. If progression occurs before 2 months, it is expected that the hazard of death given progression is higher compared to the case where progression is experienced after 2 months. Table 3.1 shows the setting values of the scale parameters $\lambda_{01}, \lambda_{02}, \lambda_{12}$ and shape parameters $\alpha_{01}, \alpha_{02}, \alpha_{12}$ for scenario A, B and C, respectively. Based on those parameter values, the Kendall's τ can be obtained for each scenario via simulation as mentioned in Section 3.2.5.

Table 3.1: Parameter values in the three simulation scenarios

	Scenario		
Parameter	A	B	C
λ_{01}	9.698	9.6978	9.6976
λ_{02}	61.296	20	61.296
λ_{12}	1.654	50	$\begin{cases} 1.654 & \text{if } t_0 < 2 \\ 2.5 & \text{if } t_0 \geq 2 \end{cases}$
α_{01}	0.675	0.675	0.675
α_{02}	1.088	1.008	1.088
α_{12}	1.009	1.080	$\begin{cases} 1.008 & \text{if } t_0 < 2 \\ 1.005 & \text{if } t_0 \geq 2 \end{cases}$
τ	0.835	0.120	0.815

We consider four censoring cases for each simulation scenario A, B and C.

In censoring cases 1 and 2, the censoring follows an exponential distribution. However, case 1 has 20% of patients whose OS time is censored while case 2 has 45 % of OS times censored. In censoring cases 3 and 4, the same levels of censoring are used, but the censoring times arise from a uniform rather than an exponential distribution. In more detail, the upper limit of the uniform distribution of censoring implies the maximum follow-up time. Regarding the normal scenario with high censoring, the maximum follow-up time is 10.5 years, where 39% and 45% of PFS and OS, respectively, were unobservable. The non-Markov case is similar, as the maximum follow-up time is 12 years, where 36% of PFS and 45% of OS are beyond the range of the follow-up. In comparison to these two scenarios, the unrealistic case maximum follow-up time is 59 years, where 7% of PFS and 45% of OS are censored.

3.3.2 Simulation results

Box plots of the estimates of the Kendall's τ from each method using 1000 simulated datasets under uniform censoring and exponential censoring are shown in Figure 3.2 below and in Figure A.2 in Appendix A.2, respectively. Further details such as the estimates of the Kendall's τ, the respective bias, standard deviation and mean squared error for every model both under uniform censoring and exponential censoring can be found in Appendix A.2.

As the patterns of the outcomes are very similar among the two censoring types, we restrict on the analysis of the simulation results under uniform censoring below.

The upper row of plots in Figure 3.2 corresponds to the realistic scenario (scenario A) and indicates that the model-based method and the IPCW method perform very well. In particular, the IPCW continues to perform quite well even when there is a higher rate of censoring. The results from the copula models are quite varied. While the Clayton copula based estimator is almost unbiased, the

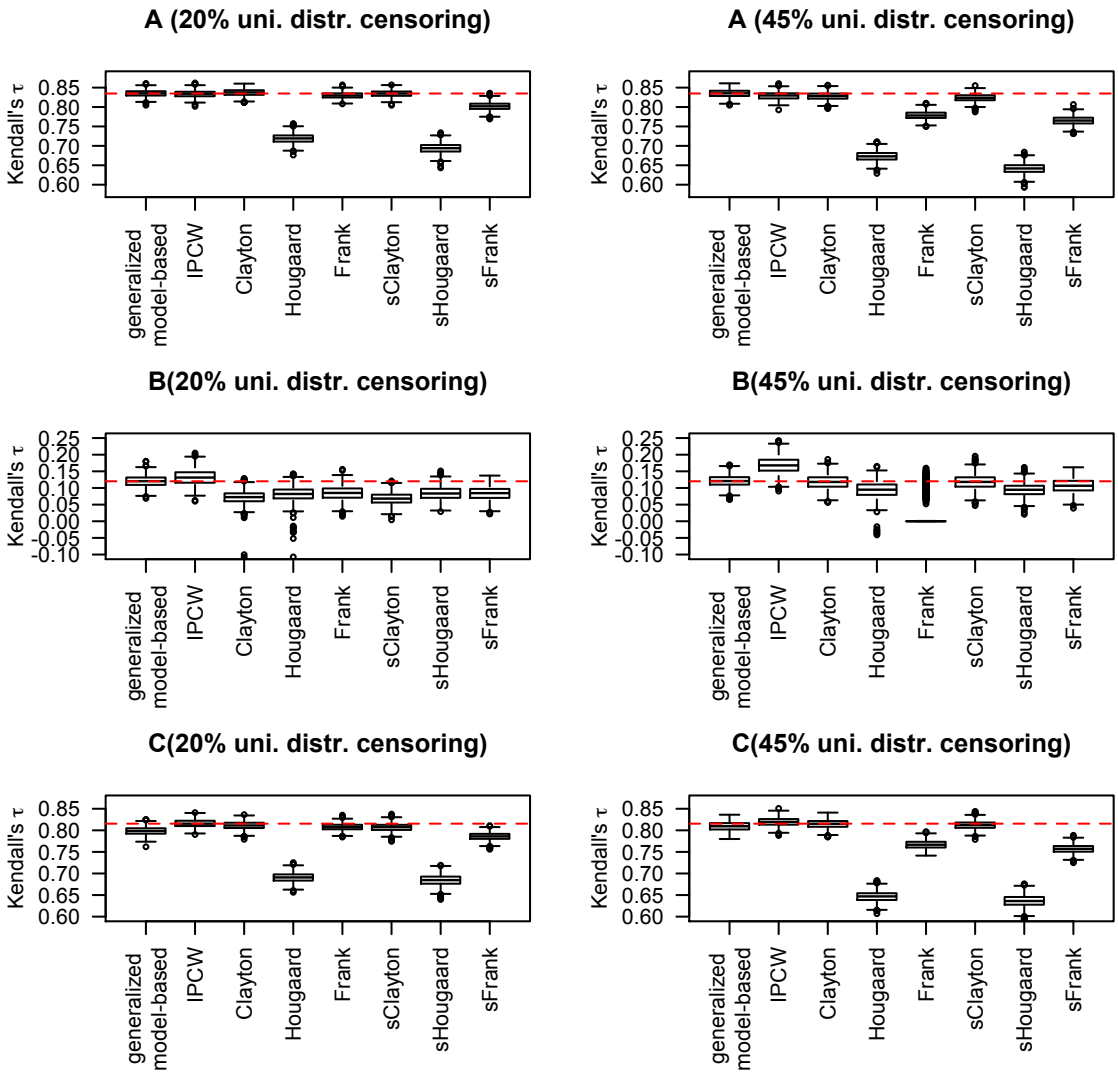

Figure 3.2: Box plots of estimates of Kendall's τ from 8 methods. Dashed red line indicates the true value.
A: normal scenario, where the parameters are used from an external dataset from a trial of treatments for colon cancer
B: unrealistic scenario
C: general semi-Markov scenario
sClayton: two stage semi parametric Clayton model; sHougaard: two stage semi parametric Hougaard model; sFrank: two stage semi parametric Frank model

Hougaard copula model captures the dependence rather poorly. As expected, the efficiency of the semi-parametric copula models is lower compared to the fully parametric models. Both the Hougaard and Frank copula estimates are

sensitive to the rate of censoring, with a greater degree of bias for higher rates of censoring. Overall, the model-based method has the lowest bias and lowest mean squared error, however as shown in Table A.1 in Appendix A.2, the Clayton copula model also has a low mean squared error being very close to the mean squared error of the generalized model-based method. Hence, the Clayton copula seems to be a very good model in this case. Further, it is not surprising that the generalized model-based method performs very well, as it corresponds to the mechanism of the data generation.

It is noticeable that the estimate of Kendall's τ is sensitive to the choice of copula function. The Clayton model performs very well, in contrast to Hougaard's model that seems to be misspecified leading to large negative biases.

Contour plots of the density functions of each of the copula models and also the implied copula for the illness-death model are presented in Figure 3.3, and offer an explanation of these differences. The Clayton, Hougaard and Frank copulas assume lower tail, upper tail and symmetric dependence, respectively. An approximation to the density of the implied copula for the true generating model is obtained through bivariate kernel density estimation from simulated PFS and OS times that are transformed by their respective marginal distribution functions. The implied copula for the true generating model indicates lower tail dependence, which is qualitatively similar to that of the Clayton model.

The middle row of Figure 3.2 shows the box plots of the estimates of Kendall's τ for each method in the unrealistic scenario (Scenario B). The model-based method continues to perform well. However, in this case the IPCW estimator is biased with a larger bias for the higher censoring rate. In the low censoring case, it is noticeable that the performance of the copula models is different in comparison to the first simulation scenario. The bias and the standard deviation of the Clayton models have increased, but the Hougaard model seems less misspecified than in the first simulation scenario.

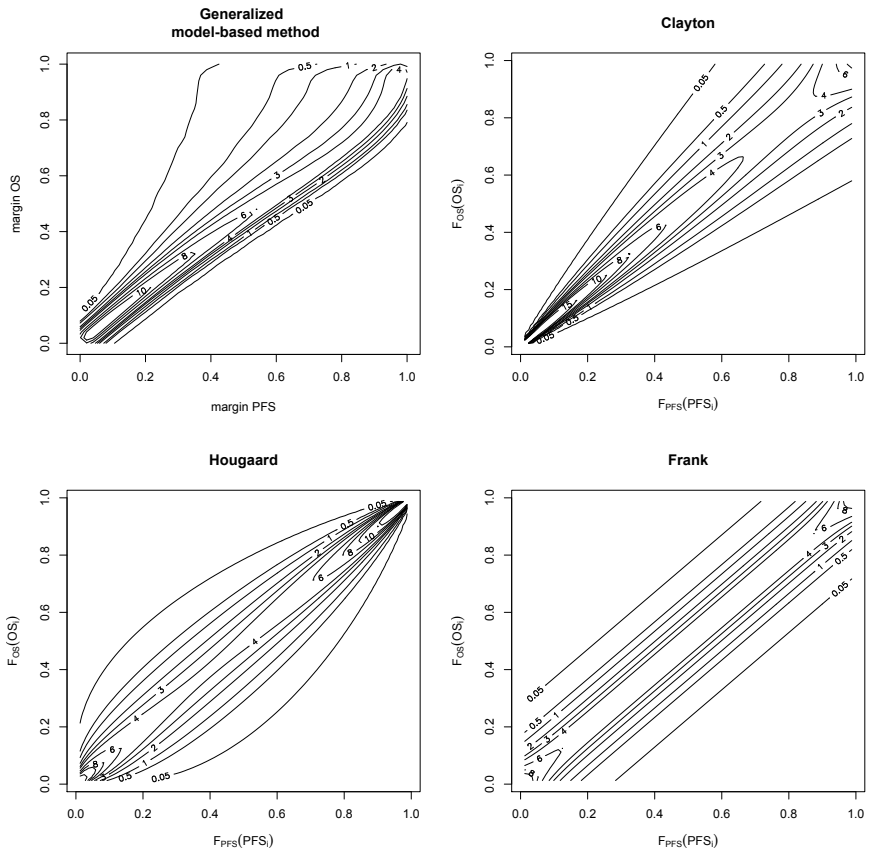

Figure 3.3: Realistic simulation case: Contour plots for the bivariate density function based on the model-based method and the survivor joint Clayton's copula model, Hougaard's copula model and Frank's copula model. Kendall's τ is 0.836.

Figure 3.4 shows the bivariate density plots for the unrealistic scenario and indicates why the performance of the copula models is sensitive to the type of data. The top-left graphic represents the density of the implied copula for the true model. The mode in the upper-left part of the plot corresponds to patients with a quick time-to-progression, but then a long follow-up to death. The band of higher density in the lower part of the plot corresponds to cases where PFS and OS are equal. The bivariate copula models have problems incorporating the combination of negative and positive correlation between PFS and OS and consequently are a poor fit to this type of data.

In addition, it is somewhat surprising that some of the copula models per-

form better in the presence of high censoring in comparison to the low censoring case. However, this can be explained by higher censoring leading to less contribution from the mode corresponding to short PFS and long OS, which the copula models are unable to accommodate.

As the copula models are not able to fit this type of data well, the model-based method is preferred in this special case. However, as with Scenario A, the superior performance of the model-based method is somewhat to be expected since it was the model by which the data were generated.

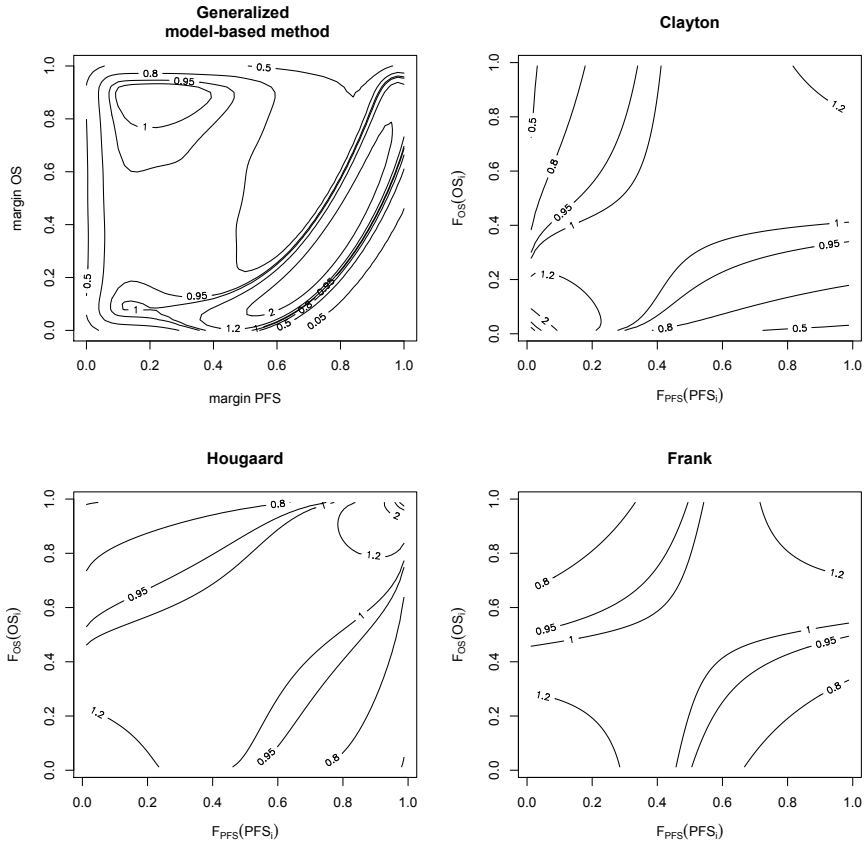

Figure 3.4: Unrealistic simulation case: Contour plots for the bivariate density function based on the model-based method and the survivor joint Clayton's Copula model, Hougaard's Copula model and Frank's Copula model. Kendall's τ is 0.119.

The results of the final simulation scenario C, based on a general non-homogeneous semi-Markov illness-death model, are shown in the lower row of

Figure 3.2. This scenario was used to investigate how incorrectly assuming a homogeneous semi-Markov model affects model-performance of the model-based method. Indeed, there is some negative bias in this case, while the performance of other methods stays broadly similar to scenario A. Due to the similar setting in scenario A and C as seen in Table 3.1, the bivariate density functions in C are expected to show the same structure as in A. The contour plots for the general semi-Markov scenario can be found in A.2 in Appendix A.2.

One way of choosing the most appropriate copula model in practice is to choose the one with the lowest AIC. Note that we can only compare AIC between the parametric copula models or between the semi-parametric copula models, but not between a parametric and semi-parametric model. Tables A.4 and A.5 in Appendix A.3 give the percentage of simulation replications for which each copula model had the lowest AIC. Broadly similar results were obtained for the parametric and semi-parametric models, except that the proportion of times the Frank copula was chosen as opposed to Clayton was higher for the semi-parametric models. In all cases the Clayton model was preferred the majority of the time, with the Hougaard model never chosen. While for scenarios A and C the Clayton model is the least biased, for scenario B in the 20% uniformly distributed censoring case the Clayton model is slightly more biased than Hougaard and Frank. The results therefore seem to indicate that the best model with respect to AIC will not necessarily correspond to the model which best estimates Kendall's τ.

3.4 Application

In order to illustrate the performance of the different methods, they were applied to data from a clinical trial of treatments for colon cancer.[55] This trial was conducted to investigate the effectiveness of two adjuvant therapies in improving surgical cure rates in stage III colon cancer. Patients were randomized

to observation, to the treatment levamisole alone or to a combination of levamisole plus fluorouracil. In terms of the measurements in this trial, the time-to-progression and time-to-death were observed in order to evaluate the difference in the hazards of recurrence and death between the treatment groups. The dataset contains the survival experiences of 929 patients who were followed up for 5 years or more (median follow-up, 6.5 years). During the trial, 425 individuals died, 54 were censored after progression and 423 were censored before progression. The maximum follow-up time was 9.1 years, by which point based on the Kaplan-Meier estimates, 43% and 46% would be yet to experience the PFS time and OS time, respectively. Hence, in this case, considerable extrapolation beyond the follow-up period is required to fully characterize the distributions.

In the analysis we initially pool together data from all treatment arms when estimating Kendall's τ and we first considered models using either Weibull transition intensities or Weibull marginal distributions for the illness-death model-based method and the parametric copula models. There was clear evidence against the constrained Weibull model used in the method of Li and Zhang[48] based on a likelihood ratio comparing the models with common and different Weibull shape parameters for the three intensities (LR = 41.6 on 2 degrees of freedom, $p < 0.001$). However, none of the Weibull-based models represented an adequate fit to the data. Comparisons of the estimated cumulative marginal hazard of PFS and OS with the Nelson-Aalen estimates are given in Figure A.3 in Appendix A.3 and indicate substantial discrepancies in all cases. To improve the model fit we considered Royston-Parmar (RP) flexible parametric models.[64] Specifically, we consider models for which the cumulative log-hazard function $\log H(t)$ is modelled as a natural cubic spline $s(x, \gamma)$ with respect to log time $x := \log t$. Note that this formulation includes Weibull hazards as a special case when there is a linear relationship between $\log H(t)$ and x. Moreover the specification of the natural spline to be linear beyond the range of the boundary knots,

implies a Weibull tail. The RP transition intensity models can be fitted directly using the *flexsurv* package in **R**.[40]

Following the guidelines in Royston and Parmar[64] we considered spline models of increasing complexity by placing the boundary knots k_{min}, k_{max} at the lowest and highest uncensored event times, respectively, and placing internal knots $k_1, \ldots, k_m$ with $k_1 > k_{min}$ and $k_m < k_{max}$ at quantiles of the distribution of uncensored event times. Given these assumptions, a natural cubic spline can be expressed as

$$s(x, \gamma) = \gamma_0 + \gamma_1 x + \gamma_2 \nu_1(x) + \ldots + \gamma_{m+1} \nu_m(x),$$

where γ_l with $l = 0, \ldots, m$ are unknown parameters to be estimated, and $\nu_j(x)$ is the jth spline basis function defined for $j = 1, \ldots, m$ as

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

where $\lambda_j = (k_{\max} - k_j)/(k_{\max} - k_{\min})$ and $(x - a)_+ = \max(0, x - a)$.

For instance, for the model with one internal knot point, the knot is placed at the median uncensored survival time to allow equal information to inform the two periods. For the copula models, the best AIC was achieved by allowing each marginal distribution to have an RP form with one internal knot. For the illness-death based models, it was only necessary to include one internal knot point for the transition to progression, with the RP model having a worse AIC for the other two transitions. Figure 3.5 shows a comparison of the fitted marginal cumulative hazard functions for PFS and OS and indicates an adequate fit compared to the Nelson-Aalen estimates.

In Table 3.2 the required parameters for our flexible RP model-based method are shown. The spline coefficients $\gamma_0, \gamma_1, \gamma_2$ refer to the non-parametric spline function representing the time-to-progression. The Weibull hazard function of death before progression is defined by the shape parameter α_{02} and scale parameter λ_{02}, while the Weibull hazard function of death after progression is given by

the parameters α_{12} and λ_{12}. The model parameters for the copula based models are provided in Table A.6 in Appendix A.

Table 3.2: Estimates of the model parameters for the generalized model-based method fit for the colon cancer data.

Parameter	RP model-based method	SE
γ_0	0.5894	0.1507
γ_1	2.4705	0.1555
γ_2	0.0997	0.0077
α_{02}	1.0600	0.1478
α_{12}	0.9619	0.0376
λ_{02}	79.8428	32.6204
λ_{12}	1.6106	0.0838

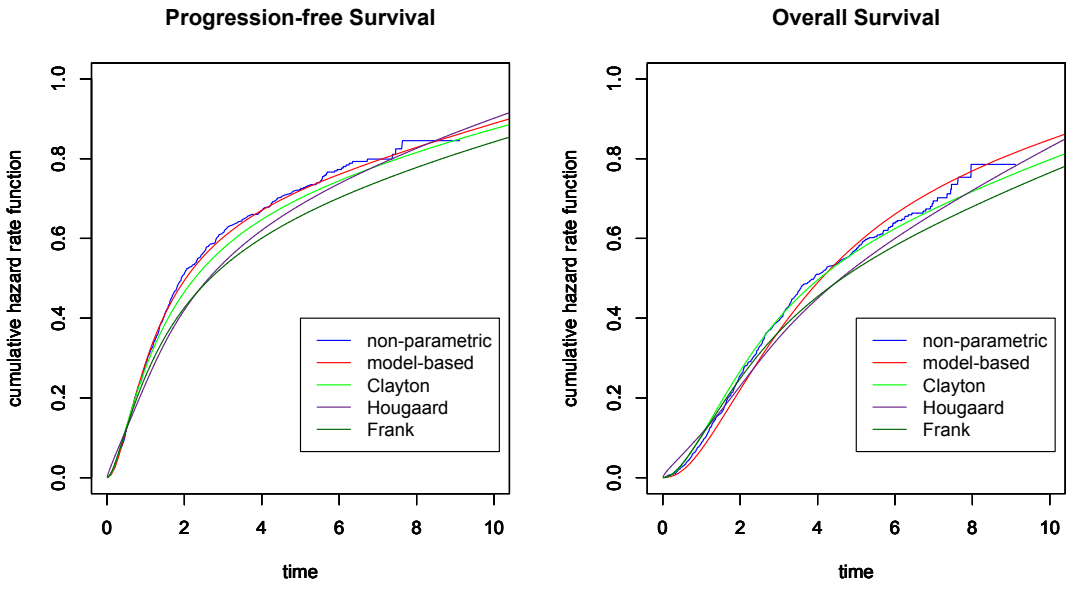

Figure 3.5: Nelson-Aalen and model-based estimates of the marginal cumulative hazard functions for PFS and OS where the parametric models are fitted using Royston-Parmar distributions

Based on AIC, the Clayton copula is preferred amongst the copula models since $(AIC_{Clayton} = 4242.255 < AIC_{Frank} = 4256.965 < AIC_{Hougaard} = 4355.2)$.

Similarly for the semi-parametric copula models, the Clayton model is preferred with ($AIC_{semiClayton} = -24.96 < AIC_{semiFrank} = -3.20 < AIC_{semiHougaard} = 181.79$).

Table 3.3: Comparison of estimates of Kendall's τ and associated standard errors for the colon cancer dataset.

model	Kendall's τ	standard error
generalized model-based method	0.836	0.01152
IPCW estimator	0.834	0.01084
one-stage fully parametric model(Clayton)	0.845	0.00921
one-stage fully parametric model (Hougaard)	0.736	0.01311
one-stage fully parametric model (Frank)	0.806	0.01003
two-stage semi-parametric model (Clayton)	0.830	0.00957
two-stage semi-parametric model (Hougaard)	0.668	0.12606
two-stage semi-parametric model (Frank)	0.767	0.03121

Table 3.3 gives the estimates of Kendall's τ derived from each of the methods. The example partially supports the findings of the simulation results. One similarity is the high sensitivity to choice of copula, as the values of the Kendall's τ differ in the different copula models as well. There is also reasonably close agreement between the model-based and the Clayton copula estimates.

Note that for this example, if we instead considered quantifying association based on Pearson correlation (as proposed by Fleischer *et al*[28] and Li and Zhang[48]) then using the flexible RP model-based method and estimating by simulation, returns an estimate exceeding 0.999. This is due to the estimate being dominated by the upper tail where there is strong agreement between PFS and OS since long survival times will correspond to those who did not progress and because the hazard of progression is estimated to decrease. The result highlights that Pearson correlation is often not a useful measure of dependence for survival

outcomes.

3.4.1 Separate estimates by treatment arm

After previously having pooled the treatment arms in our analysis, we can instead take treatment into account in our calculations. As mentioned above the colon data contains three treatment arms: the observation group, the group with levamisole (Lev) alone and the group with a combination of levamisole and fluorouacil (Lev+5FU). It is of interest whether Kendall's τ between PFS and OS is different for each treatment arm. Burzykowski *et al*[10] incorporated covariates into the calculation of Kendall's τ for copula models by allowing covariates to affect the marginal distributions of PFS and OS. In this way, a common Kendall's τ applies to all groups. Following this procedure, assuming proportional hazards for PFS and OS between each of the treatment groups, the Kendall's τ under a Clayton copula is 0.8464 (SE = 0.0093) which is almost unchanged from the estimate ignoring treatment effects.

Alternatively, we can fit completely separate models to each treatment arm. Table 3.4 shows the Kendall's τ for each treatment arm based on the RP illness-death model-based method and RP Clayton model, respectively. There is evidence that the treatment affects the degree of dependence between PFS and OS. In particular, Lev and Lev+5FU decrease the hazard of progression and hence increases the proportion of patients for whom PFS equals OS.

Table 3.4: RP model-based method: Kendall's τ and standard error in the three treatment arms

	RP model-based method		RP Clayton model		IPCW method	
Treatment arm	Kendall's τ	se	Kendall's τ	se	Kendall's τ	se
Control group	0.786	0.023	0.810	0.017	0.802	0.020
Treatment "Lev" group	0.804	0.023	0.816	0.018	0.805	0.021
Treatment "Lev+5FU" group	0.903	0.015	0.912	0.011	0.901	0.014

3.5 Discussion

The relationship between progression-free survival and overall survival can be investigated by copula based approaches, a non-parametric IPCW approach or illness-death model-based methods. As the copula based approaches ignore the fact that PFS cannot be longer than OS and the non-parametric IPCW method requires a strong assumption about censoring, another approach without these above drawbacks is of interest. The illness-death model-based method proposed by Li and Zhang[28] only offers a partial solution to these issues. In this chapter, we generalized the model-based method to allow estimation of Kendall's τ for general parametric models, potentially allowing dependence both on time since progression and time since randomization.

The simulation results in the previous section give insight on the issue of the most appropriate method for quantifying the association between PFS and OS in a series of scenarios. One notable result is the generally good performance of the Clayton copula model in realistic scenarios, being close to be unbiased and having efficiency similar to the illness-death model approach.

However, while the Clayton copula appears to perform well, it is clear that in general the estimate of Kendall's τ is sensitive to the choice of copula. These varied outcomes are primarily due to the different tail dependencies of the copulas. The Clayton copula may be most appropriate in cancer survival, as it focuses on the dependence in the lower tail of the bivariate density function, which captures the common situation where a quick time-to-progression leads to quick time-to-death. However, as seen in the unrealistic scenario, there are illness-death model scenarios where the Clayton copula model will be biased. If a copula based approach is to be used in practice, it is important to consider several different copula models and adopt the one with the best fit. However, results from the simulation study suggested AIC is not necessarily a reliable criterion

for finding the model with the least bias for Kendall's τ. Furthermore, it is questionable whether it is sensible to apply copula models even if they often produce reasonable estimates, given that they do not offer an admissible model for PFS and OS.

The generalized illness-death model-based method also relies upon the underlying model being close to correctly specified, with the simulations showing some bias arising if a homogeneous semi-Markov model is assumed when the true model is non-homogeneous. Assessment of the appropriateness of the model should therefore be considered. In addition, parametric assumptions about the transition intensities have to be taken into account. A fully non-parametric approach is not possible as the hazard functions cannot be identified beyond the maximum follow-up time. The use of hazard functions based upon flexible natural cubic splines reduces the danger of estimates being sensitive to the particular parametric specification. However, they still rely on the assumption of log-linearity of the hazards beyond the last knot point. Alternative non-parametric approaches aiming a restricted version of Kendall's tau will be discussed in the overall discussion in Chapter 6.

Incorporating the treatment effect on PFS and OS was not the focal point in this thesis. However, it is of interest to what extent the treatment is relevant for the relationship between PFS and OS. There are various ways in which covariates could be accommodated into the models for Kendall's τ. In the copula based models, such as in the work of Burzykowski *et al*,[10] covariates are allowed to affect the marginal distributions of PFS and OS, but the dependence between PFS and OS is common across all patients. For the illness-death models, the natural way to accommodate covariates effects is to allow separate effects for each of the transition intensities. This results in a different estimated Kendall's τ between PFS and OS depending on the value of covariates. In the colon cancer example there was clear evidence of the treatment affecting the degree of dependence be-

tween PFS and OS, with a higher estimated Kendall's τ in the Lev+5FU group. Since we would usually expect an effective treatment to alter the proportion of patients who progress before death, it seems reasonable that the Kendall's τ between PFS and OS would also differ. Hence, copula based methods, if pursued, should also incorporate covariate effects for τ itself. The next chapter concerns joint modelling of PFS and OS with incorporating treatment effects. The primary outcome of interest is then the measurements of the group difference of OS, but also the Kendall's τ will be considered.

CHAPTER 4

Joint modelling of PFS and OS in a Bayesian framework

4.1 Introduction

Chapter 3 explored joint models of PFS and OS to deduce the relationship between progression-free and overall survival and compared Copula-based, non-parametric and illness-death model-based methods. This present chapter is partially similar to the previous, but differ in its main focus. In this chapter, we also consider parametric joint modelling of PFS and OS, and some of the models overlap with the models from the previous chapter. However, in this chapter we consider the models within a Bayesian framework. Furthermore, our main aim in the modelling is to incorporate treatment effects to infer the group difference between the treatment groups for PFS and OS.

The remainder of the chapter is structured as follows. In Section 4.3, we describe two copula-based approaches and a multi-state model-based approach for joint modelling for PFS and OS. In addition, it is shown how the group difference and relationship in terms of AHR and Kendall's τ, respectively, is derived by each of the considered methods. Section 4.4 summarizes the results of simulation studies, where the methods are compared regarding group difference measurements and correlation values. In Section 4.5, the methods are illustrated on data from clinical ovarian cancer trials. This chapter concludes with a discussion.

4.2 General background and computation for Bayesian statistics

This subsection briefly summarizes the general Bayesian approach to estimation and computation in order to provide relevant background for the content in this chapter.

The idea of the Bayesian approach arises from Bayes' theorem.[5] The key

characteristic of the Bayes approach is the incorporation of external information into the statistical model.

In a frequentist approach, suppose the probability distribution $f(y|\theta)$ for observed data $y = (y_1, \ldots, y_k)$ given a vector of unknown parameters $\theta = (\theta_1, \ldots, \theta_k)$, where θ refers to a fixed but unknown parameter. In a Bayesian approach, θ is considered as a random quantity conditional on a prior distribution $\pi(\theta|\mu)$, where the μ refers the hyperparameters of the prior. The prior distribution is a probability distribution for θ containing information about it without involving any information from the observed data y. Here, we suppose that the hyperparameters are known, therefore $\pi(\theta) = \pi(\theta|\mu)$. Inference for θ is then yielded by the Bayes Theorem as follows:

$$p(\theta|y) = \frac{p(y,\theta)}{p(y)} = \frac{p(y,\theta)}{\int p(y,\theta)d\theta} = \frac{f(y,\theta)\pi(\theta)}{\int f(y|\theta)\pi(\theta)d\theta}. \tag{4.1}$$

In practice, the integral in the denominator is often high dimensional and intractable. As such Bayesian estimation requires specific methods for generating posterior samples. Markov chain Monte Carlo (MCMC) algorithms such as the Metropolis-Hastings algorithm[37,54] and the Gibbs sampler[34,35] are typically used for generating samples from posterior distributions. A brief overview of these algorithms will be presented below.

4.2.1 The Gibbs sampler

Suppose $\theta = (\theta_1, \ldots, \theta_k)$. The Gibbs sampler can be applied to sample from $p(\theta_1, \ldots, \theta_k)$ if the full conditional distribution for each parameter given the data and all other parameters is known and in a form from which it is straightforward to sample. The conditional distribution for each parameter being conditional on the other parameters and the observed data can be written as $p(\theta_i|\theta_i \neq \theta_j, y)$. After choosing a set of initial values for all $\theta_1^{(0)}, \ldots, \theta_k^{(0)}$, the algorithm follows the steps as shown below.[6] For every iteration $(t \in 1:T)$, apply:

Step 1: draw $\theta_1^{(t)}$ from $p(\theta_1|\theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_k^{(t-1)}, y)$

Step 2: draw $\theta_2^{(t)}$ from $p(\theta_2|\theta_1^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_k^{(t-1)}, y)$

$\vdots$

Step k: draw $\theta_k^{(t)}$ from $p(\theta_k|\theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_{k-1}^{(t-1)}, y)$.

If t is sufficiently large, then the simulated draws reflect the posterior distribution: $(\theta_1^{(t)}, \ldots, \theta_k^{(t)}) \overset{\text{approx}}{\sim} p(\theta_1, \ldots, \theta_k|y)$.

4.2.2 Metropolis-Hastings algorithm

The Gibbs sampler is straightforward, but it requires the full conditional distributions $p(\theta_1, \ldots, \theta_k)$ for each parameter to be of known form in order to sample from each distribution. However, cases might occur that a closed form of the full conditional distribution may not known.[6] The Metropolis-Hastings algorithm works, even if the closed form of full conditional posterior distribution for each parameter does not exist. It can be categorized as a rejection algorithm, as it includes a step of rejection sampling from specific candidate density. Following the description of the Metropolis-Hastings algorithm in Berry *et al*,[6] suppose we have joint posterior distribution $p(\theta|y) \propto h(\theta) \equiv f(y|\theta)p(\theta)$. A candidate density $q(\theta^*|\theta^{(t-1)})$ needs to be specified that is a valid density function for every realization of the conditioning variable $\theta^{(t-1)}$.

After choosing an initial value for $\theta^{(0)}$ at iteration $t = 0$, the following steps in the algorithm are as follows:

For every iteration: $t \in 1 : T$, repeat:

Step1: draw θ^* from $q(\cdot|\theta^{(t-1)})$

Step2: compute an acceptance ratio: $r = \frac{h(\theta^*)}{h(\theta^{t-1})} = \exp(\log h(\theta^*) - \log h(\theta^{(t-1)}))$

Step3: if $r \geq 1$ set $\theta^{(t)} = \theta^*$.

$$\text{If } r < 1, \text{ set } \theta^{(t)} = \begin{cases} \theta & \text{with probability} \quad r \\ \theta^{(t-1)} & \text{with probability} \quad 1 - r. \end{cases}$$

If t is sufficiently large, then the simulated draws reflect the posterior distribution: $\theta^{(t)} \overset{\text{approx}}{\sim} p(\theta|y)$.

4.3 Approaches to joint modelling of progression-related measurements and overall survival

In this section different parametric approaches for joint modelling of PFS and OS adapted to the Bayesian framework will be shown. Two of them are copula-based approaches for semi-competing risks data. The third model is a multi-state model-based approach. We will also use a Cox model approach in this chapter.

The standard approach to estimate the group difference between control and treatment, is the hazard ratio. However, despite the assumption of proportional hazards for each transition between control and treatment group, the three-state model leads to a non-proportional hazards assumption for overall survival between the treatment groups. Therefore we will focus on the average hazard ratio. It describes the average of the hazard ratio over through an estimated weight function. The AHR is characterized by the choice of the weight function. The general definition of the AHR is given in (2.9) in the Preliminary Methods. Schemper *et al*[68] chose a weight function, such that (2.9) simplifies to

$$AHR = \frac{\int_0^\infty h_{tr}(t)S_c(t)S_{tr}(t)dt}{\int_0^\infty h_c(t)S_c(t)S_{tr}(t)dt}. \tag{4.2}$$

Schemper *et al* showed that (4.2) is equal to the concordance odds definition of the AHR (see Section 2.3 for more details).

4.3.1 Copula based approach: Clayton model

One approach to jointly model PFS and OS is based on the Clayton copula function whose details can be found in Section 3.2.2, whereas the general idea of copulas are summarized in Section 2.2 in the Preliminary methods. The

marginal distributions of PFS and OS are assumed to each have separate Weibull distributions. A corresponding bivariate joint survival copula function based on Weibull distributions can be obtained by combining the marginal distribution for PFS and OS with a copula function.[10] The scale and shape parameters in the two hazard functions for PFS and OS as well as the copula dependence parameter δ can be jointly estimated by using maximum likelihood estimation within Bayesian framework. With respect to Kendall's τ, a useful aspect of Archimedean copula methods is that there is a direct link between Kendall's τ and the dependence parameter δ (Genest and MacKey).[36] As we model the survival times in a Bayesian framework, the prior distributions for the scale parameters $\log(\lambda_{PFS})$, $\log(\lambda_{OS})$, the treatment effects $\log(\theta_{PFS})$, $\log(\theta_{OS})$ and the shape parameters $\log(\alpha_{PFS})$, $\log(\alpha_{OS})$ for both the control and the treatment group have to be specified.

In general, copula-based approaches are convenient to jointly model survival data. However, for modelling PFS and OS the copula model does not restrict the ordering of those endpoints even though by definition the PFS time must be less than or equal to the OS time. Therefore, it is expected that the copula model leads to biased results in modelling the survival endpoints. Surprisingly, the Clayton copula-based method we considered in Chapter 3 for estimating the rank correlation coefficient Kendall's τ was seen to be close to unbiased in most of the scenarios considered. However, the performance of a copula-based approach is sensitive to the choice of copula. It is therefore important to consider several different copula models and adopt the one with the best fit.

Incorporating a treatment effect, we assume the treatment does not affect the shape parameter, but the baseline scale parameter λ as follows

$$\log(\lambda_T) = \log(\lambda_C) + \theta. \tag{4.3}$$

For the calculation of AHR for PFS and OS, we use (4.2) by plugging in the hazard functions and density function of PFS and OS calculated from the sample.

The model estimation in a Bayesian framework results in a sample from the posterior distribution for the model parameters to be estimated. In order to obtain an estimate of the AHR, we compute the AHR for each sample from that posterior distribution and then take the mean of all AHR's.

4.3.2 Normal induced copula estimation model

Fu *et al*[32] proposed a Bayesian semi-competing risks approach to jointly model TTP (time-to-progression) and OS and to explore the relationship between PFS and OS. As this approach uses a Gaussian copula to link the distribution of TTP and OS, it is called a normal induced copula estimation model (NICE model). Compared to the previous copula-based approach in Section 4.3.1, for this model the copula models the joint distribution with respect to TTP and OS, not PFS and OS. In particular, this approach allows the ordering of PFS and OS to be preserved in contrast to the Clayton copula-based approach. In order to construct the likelihood, the four possible observed cases are considered: (1) Either TTP and OS are observed, (2) TTP is observed but OS is censored and (3) TTP is censored by OS and (4) both TTP and OS are censored. This model including the likelihood function and the estimation method within Bayesian framework is described in detail in Fu *et al*. The model induces the relationship between PFS and OS. Correlation among these variables such as Kendall's τ and Spearman's ρ can be directly derived from the joint model. Fu *et al* estimate correlation values via Monte Carlo integration for single-arm trial based on simulation method. We adapt this NICE model later in our simulation. However, instead of exponential distribution for TTP and OS we use a Weibull distribution in order to reflect the survival course of a patient more realistically. Further, for the same reasons as in Chapter 3, we only consider Kendall's τ as outcome measures representing the correlation. As the NICE model is considered for a one-arm study and we consider here a two-arm trial, we apply the NICE model to both treatment groups

separately. After estimating the survival functions and hazard functions for PFS and OS for each treatment group, the AHR can be obtained by using Formula (4.2).

4.3.3 Multi-state model-based approach

The following approach is based on a multi-state model framework often used to analyze event history data of patients with cancer (please refer to Section 2.1 for full details).

Suppose we have a data set comprising N patients. Incorporating the individual's covariate effects on the hazards, we let the treatment affects the scale parameters $\tilde{\lambda}_{01,i}$, $\tilde{\lambda}_{02,i}$ and $\tilde{\lambda}_{12,i}$ given subject $i \in N$ shown as follows

$$\lambda_{01,i} = \tilde{\lambda}_{01} \exp(\beta_{12} X_i),$$
$$\lambda_{02,i} = \tilde{\lambda}_{02} \exp(\beta_{02} X_i),$$
$$\lambda_{12,i} = \tilde{\lambda}_{12} \exp(\beta_{12} X_i),$$

where vector X_i is the vector of covariates for all N individuals and β_{01}, β_{02} and β_{12} are the corresponding regression coefficients. The baseline hazards $\tilde{\lambda}_{01}, \tilde{\lambda}_{02}$ and $\tilde{\lambda}_{12}$ are fitted using the assumption of an exponential distribution or a Weibull distribution. Finally, the prior distribution has to be specified for all $\beta' s, \lambda' s$ and $\alpha' s$. Estimating the parameters of the multi-state model in Bayesian framework involves determining the likelihood contributions for the 4 possible cases of observed patient history.

The reason why we consider the average hazard ratio rather than the standard hazard ratio based on the illness-death model framework is explained in Section 2.3. Briefly summarized, despite the assumption of proportional hazards for each transition between control and treatment group, the three-state model induces a non-proportional hazards assumption for overall survival be-

tween the treatment groups. Consequently, the proportional hazards assumption is violated. Therefore, we use the average hazard ratio as an alternative tool to estimate the group difference. This type of analysis is a more flexible approach than the standard hazard ratio estimates, as it incorporates the effect of time on the group difference. It works well even if the proportional assumption of overall survival between the treatment groups is violated.[63]

Once the parameters for the three hazard functions are estimated, the estimate for the average hazard ratio can be derived. In addition, the steps of how to obtain the average hazard ratio will be described. Mathematically, the steps to estimate the AHR are as follows: deriving the transition probabilities from the estimated transition intensities of the illness-death model, the survival functions for OS can be calculated. We consider the transition probabilities $P_{00}(t_1, t_2)$ and $P_{11}(t_1, t_2)$ implying the probability to stay in state 0 and in state 1, respectively, within the time interval $[t_1, t_2]$. They can be rewritten as survival functions such as

$$P_{00}(t_1, t_2) = S_0(t_2 - t_1),$$
$$P_{11}(t_1, t_2) = S_1(t_2 - t_1).$$

The explicit expressions of the transition probabilities are given in Formulas (2.4) and (2.5), respectively, in the Preliminary Methods. Note, that the expression of $P_{11}(t_1, t_2)$ only works for a semi-Markov model if t_1 is the time the patient entered state 1.

Then the survival function for overall survival $S(t)$ and its estimate $\hat{S}(t)$ can be expressed by

$$\begin{aligned} S(t) &= P_{00}(t) + \int_0^t P_{00}(0,u)\pi_{01}(u)P_{11}(u,t)du, \\ \hat{S}(t) &= \exp\Big(-\Big(\frac{t}{\lambda_{01}}\Big)^{\alpha_{01}} - \Big(\frac{t}{\lambda_{02}}\Big)^{\alpha_{02}}\Big) + \int_0^t \Big(\frac{\alpha_{01}}{\lambda_{01}}\Big) * \Big(\frac{u}{\lambda_{01}}\Big)^{(\alpha_{01}-1)} \\ &\qquad \exp\Big(-\Big(\frac{u}{\lambda_{01}}\Big)^{\alpha_{01}} - \Big(\frac{u}{\lambda_{02}}\Big)^{\alpha_{02}}\Big) \exp\Big(-\Big(\frac{1}{\lambda_{12}}\Big)^{\alpha_{12}}(t-u)^{\alpha_{12}}\Big) du. \end{aligned} \tag{4.4}$$

The standard method to obtain an estimate of AHR is iteration-based. Let B be the number of iterations and $\theta_1, \ldots, \theta_B$ the samples from the posterior distribution $\pi(\theta|x)$. The AHR can be computed for each sample of the model parameters from the posterior distribution such as $AHR_1 = AHR(\theta_1), \ldots, AHR_B = AHR(\theta_B)$ by using Formula (4.2) for the AHR. Then the posterior mean of the AHR is estimated as

$$\overline{AHR} = \frac{1}{B}\sum_{i=1}^{B} AHR_i,$$

and the variance is

$$Var(AHR) = \frac{1}{B-1}\sum_{i=1}^{B}(AHR_i - \overline{AHR})^2$$

A credible interval can be constructed either by assuming the posterior is approximately normal and using the estimated standard deviation, or else through the percentiles of the AHR_b. This iteration-based option is quite straightforward as we get the estimated posterior parameters for all iterations directly, but in practical terms it is very time consuming.

An alternative approach to get an estimate of the AHR and its posterior standard deviation exists. We consider the mean of the posterior estimates for each model parameter and then we compute the AHR using Formula (4.2) as follows $\widehat{AHR} = AHR(\overline{\underline{\theta}})$, where $\overline{\underline{\theta}} = \frac{1}{B}\sum_{i=1}^{B}\theta_i$. In order to obtain the standard error the multivariate delta method can be applied to estimate the posterior variance of AHR. A finite differences numerical approximation can be used to find the derivative of AHR with respect to the model parameters. Having the gradient vector $\frac{\partial AHR}{\partial \theta}\Big|_{\theta=\overline{\theta}}$ the variance of the AHR defined by multivariate delta method is then

$$Var(AHR(\theta)) \approx \frac{\partial AHR}{\partial \theta}\Big|_{\theta=\overline{\theta}} Var(\theta) \frac{\partial AHR}{\partial \theta}\Big|^{T}_{\theta=\overline{\theta}},$$

where $Var(\theta) = \frac{1}{B-1}\sum_{i=1}^{B}(\theta_i - (\overline{\theta}))(\theta_i - (\overline{\theta}))^T$ is the covariance matrix. This method is quicker than the first method, but is based on a first order Taylor approximation. The first method is more accurate, but it is more time-consuming

as AHR has to be numerically computed B times. Therefore, we consider the second method.

In this chapter, one measure outcome we look at is the Kendall's τ for quantifying the relationship between PFS and OS. The Appendix B.1, in particular Subsection B.1.1, contains a description how to obtain the estimate for Kendall's τ due to a lack of its closed form expression based on the underlying model. In addition, the simulation-based method for how to get the standard error of the Kendall's τ is shown there.

Beside the above methods for joint modelling of PFS and OS, we also consider a standard Cox model approach. This approach is fitted using the standard Cox regression model and is based on assuming proportional hazards with respect to overall survival. The group difference is then estimated by the hazard ratio.

4.4 Simulation

4.4.1 Simulation Set up

A simulation has been conducted to test the performance of the Gaussian copula model, the Clayton copula model, the multi-state model based approach and the Cox model approach. We simulate 1000 datasets with a sample size of 1000 for each data, with 500 patients in each of the control and treatment groups. No other explanatory variables apart from treatment are considered in the simulations. As the illness-death model reflects the underlying survival process of patients with cancer, it is assumed to be the true model. In all four simulation scenarios, we assume the data are generated from a *homogeneous semi-Markov* model with Weibull transition intensity functions. Let the scale parameter be $\lambda_{01}, \lambda_{02}$ and λ_{12} as well as the shape parameter be α_{01}, α_{02} and α_{12} for each tran-

sition then the transition intensities are given by

$$\pi_{01}(t) = \alpha_{01}\Big(\frac{1}{\lambda_{01}}\Big)^{\alpha_{01}} t^{\alpha_{01}-1},$$
$$\pi_{02}(t) = \alpha_{02}\Big(\frac{1}{\lambda_{02}}\Big)^{\alpha_{02}} t^{\alpha_{02}-1},$$
$$\pi_{12}(s) = \alpha_{12}\Big(\frac{1}{\lambda_{12}}\Big)^{\alpha_{12}} s^{\alpha_{12}-1},$$

where t and s refer to time since randomization and to time since progression, respectively.

The simulation scenarios mainly differ with respect to which transitions are assumed to be affected by the treatment. As there is not one rule of pattern how the treatment affects the transitions, we choose different patterns to be the true model. In scenario 1, we choose values of the shape and scale parameters such that we have a situation with a treatment effect both on transition from study entry to progression and from progression to death prior to progression. In scenario 2, we consider a situation where treatment affects the transition between state 0 and 1, while death given progression is not affected by the treatment. In scenario 3, we assume a treatment effect on all three transitions and test for example how the parametric multi-state model-based approach assuming only treatment effects on transitions $0 \rightarrow 1$ and $1 \rightarrow 2$ performs. Scenario 4 represents a situation where the treatment affects transitions π_{01} and π_{12} such that the true hazard ratio decreases over time much faster than in the first scenario in order to see whether that setting makes a difference in terms of the performance of the methods.

In Table 4.1, the setting values of the scale parameters and shape parameters of the Weibull distributions used in each scenario are summarized. Graphics of the hazard function, survival function and cumulative hazard function of OS for scenarios 1-4 can be found in Appendix B.

Table 4.1: Parameter values for the simulation scenarios 1-4

	Scenario			
Parameter	1.Scenario	2.Scenario	3.Scenario	4.Scenario
$(\lambda_{01}, \lambda_{02}, \lambda_{12})_{control}$	(2.5, 9, 2.3)	(2.5, 9, 2.5)	(2.5, 7, 2.1)	(1.5, 9, 2.3)
$(\lambda_{01}, \lambda_{02}, \lambda_{12})_{treat}$	(5, 9, 2.6)	(5, 9, 2.5)	(3.1, 10, 2.5)	(5, 9, 6)
$(\alpha_{01}, \alpha_{02}, \alpha_{12})_{control}$	(0.8, 1, 1.2)	(0.8, 1, 1.2)	(0.9, 1, 1)	(0.7, 1, 1.2)
$(\alpha_{01}, \alpha_{02}, \alpha_{12})_{treat}$	(0.8, 1, 1.2)	(0.8, 1, 1.2)	(0.9, 1, 1)	(0.7, 1, 1.2)
low censoring	U[4, 10]	U[4, 10]	U[3, 10]	U[6, 10]
high censoring	U[2, 7]	U[3.5, 5]	U[2, 6]	U[2.5, 7]

As the rate of censoring might influence the performance of the methods, we consider two censoring cases for each simulation scenario. In case 1, 20% of patients are censored with respect to OS while in case 2 around 40% of OS times are censored. The censoring times arise from a uniform distribution.

The following section addresses the setting about the prior distribution for each considered method. The program JAGS (see Section 4.3 for more details on computation of the posterior distribution in Bayesian analysis) was used for the Clayton-based model to obtain the posterior distribution of the baseline scale parameters λ_{PFS} and λ_{OS}, the treatment effects θ_{OS} and θ_{PFS} and the dependence parameter ξ. Furthermore, the posterior distribution of the shape parameters $\alpha_{PFS_{control}}, \alpha_{OS_{control}}, \alpha_{PFS_{treat}}$ and $\alpha_{OS_{treat}}$ for the control and treatment group will be generated. The initial values of all parameters were set to zero. Like in the Clayton-based model approach, the posterior distribution of the parameters to be estimated were generated by JAGS within the framework of the multi-state model-based approach. The various parameters include scale parameters $\lambda_{01}, \lambda_{02}$ and λ_{12}, the baseline shape parameters α_{01}, α_{02} and α_{12}, and

the treatment effects θ_{01}, θ_{02} and θ_{12}. In both approaches, the prior density distribution of all parameters were set to be weakly informative. The initial values for all parameters were all set to zero prior density. For the Clayton-based approach and the multi-state model based approach, the output of the posterior results were based on 10000 iterations after a burn-in period of 5000.

The summary of the setting for the prior distributions for all parameters and the two approaches is shown in Table 4.2. Following the standard recommendation in the Clayton copula-models we parameterize the model on the log scale. Regarding the scale parameter for each transition in the multi-state model-based method, we used a half normal distribution truncated at 0.05 rather than a log-normal distribution because of the potential for the scale parameter λ_{02} to get close to 0.

Table 4.2: Setting of prior distribution of the parameter values for Clayton copula-based and multi-sate model-based approach

Approach	
Clayton copula-based (in JAGS)	multi-state model-based (in JAGS)
$\log(\lambda_{PFS}) \sim N(0.2, 3^{-2})$	$\lambda_{01} \sim N(0, 0.1)I[0.05,]$
$\log(\lambda_{OS}) \sim N(0.2, 3^{-2})$	$\lambda_{02} \sim N(0, 0.1)I[0.05,]$
$\theta_{PFS} \sim N(0.1, 3^{-2})$	$\lambda_{12} \sim N(0, 0.1)I[0.05,]$
$\theta_{OS} \sim N(0.1, 3^{-2})$	$\theta_{01} \sim N(0, 0.1)$
$\log(\alpha_{PFS_{control}}) \sim N(0.2, 3^{-2})$	$\theta_{02} \sim N(0, 0.1)$
$\log(\alpha_{OS_{control}}) \sim N(0.2, 3^{-2})$	$\theta_{12} \sim N(0, 0.1)$
$\log(\alpha_{PFS_{treat}}) \sim N(0.2, 3^{-2})$	$\alpha_{01} \sim N(0, 0.01)I[0.05,]$
$\log(\alpha_{OS_{treat}}) \sim N(0.2, 3^{-2})$	$\alpha_{02} \sim N(0, 0.01)I[0.05,]$
$\xi \sim uni(-1, 50)$	$\alpha_{12} \sim N(0, 0.01)I[0.05,]$

Instead of using the JAGS program to generate the posterior samples, the Metropolis–Hastings algorithm was used for the Gaussian copula-based approach to get the posterior distribution of the various parameters to be estimated. In

order to perform the NICE model estimation, we adapted the code by Fu *et al*.[32] Therefore, we will use the posterior mode as the estimate for each iteration while assuming a non-informative prior in terms of improper uniform priors. Because the Bayesian estimate uses a non-informative prior, it will be equivalent to the maximum likelihood estimate (MLE) of the NICE model. Compared to the NICE model in the paper, where exponential hazards for the transitions are assumed, we instead use Weibull hazards. Therefore we also need posterior estimates for the shape parameters. The scale and shape parameters $\lambda_{TTP}, \lambda_{OS}$, α_{TTP} and α_{OS} are required for both control and treatment group. In addition, the posterior density for the dependence parameter ρ of the Gaussian copula will be generated. The initial values for were all set to be $\lambda_{TTP} = 0.1, \lambda_{OS} = 0.1, \alpha_{TTP} = 0.1, \alpha_{OS} = 0.1$ and $\rho = 0.4$ for both treatment and control group. The outcome of all posterior distributions of the parameters is the result from 20000 MCMC simulations with a burn-in period of 1000 iterations.

4.4.2 Simulation Results

Figure 4.3 shows the estimated values of the average hazard ratio and its standard error from each method using 500 simulated data sets for the four scenarios under high censoring. The results under low censoring can be found in Table B.1 in Appendix B.1.2.

In Scenario 1, where treatment effects on transitions $0 \to 1$ and $1 \to 2$ are assumed, the Gaussian copula model seems to perform very well. The Gaussian copula model has the lowest standard error and the smallest bias compared to the other models. It is surprising, as it was expected that the model-based method would provide the least biased AHR estimate, as the scenario is generated from this model. The multi-state model has a slightly smaller standard error, but has a higher bias than the Cox model. The two versions of Clayton copula models give similar results, but their AHR are overestimated and biased. As

in scenario 1, the data in scenario 4 is also based on the assumptions that the treatment affects transitions $0 \to 1$ and $1 \to 2$, but with a higher treatment effect on both transitions. As expected the standard error and bias in all models are smaller than in scenario 1. The AHR in Gaussian copula model approach seems to be unbiased with the smallest standard error. The model-based method performs also well, and has a lower standard error and bias than in the Cox model. Due to the higher intensity of the treatment effect on the transitions , the performance of the two versions of Clayton copula methods have improved in scenario 4 compared to scenario 1.

Table 4.3: Simulation results for all 4 scenarios, high censoring

			Scenario			
	Model	Summary	1	2	3	4
		max. follow-up time	7	5	6	7
		true AHR	0.7556	0.8167	0.7406	0.4337
1	model-based method	AHR_{OS}	0.7407	0.7965	0.7348	0.4239
		SE_{AHR}	0.0592	0.0646	0.0599	0.0358
		$BIAS_{AHR}$	-0.0149	-0.0202	-0.0057	-0.0098
2	Clayton-copula based methods	AHR	0.8415	0.9172	0.7681	0.4504
	(non-proportional hazards assumption)	SE_{AHR}	0.0618	0.0700	0.0579	0.0373
		$BIAS_{AHR}$	0.0860	0.1005	0.0275	0.0168
3	Clayton-copula based methods	AHR	0.8470	0.9324	0.7799	0.4352
	(proportional hazards assumption)	SE_{AHR}	0.0596	0.0676	0.0563	0.0349
		$BIAS_{AHR}$	0.0915	0.1157	0.0393	0.0016
4	Gaussian-copula based approach	AHR	0.7566	0.8125	0.7420	0.4337
		SE_{AHR}	0.0384	0.0416	0.0376	0.0259
		$BIAS_{AHR}$	0.0010	-0.0043	0.0014	1.75E-05
5	Cox Model	HR_{COX}	0.7541	0.8085	0.7460	0.4206
		$SE_{HR_{COX}}$	0.0613	0.0661	0.0609	0.0360
		$BIAS_{HR_{COX}}$	-0.0015	-0.0083	0.0054	-0.0131

Scenario 2 assumes treatment affects only transition $0 \to 1$, and not $1 \to 2$. However, the structure of the results are still similar as in the previous scenarios. The Gaussian copula model seems to be preferred. The model-based method shows similar values as the Cox model. Furthermore, the copula-based methods differ slightly in their results, but with biased estimates and higher standard

errors than in the models. In scenario 3, where the data are based on assumption with treatment effects on all transitions, the performance of all models are very similar. Beside the AHR, the weighted Kendall's τ was calculated for the model-based method and the Clayton copula-based model. However, it is doubtful that the weighted Kendall's τ is meaningful in terms of the relationship when there is a situation with two or more treatment groups. The weighted Kendall's tau for the Clayton and multi-state model-based method are shown in Tables B.2 and B.2 under high and low censoring, respectively, in Appendix B.1.2. Comparing them, the performance of the model-based methods is superior compared to the two Clayton copula models in all scenarios.

4.5 Application

We applied the above methods to a dataset pooling data from 4 double-blind randomized clinical trials in advanced ovarian cancer.[59] The trials were conducted to compare the efficacy of cyclophosphamide plus cisplatin (CP) versus cyclophosphamide plus adriamycin plus cisplatin (CAP) to treat advanced ovarian cancer. In order to assess survival difference in the treatment of the ovarian cancer, overall survival is considered to be the primary endpoint and defined as time from randomization to death from any cause. In addition, the potential surrogate endpoint progression-free survival was also measured in all four trials and is defined as the time from randomization to death or clinical progression depending on what occurs first. The sample size of the pooled data is 1194, where 606 patients were assigned to cyclophosphamide plus cisplatin (CP) and 586 patients were assigned to cyclophosphamide plus adriamycin plus cisplatin (CAP). The characteristics of the single trials are shown in Omura *et al*.[59] The dataset is provided by the *frailtypack* package in **R**. The maximum folllow-up time was almost 14 years. As the dataset from the **R** package considers another time scale than the original data in Omura *et al*,[59] we multiplied the observed

times in the data from the package by 7 to get the same time scale as in the original data.

Table 4.4 shows the results of the AHR estimate and its standard error. There it seems that the model-based methods and the Gaussian-copula based approach seemed to perform well in terms of estimating the group difference being close to the hazard ratio of the Cox model. By contrast, the Clayton copula models seem to result in a weak performance due to potentially overestimated average hazard ratios. In order to support the values we illustrated the survival plots 4.1 for every method. However, they did not show an adequate fit to the data, as the survival curves indicate that none of the methods perform well. One way to solve this issue would be to incorporate Royston-Parmar (RP) flexible parametric models[64] into the models we considered. In Chapter 3 we applied the Royston-Parmar parametric model, as we experienced the same issue. Please refer to Section 3.4 for more details how the Royston-Parmar spline functions work.

Table 4.4: Application results for the Ovarian data set

	Model	AHR	posterior SD_{AHR}
1	model-based method	0.8383	0.0558
2	Clayton-copula based methods (non-proportional hazards assumption)	1.0229	0.0596
3	Clayton-copula based methods (proportional hazards assumption)	1.0434	0.0455
4	Gaussian-copula based approach	0.8511	0.0368
5	Cox Model	0.8484	0.0552

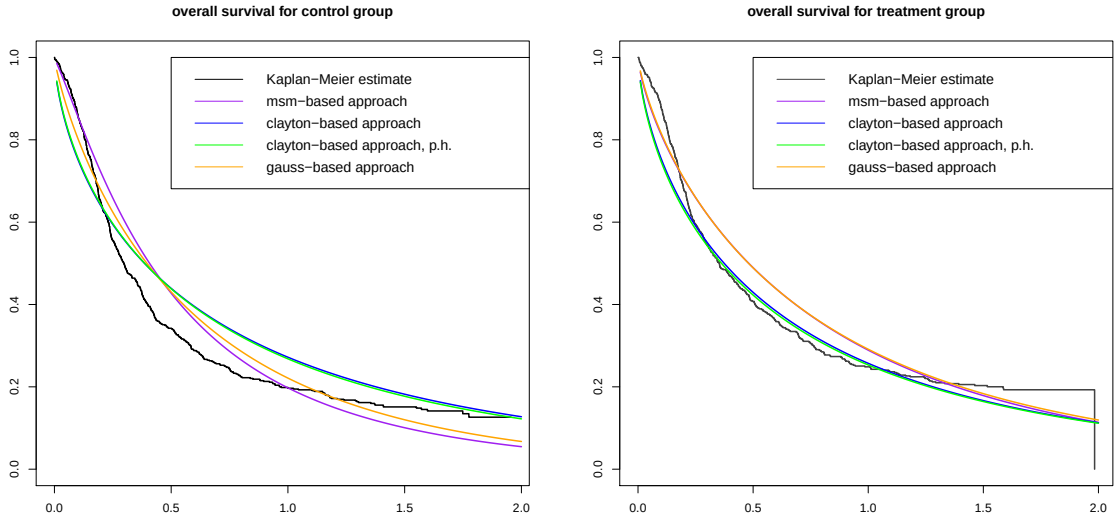

Figure 4.1: Survival plots for control group and treatment group based on all methods

4.6 Discussion

A multi-state model based approach and copula-based approaches to jointly model progression-free related measurements and overall survival were explored within a Bayesian framework. We investigated the performance of those approaches by focusing on the survival difference of both groups, assuming a two-arm clinical trial. We expected the multi-state model-based approach to be the most reasonable, convenient and precise approach in estimating the parameters and treatment effect on overall survival. We investigated whether the multi-state model-based approach still shows a benefit in the analysis of overall survival.

The motivation to use the average hazard ratio to express the group difference is that it is still relevant even without proportional hazards compared to the standard hazard ratio. Regarding the multi-state model, there will be definitely non-proportional hazards for PFS and OS. As the Clayton copula-based approach estimates PFS and OS directly, we have the choice to decide whether to allow non-proportional or proportional hazards for the endpoints. Under

an assumption of Weibull distributed data, assuming non-proportional hazards means allowing different shape parameters in the treatment groups for PFS and OS, respectively. In contrast, the assumption of proportional hazards implies assuming the same shape parameter for PFS in both treatment groups and the same shape parameter for OS in both treatment groups. The first assumption is more realistic and flexible but leads to some loss of efficiency.

In this chapter, we consider joint modelling of PFS and OS in a Bayesian framework. If there is prior information, then it can be easily incorporated into the model estimation in the Bayesian framework to make posterior distribution more informative. In theory, approaches might inform trial designs based on external information. The work about overall survival based on external information allows applications in some Bayesian contexts such as using posterior predictive sample to generate and design future trials. However, the external information is supposed to be informative and reliable. If there is information from several past trials, then the effect from a new trial is expected to come from the population effects observed in the previous trials. However, as we just use data from one trial, it is not sufficient to predict anything. Therefore, one extension might be to extend the approaches to approaches incorporating results from more several trials.

The simulation results show some surprising outcomes. Interesting is the extremely good performance of Gaussian copula model in terms of estimating the AHR, being close to be unbiased and being substantially more efficient than the multi-state model. That is surprising, as the data generation was based on the multi-state model for all scenarios. The better efficiency for the Gaussian copula may be due to only requiring estimation of two treatment effects whereas the multi-state model requires three.

The multi-state model and the Cox model result in a very similar performance. Whilst the bias was lower in the Cox model, the multi-state model based

methods showed a slight gain in efficiency in some scenarios. One notable result is the weak performance of the Clayton copula-based model. Overall, the simulation results show an biased and overestimated AHR and a loss in efficiency. That is in contrast to the good performance of the Clayton copula in Chapter 3. In that chapter it was noted that the performance of the copula method depended on the choice of copula and the degree of association between PFS and OS. Here, the poor performance of the Clayton copula in estimating the AHR might also be related to the underlying model misspecification. The application results are not meaningful, as none of the models showed a good fit to the data we used. Comparisons of the estimated survival plots of OS indicated discrepancies in all cases. Therefore the estimated AHR based on the copula do not provide a reliable conclusion and comparison, as the models seem to be misspecified in modelling the survival function. Further work regarding this issue relates to modifying the models by incorporating the idea of Royston-Parmar models.

CHAPTER 5

Gaining efficiency in oncology trials using multi-state model-based methods for modelling the survival times

5.1 Introduction

It is of interest to establish more efficient ways of estimating the treatment effects on overall survival. Instead of using surrogate endpoints with the intention to replace the true outcome of interest, auxiliary variables such as the information of time-to-progression of the disease can be used to provide improved efficiency in analyzing the effect of treatment on overall survival. This chapter explores a multi-state model-based approach for gaining efficiency in testing for, and estimating the treatment effects on, overall survival in oncology trials compared to the standard methods based on Cox regression or the log-rank test. In consideration of the multi-state model-based approach, the previous two chapters and this chapter differ in parametric assumptions. While Chapters 3 and 4 assume fully-parametric multi-state model-based approaches to model progression-free related measurements and overall survival, the framework of this chapter is semi-parametric.

Conlon *et al*[18] utilized the time-to-progression as auxiliary information for overall survival by using separate models for time-to-progression and time-to-death. A cured fraction model was used to model the time-to-progression. Based on the model for time-to-death multiple imputation was applied to generate death times for censored subjects from their conditional distributions, and these new calculated data were used in the primary analysis of overall survival. Conlon *et al*[19] extended the approach by deriving a multi-state model with an incorporated cured fraction for progression to jointly model time-to-progression and time-to-death. This joint multi-state model is parametric, with Weibull intensities are assumed. The cured fraction is modelled using the mixture model formulation of the cure model. The model is semi-Markov conditional on the patient not being in the cured fraction is used.

This parametric model is used to impute death times for censored sub-

jects aiming to improve the efficiency of the analysis of overall survival. In a simulation and in a particular example,[19] the multiple imputation procedure within the model was shown to provide an improvement in efficiency of the analysis of overall survival. However, this approach is somewhat dependent on the assumption of a cured fraction, as the approach benefits from involving a cured fraction. Further, the cure model assumes treatment only affects the probability of being cured and that only affects time-to-progression, not any post-progression survival. A consequence might be that the general analysis of treatment effects on OS based on the presence of a cured fraction might be improved. This method is then more suitable for clinical trials where the assumption of a cured fraction is plausible and there are no direct treatment effects on time-to-death after progression.

In this chapter, we focus on a semi-parametric method based on a multi-state model with three states as randomization, progression and death. Prior to commencement of a trial it may be possible to a priori exclude the possibility of treatment effects on some of the transitions between states. In particular, a necessary condition for surrogacy of PFS is that the treatment only affects the transition to progression.[61] However, we may be willing to assume no treatment effect on death prior to progression but not on death post-progression. It is of interest to vary the method by assuming different parametric assumptions regarding treatment effects among the transitions in order to see how different settings of circumstances affect the efficiency in analyzing the overall survival.

The treatment effects will be modelled by individual Cox models for each transition of the illness-death model. The dependence of the transitions on time can be specified by either Markov or semi-Markov assumptions (see Section 2.1 for more details). The performance of the semi-parametric multi-state model-based method might be affected by the particular assumption. Whether a semi-Markov or Markov assumption is most appropriate depends on the underlying

process. While a semi-Markov model resets the clock to zero upon entry into a new state, the Markov model uses a "clock forward"-model. In other words, the hazard of death given progression depends in a semi-Markov on time since progression, rather than time since randomization in a Markov model. The Markov assumption is computationally/theoretically simpler but the paper by Shu *et al*[70] indicates that a semi-Markov model is more efficient, at least with respect to the overall survival estimates.

To estimate the treatment effect on OS, the hazard ratio is the standard measure outcome in many clinical trials. However, if separate treatment effects for each transition in an illness-death model are assumed, then OS will not have proportional hazards (see Section 2.3 for more details). And that case of non-proportional hazards might induce a hazard ratio which won't be that meaningful.[63] The average hazard ratio proposed by Kalbfleisch and Prentice[43] is a generalization of the hazard ratio, appropriate even under non-proportional hazards. The average hazard ratio is based on a flexible weighting function to incorporate the impact of time on the treatment effect(see Section 2.3 for more details).

A reasonable comparison of the semi-parametric approach mentioned above is the Cox regression model. Typical analysis of the treatment effect on OS would be through the use of a log-rank test, with treatment effect estimates based on the hazard ratio estimated through Cox model possibly supplemented by estimated differences in survival at specific time points. However, such analyses do not utilize the early information on time-to-progression. A further reason why the log-rank test may not be efficient is that there may be non-proportional OS hazards between treatment groups. The situation of non-proportional hazards may affect the estimates of the group difference based on the Cox model. Struthers and Kalbfleisch[73] stated that under non-proportional hazards the Cox model estimates an average hazard ratio where the weights are dependent on the censoring distribution.

The remainder of the chapter is structured as follows. In Section 5.2, we discuss a typical analysis of the treatment effect on overall survival and review the Colon approach.[19] Subsequently, we describe both the semi-Markov and Markov multi-state model-based method and how we estimate the treatment effect on overall survival. Section 5.3 shows results of simulation studies, whereas the performance of the methods, in terms of bias and efficiency, are investigated and compared through several simulation scenarios. In Section 5.4, the semi-parametric multi-state model-based method is illustrated on the colon cancer dataset. The chapter concludes with a discussion.

5.2 Approaches to improve the efficiency in analyzing the overall survival

In this section, we focus on various approaches to investigating the treatment effects on overall survival in oncology trials.

5.2.1 Semi-parametric semi-Markov illness-death model

The multi-state model can be used as a framework for approaches intending to be efficient in estimating the treatment effect on overall survival. It is the basis for the approach we consider in this section in terms of jointly modelling the time-to-progression and the time-to-death. Here, we consider an illness-death model as depicted in Figure 2.1 in the Preliminary Methods. The states ”0”, ”1”, and ”2” correspond to ”randomization”,”progression” and ”death”, respectively, and the three transition intensities are given by π_{01}, π_{02} and π_{12}.

In this section, a semi-parametric semi-Markov model-based approach will be presented. Investigating to what extent this model improves the efficiency in estimating the treatment effect on overall survival will be of main interest. In particular, we will focus on how parametric assumptions about the treatment effect

and information about the time until progression can benefit the analysis of overall survival. Considering the characterization of the semi-parametric illness-death model, the hazard functions are modelled by assuming a non-parametric hazard function, common baseline hazards and proportional hazards among the treatment groups. Different parametric assumptions about the treatment effect can be made, when fitting the model using the standard Cox regression model. Here, we consider the following four possible assumptions about the treatment effect: (1) separate treatment effects on all transitions; (2) treatment affects both transition $0 \to 1$ and transition $1 \to 2$; (3) treatment only affects transition $0 \to 1$; (4) same treatment effect on all transitions.

We estimate the cumulative baseline transition hazards for each transition from the Cox regression model, separately. This procedure can be extended to the case with covariates such as treatment in order to obtain the cumulative transition hazards for the treatment group. The estimation of the $1 \to 2$ transition intensity differs between the Markov and semi-Markov cases. Using the Cox regression model, the extent how the Cox partial likelihood affects the risk set depends on the assumption regarding Markov and semi-Markov. For example, if a patient dies 5 months after progression, the risk set in a semi-Markov case are the patients who are alive 5 months after progression. However in a Markov case, it does not matter whether 5 months after progression or 3 years after randomization, so the risk set are all the patients in the progression state. .

The derived estimates are the basis to study the treatment effect on overall survival. Note that, despite the assumption of proportional hazards for each transition between control and treatment group, the three-state model leads to a non-proportional hazards assumption for overall survival between the treatment groups (see for more information Section 2.3). Consequently, the proportional hazards assumption is violated and as a consequence our analysis will be in terms of the average hazard ratio. This type of analysis is a more flexible

approach than the standard hazard ratio estimates as it incorporates the effect of time on the group difference. It works well even if the proportional assumption of overall survival between the treatment groups is violated.[63] However, the method still requires proportional hazards within each individual transition intensity so will not necessary work well in all cases.

After estimating the parameters for the three hazard functions, the estimate for the average hazard ratio can be derived. In addition, the steps of how to obtain the average hazard ratio will be described. Mathematically, the steps to estimate the AHR are as follows: deriving the transition probabilities from the estimated transition intensities of the illness-death model, the survival functions for OS can be calculated. We consider the transition probabilities $P_{00}(t_1, t_2)$ and $P_{11}(t_1, t_2)$ implying the probability to stay in state 0 and in state 1, respectively, within the time interval $[t_1, t_2]$. They can be rewritten as survival functions such as

$$P_{00}(t_1, t_2) = S_0(t_2 - t_1),$$
$$P_{11}(t_1, t_2) = S_1(t_2 - t_1).$$

The explicit expressions of the transition probabilities are given in Formulas (2.4) and (2.5), respectively, in the Preliminary Methods. Note, that the expression of $P_{11}(t_1, t_2)$ only works for a semi-Markov model if t_1 is the time the patient entered state 1.

In a semi-Markov illness-death model, the survival function for overall survival $S(t)$ and its estimate $\hat{S}(t)$ can be expressed by

$$\begin{aligned} S(t) &= P_{00}(t) + \int_0^t P_{00}(0, u)\pi_{01}(u)P_{11}(u, t)dt, \\ \hat{S}(t) &= \hat{P}_{00}(t) + \sum_{j, t_j \leq t} \hat{P}_{00}(0, t_j)\hat{\pi}_{01,j}\hat{P}_{11}(t_j, t) \\ &= \hat{S}_0(t) + \sum_{j, t_j \leq t} \hat{S}_0(t_j)\hat{\pi}_{01,j}\hat{S}_1(t - t_j), \end{aligned}$$

where $\hat{\pi}_{01}$ is the increment of $\hat{\Pi}_{01}(t)$, the estimate of the cumulative hazard function.

For the definition of AHR, let $f(t)$ be the marginal density and $h(t)$ be the hazard function for both groups. The survival function for both the treatment group and the control group as well as the hazard function for both the treatment group and the control group are donated by $S_{tr}(t)$, $S_c(t)$, $h_{tr}(t)$ and $h_c(t)$, respectively.

The general definition of the AHR[43] is given in Formula (2.9) in the Preliminary Methods. One component in this expression is the weight function, which characterizes the influence of time on the hazard ratio. Computing the AHR with a particular choice of weighting function[68] leads to the concordance odds definition of AHR shown in Formula (2.10).

Here, we consider a time-restricted AHR written by AHR_{res}. As we work here with non-parametric survival estimates, the AHR is calculated until the maximum follow up time. The concordance odds definition of the restricted AHR is relevant for the further work in this chapter, so we need to define the weight function as in Schemper *et al*[68] to simplify the general definition of the AHR to the concordance odds definition of the AHR. The weight function is $w(t) = \frac{S_c(t)f_{tr}(t)+S_{tr}(t)f_c(t)}{f_c(t)+f_{tr}(t)}$. Then the version of the average hazard ratio in Section 2.3 modifies to

$$\begin{aligned} AHR_{res} &= \frac{\int_0^\tau \frac{h_{tr}(t)}{h_c(t)+h_{tr}(t)} w(t) f(t) dt}{\int_0^\tau \frac{h_c(t)}{h_c(t)+h_{tr}(t)} w(t) f(t) dt} \\ &= \frac{\int_0^\tau h_{tr}(t) S_c(t) S_{tr}(t) dt}{\int_0^\tau h_c(t) S_{tr}(t) S_c(t) dt} \\ &= \frac{\mathbb{P}[T_{tr} < T_c, T_{tr} < \tau]}{\mathbb{P}[T_c < T_{tr}, T_c < \tau]}, \end{aligned} \tag{5.1}$$

where $\mathbb{P}[T_{tr} < T_c, T_{tr} < \tau]$ is the concordance probability with randomly chosen survival times T_{tr} and T_c from the treatment group and the control group, respectively, conditionally on T_{tr} being shorter than the censoring time. Formula (5.1) is equivalent to using a weight function that is as $w(t)$ for $t < \tau$ and then

identically 0 afterwards. Mathematically, the estimate of the concordance probability $\hat{\mathbb{P}}[T_{tr} < T_c, T_{tr} < \tau]$ can be estimated by

$$\hat{\mathbb{P}}[T_{tr} < T_c, T_{tr} < \tau] = \frac{\sum_{t_j \leq t} \hat{S}_c(t_j) d\hat{S}_{tr}(t_j)}{(1 - \hat{S}_{tr}(\tau))\hat{S}_c(\tau)}.$$

After estimating the concordance probabilities, the AHR can then be estimated as follows

$$\widehat{AHR}_{res} = \frac{\widehat{\mathbb{P}}[T_{tr} < T_c, T_{tr} < \tau]}{\widehat{\mathbb{P}}[T_c < T_{tr}, T_c < \tau]}. \tag{5.2}$$

In order to test the significance of the estimated AHR_{res}, the standard error needs to be computed. One way to obtain the standard errors/confidence intervals of the estimated average hazard ratio is by the non-parametric bootstrap approach. As it is a simulation based approach, we firstly generate B samples from the original set of data. We resample the data by sampling all the data from an individual, at which we allow replacement. B is supposed to be large in order to provide robust estimates of the standard error. The next step is to estimate the average hazard ratio $\widehat{AHR}_{res_b}$ for every $b \in [1, B]$. The standard error of the $\widehat{AHR}_{res}$ is then yield by

$$\widehat{SE}(AHR_{res}) = \frac{1}{1 - B} \sum_{b=1}^{B} (\widehat{AHR}_{res_b} - \overline{AHR}_{res})^2,$$

where $\overline{AHR}_{res}$ is the mean of all average hazard ratio estimates for all $b \in \{1, \ldots, B\}$, such that $\overline{AHR}_{res} = \frac{1}{B} \sum_{b=1}^{B} \widehat{AHR}_{res_b}$. Then the confidence intervals can be constructed based upon the standard error assuming approximate normality. Alternatively, Shu *et al*[70] presented an asymptotic theory for a Cox semi-Markov illness-death model, which could in principle be extended to derive the variance of the AHR.

It is of interest to establish whether the semi-parametric model-based method gains any efficiency by making parametric assumptions about the treatment effects on the transitions compared to the Cox regression model. If the assumptions represent the structure of observed data well, then we may gain efficiency.

Further, another characteristic of the model which could potentially improve the performance of the analysis of overall survival is the assumption of a semi-Markov instead of a Markov model. In a Markov illness-death model based method, there is no clock reset to 0 after entering state 1. Hence, the entry into state 2 only depends on the time spent in current state 1, though not on how long the patient stayed in the previous state 0. According to the results of the simulation study in Shu *et al*, some apparent improvement in MSE for the semi-Markov model estimate of the survivor function could be found when either a semi-Markov model was true or both a Markov and semi-Markov model held. Here we investigate by simulation whether a similar result holds for the AHR between two groups.

5.3 Simulation

5.3.1 Simulation set up

In this section, the performance of the methods is studied through simulation. The motivation for choosing the following simulation scenarios is to explore the relative performance of the semi-parametric model-based method and the Cox model in a range of situations. One objective is to investigate how sensitive the models are in case of any model misspecification. Another focus of investigation is whether the assumption of a semi-Markov multi-state model leads to an improved efficiency in the analysis of OS compared to the assumption of a Markov multi-state model. The role of censoring is also an aspect to be considered in terms of how the intensity of censoring changes the performance of the semi-parametric model-based methods and Cox approach.

It is assumed that the true underlying model is an illness-death model as this plausible framework reflects the underlying disease process. In the first four simulation scenarios the data are generated from a *homogeneous semi-Markov*

model with Weibull distributed transition intensities. As we incorporate treatment effects on the hazards, we let the treatment affects the scale parameter of each transition. Assuming N patients, the scale parameters incorporating the treatment effects are given by

$$\lambda_{01,i} = \tilde{\lambda}_{01} \exp(\beta_{12} X_i),$$
$$\lambda_{02,i} = \tilde{\lambda}_{02} \exp(\beta_{02} X_i),$$
$$\lambda_{12,i} = \tilde{\lambda}_{12} \exp(\beta_{12} X_i),$$

where $\tilde{\lambda}_{01}$, $\tilde{\lambda}_{02}$ and $\tilde{\lambda}_{12}$ are the baseline scale parameters, the vector X_i represents the vector of covariates for all N individuals and β_{01}, β_{02} and β_{12} are the corresponding regression coefficients. Mathematically, the Weibull transition intensities with the shape parameters $\alpha_{01}, \alpha_{02}, \alpha_{12}$ for all three transitions are given here by

$$\pi_{01}(t) = \big(\frac{1}{\lambda_{01}}\big)^{\alpha_{01}} \alpha_{01} t^{\alpha_{01}-1},$$
$$\pi_{02}(t) = \big(\frac{1}{\lambda_{02}}\big)^{\alpha_{02}} \alpha_{02} t^{\alpha_{02}-1},$$
$$\pi_{12}(s) = \big(\frac{1}{\lambda_{12}}\big)^{\alpha_{12}} \alpha_{12} s^{\alpha_{12}-1},$$

where t and s refer to time since randomization and to time since progression, respectively.

The first four scenarios differ with respect to which transitions are affected by the treatment. As there is not one rule of pattern how the treatment affects the transitions, we choose different patterns to be true model. In scenario 1 we choose values of the shape and scale parameters such that we have a situation with a treatment effect both on transition from study entry to progression and from progression to death prior to progression. An illustration of the hazard functions for each transition intensity and for overall survival for scenario 1 are shown in Figure 5.1 and in Figure 5.2, respectively. In scenario 2 we consider a situation where treatment affects the transition between state 0 and 1.

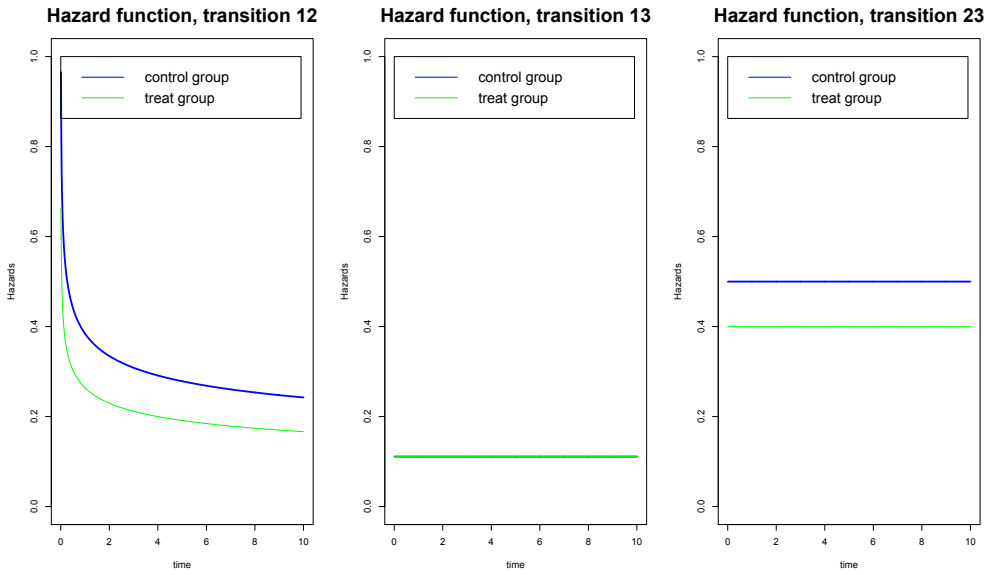

Figure 5.1: Hazard functions for simulation scenario 1

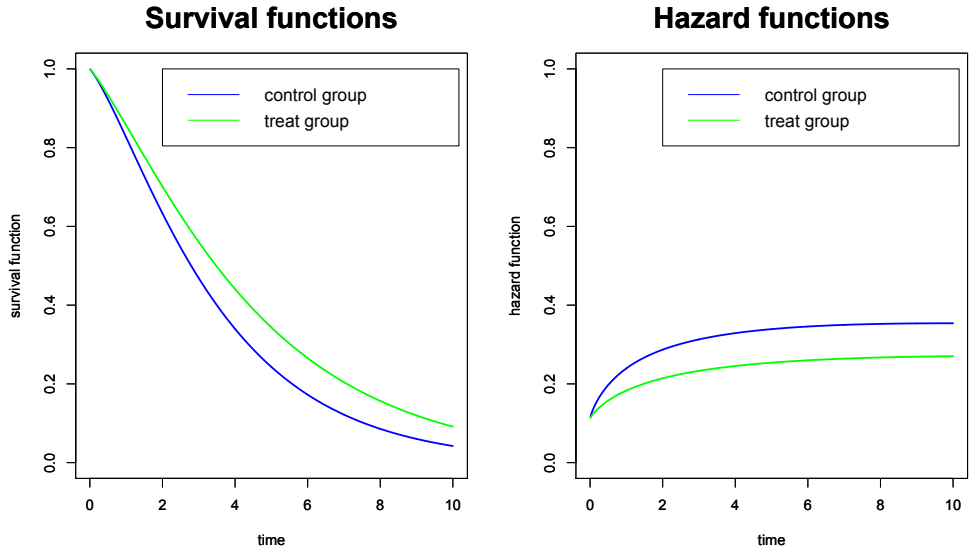

Figure 5.2: Survival function of OS and cumulative Hazard function for simulation scenario 1

The two following scenarios are considered in order to investigate the robustness of the semi-Markov setting in case of some special types of misspecification of the illness-death model. In scenario 3 we assume a treatment effect on all three transitions and test, for example, how the semi-parametric multi-state

model based approaches assuming only treatment effects on transition $0 \rightarrow 1$ and $1 \rightarrow 2$ perform.

In scenario 4 we do not assume proportional hazards for each transition. Considering this case, we seek to explore how sensitive the performance of the semi-parametric model is to violation of the proportional hazards assumption for each transition in an illness-death model.

The simulation scenarios are now chosen to focus in particular on the performance between a semi-Markov and a Markov semi-parametric model-based approach. It is expected that a semi-Markov illness-death model reflects the underlying disease process better than a Markov model. Assuming an exponential distribution as in scenarios 1, 2 and 3 time-to-death given progression represents both a semi-Markov and Markov framework for the transition from progression to death.

Scenario 5 is then simulated based on a Markov illness-death model. By the use of a truncated piecewise exponential distribution for transition to death given progression, survival times given progression can be generated within a Markov framework. The transitions $\pi_{01}(t)$ and $\pi_{02}(t)$ have then the same definition as in scenario 1-4. In order to create a fully Markov illness-death model it is also possible to consider a left truncated Weibull distribution for time-to-death given progression. However, the reason why we show a scenario based on a truncated piecewise exponential distribution is that the temporal effect of a left truncated Weibull distribution is often quite concentrated near $t = 0$ and there is a less pronounced hazard ratio at later times.

A left truncated distribution for transition from progression to death implies a Markov scenario, as the time-to-progression is considered as censored and hence it doesn't reset the clock to 0 upon entry into state 1. In order to construct a left truncated distribution, we choose a piecewise exponential hazard

function $h(t)$ and a time point t_0 such that

$$h(t) = \begin{cases} \lambda_1 & \text{if} \quad t < t_0 \\ \lambda_2 & \text{if} \quad t > t_0. \end{cases}$$

The two exponential times $T_1^* \sim Exp(\lambda_1)$ and $T_2^* \sim Exp(\lambda_2)$ contribute then time-to-death as follows. If progression occurs at time t_p, then

$$T^* = \begin{cases} T_2^* & \text{if} \quad t_p > t_0 \\ \min(T_1^*, t_0 - t_p) + I(T_1^* > t_0 - t_p) * T_2^* & \text{if} \quad t_p < t_0. \end{cases}$$

The time-to-death given progression since randomization is the sum of times t_p and T^*.

Scenario 5 is created with treatment effects on transition $0 \to 1$ as well as $1 \to 2$. The robustness of the semi-Markov model will be tested by applying it to scenario 5.

Instead of simulating survival times given progression based on a Markov assumption, we now consider fully semi-Markov processes in scenarios 6 and 7. Having a semi-Markov case makes it possible to investigate whether the semi-parametric semi-Markov illness-death model-based method works better than the Markov version. This scenarios are chosen in order to investigate whether the semi-Markov approach is more efficient than the Markov approach when the process is truly semi-Markov.

In scenarios 6 and 7, the hazard function for time-to-death given progression follows a Gamma distribution which makes it possible to define more extreme semi-Markov cases easily. We can choose shape parameter k and scale parameter θ such that time-to-death given progression follows a semi-Markov assumption. For example we construct a case where within the first year after progression death is very unlikely, though after one year the hazard function increases quickly. That setting implies a strong semi-Markov effect even if the scenario is somewhat unrealistic as it implies an initial decrease in the hazard of

death following progression. In scenario 6, we assume the treatment affects all three transitions. We consider two censoring cases for each simulation scenario. Case 1 has 20% of patients whose OS time is censored while case 2 has 40% of OS times censored. The censoring times arise from a uniform distribution.

Tables 5.1, 5.2 and 5.3 show the setting values of the scale parameters and shape parameters according to the distributions used in each scenario. The illustration of the hazard function and survival function and cumulative hazard function of OS for scenarios 2-7 can be found in Appendix C.1.

Table 5.1: Parameter values for simulation scenarios 1-4

	Scenario			
Parameter	1.Scenario	2.Scenario	3.Scenario	4.Scenario
$(\lambda_{01}, \lambda_{02}, \lambda_{12})_{control}$	(2.5,9,2.1)	(2.5,9,2.5)	(2.5,7,2.1)	(2,9,2.2)
$(\lambda_{01}, \lambda_{02}, \lambda_{12})_{treat}$	(5,9,2.8)	(5,9,2.5)	(3.1,10,2.5)	(3.7,9,2.4)
$(\alpha_{01}, \alpha_{02}, \alpha_{12})_{control}$	(0.8,1,1)	(0.8,1,1)	(0.9,1,1)	(0.9,1,1.13)
$(\alpha_{01}, \alpha_{02}, \alpha_{12})_{treat}$	(0.8,1,1)	(0.8,1,1)	(0.9,1,1)	(0.7,1,1.2)
low censoring	U[3,12.5]	U[3,13]	U[3,10]	U[3,10]
high censoring	U[2,6.7]	U[3.5,5]	U[2,6]	U[3,5]

Table 5.2: Parameter values for simulation scenario 5

Parameter	5.Scenario
$(\lambda_{01}, \lambda_{02})_{control}$	(2,7)
$(\lambda_{01}, \lambda_{02})_{treat}$	(3,7)
$(\alpha_{01}, \alpha_{02})_{control}$	(0.9,1)
$(\alpha_{01}, \alpha_{02})_{treat}$	(0.9,1)
$(\lambda_1, \lambda_2)_{control}$	(1.5,7)
$(\lambda_1, \lambda_2)_{treat}$	(2,8)
t_0	2
low censoring	U[3,10]
high censoring	U[2,8]

Table 5.3: Parameter values for simulation scenarios 6 and 7

	Scenario	
Parameter	6.Scenario	7.Scenario
$(\lambda_{01}, \lambda_{02}, \lambda_{12})_{control}$	(1.5,9,0)	(1.5,9,0)
$(\lambda_{01}, \lambda_{02}, \lambda_{12})_{treat})$	(2.3,9)	(2.3,9,0)
$(\alpha_{01}, \alpha_{02}, \alpha_{12})_{control}$	(0.9,1,0)	(0.9,1,0)
$(\alpha_{01}, \alpha_{02}, \alpha_{12})_{treat}$	(0.9,1,0)	(0.9,1,0)
$(\theta_{12})_{control}$	0.9	0.8
$(\theta_{12})_{treat}$	0.9	0.8
$(k_{12})_{control}$	6	7
$(k_{12})_{treat}$	7	7
low censoring	U[6,14]	U[6,13]
high censoring	U[3,12]	U[3,12]

5.3.2 Results

Results with the mean estimate of AHR, and the empirical SE and bias from each method using 1000 simulated data sets for the above eight scenarios under high censoring are shown in Table 5.4. The respective results for the same scenarios, but under low censoring, can be found in Table C.1 in Appendix C.1. The semi-parametric model-based methods are flexible regarding the setting of treatment effect assumptions. They can be modelled by assuming different scenarios of the observed data. In the following, we consider scenarios based on three types of treatment effect assumptions: (A) treatment effect on transitions $0 \to 1$ and $1 \to 2$; (B) treatment effect on transition $0 \to 1$; (C) independent treatment effect on all three transitions. Table 5.4 shows the performance of the model-based methods, where the methods are modelled such that they use the correct treatment effect assumption considering the corresponding true scenario.

Table 5.4: Results for all scenarios, high censoring

Scenario	True HR	maximum follow-up time	Model								
			semi-Markov model-based method			Markov model-based method			Cox model		
			AHR	SE	BIAS	AHR	SE	BIAS	AHR	SE	BIAS
true scenario: treatment effect on transitions $0 \to 1$ and $1 \to 2$:											
1. scenario	0.7059	6.7	0.7062	0.0422	0.0004	0.7058	0.0422	-0.0001	0.7058	0.0582	-2.84E-05
4. scenario	0.7541	5	0.7676	0.0460	0.0136	0.7587	0.0459	0.0046	0.7545	0.0610	0.0004
5. scenario	0.8197	8	0.7970	0.0506	-0.0227	0.7994	0.0474	-0.0203	0.8062	0.0659	-0.0135
6. scenario	0.7359	12	0.7436	0.0412	0.0078	0.7409	0.0437	0.0050	0.7280	0.0597	-0.0079
true scenario: treatment effect on transition $0 \to 1$:											
2. scenario	0.8182	5	0.8152	0.0273	-0.0030	0.8148	0.0274	-0.0034	0.8198	0.0673	0.0016
7. scenario	0.9484	12	0.9288	0.0176	-0.0196	1.0074	0.0108	0.0590	0.9376	0.0754	-0.0109
true scenario: treatment effect on transitions $0 \to 1$, $1 \to 2$ and $0 \to 2$:											
3. scenario	0.7406	6	0.7416	0.0607	0.0010	0.7410	0.0607	0.0004	0.7447	0.0609	0.0042

Scenarios 1, 4, 5 and 6 assume that the treatment affects transitions $0 \to 1$ and $1 \to 2$. Both model-based methods show gains in efficiency in terms of the standard error in contrast to the Cox model. However the bias in the Cox model is predominantly smaller than in model-based methods. Comparing the semi-Markov model-based method with the Markov model-based method, they result in a similar performance. Despite a slight difference in terms of the standard error and bias between the two model-based methods, there is not a clear gain in efficiency for one of those specific methods.

Scenario 5 assumes a non-homogeneous Markov illness-death model. This scenario was used to investigate how incorrectly assuming a non-homogeneous semi-Markov model affects the performance of the method. It is somewhat surprising that the results of the semi-Markov model-based method indicate a robustness to such a type of misspecification since the individual survivor function estimates would certainly be expected to be biased. In order to understand why the AHR, its standard error and bias are so close to the corresponding estimates in the Markov model-based method, we examine the average estimated survival curves. Figure 5.3 shows the average estimated survival curves of OS between the treatment groups based models for scenario 5. Both survival curves corresponding to the semi-Markov model-based method are biased implying that indeed the Markov case affects the survival curve of a semi-Markov model-based method. The curves of the log hazard ratio over time for the semi-parametric model-based method based on both semi-Markov case and the Markov case are very similar, illustrated in Figure 5.4. A possible explanation for the good performance of the semi-Markov method despite the bias might be that the AHR smooths the bias of the survival curves out, as we assume non-proportional hazards of the survival function between the treatment groups.

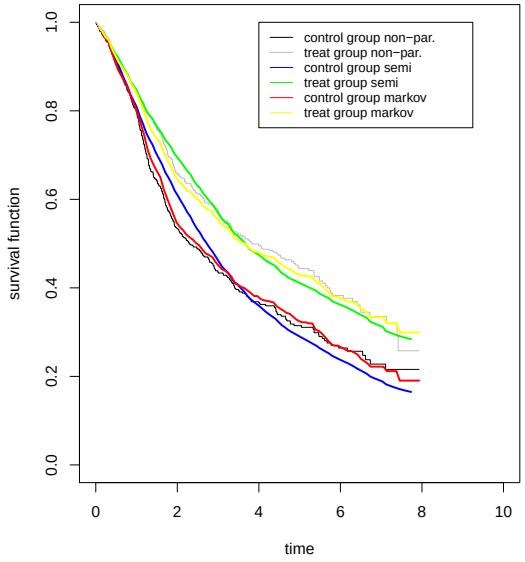

Figure 5.3: The survival curve of OS based on Kaplan-Meier, various semi-parametric model-based methods

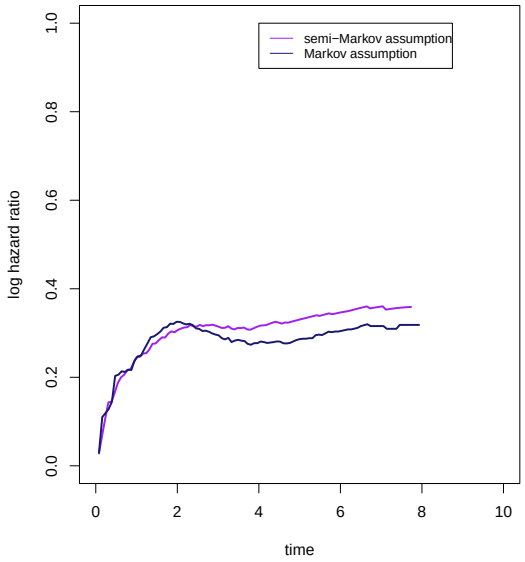

Figure 5.4: The log hazard ratio of OS based on semi-parametric model-based methods with semi-Markov assumptio versus Markov assumption

Scenario 6 is generated based on a semi-Markov assumption, therefore a better performance for the semi-Markov model might be expected. However,

the results seem to indicate very similar performance for the Markov and semi-Markov models. The standard error is slightly lower in the semi-Markov model, but the bias is higher in contrast to the model-based method with the Markov model.

Scenarios 2 and 7 are generated by assuming a treatment effect only on transition $0 \to 1$. The outcome is very similar to the previous results. The model-based methods are substantially more efficient in terms of the standard error compared to the Cox model which shows a slight benefit regarding the bias though. Focusing on the difference of performance among the two model-based methods, there is not a clear advantage for one of the methods. In scenario 3, all transitions are affected by the treatment with a Weibull distributed hazard for time-to-progression and exponential distributed hazard for time-to-death given progression and prior death. In that scenario, the semi-parametric model-based methods perform better in terms of both the bias and the standard error. However, there is not really a difference between the semi-parametric model-based methods regarding the Markov and semi-Markov assumption.

Besides the efficiency of the model-based methods, it is of interest to assess how these methods perform in case of any model misspecification about the assumptions regarding treatment effects. Specifically we consider cases where the absence of a treatment effect is assumed for one or more transitions in which a non-zero effect in fact exists. We show this type of investigation relating to scenario 7 in the following table. Table 5.5 shows what happens to the power of the model-based methods applied scenario 7, where a treatment effect only on transition $0 \to 1$ is assumed. The considered types of the model-based method are A, B, C and D. Assumption B is the true one, as it assumes the treatment effect on transition $0 \to 1$. The results of the model-based methods and the Cox model using assumption B are already given in Table 5.4.

Table 5.5: Investigation of performance of the model-based method for scenario 7, when assumption about treatment effect is misspecified, where
assumption A: treatment effect on transitions $0 \to 1$ and $1 \to 2$;
assumption B: treatment effect on transition $0 \to 1$;
assumption C: independent treatment effect on all transitions;
assumption D: same treatment effect on all transitions.

Scenario 7 (maximum follow up=12, true HR=0.9484)				
Results	Semi-Markov model-based method with different treatment effect assumptions			
	A	B (true assumption)	C	D
AHR	0.9300	0.9288	0.9307	0.7907
SE_{AHR}	0.0505	0.0176	0.0719	0.0460
Power	30.8000	100.0000	19.7000	98.6000
Results	Markov model-based method with different treatment effect assumptions			
	A	B (true assumption)	C	D
AHR	0.9268	1.0074	0.9276	0.7889
SE_{AHR}	0.0539	0.0108	0.0741	0.0434
Power	29.3000	8.4000	19.7000	99.3000
Results	Cox Model			
HR_{COX}	0.9376			
$SE_{HR_{COX}}$	0.0754			
Power	13.8000			

Comparing the power for that particular scenario among the methods, the semi-Markov model-based method results in a power of 100%, the Markov model-based method shows a power of 8.5%, even lower than the power of the Cox model. A possible explanation for this noticeable difference among the model-based methods might be that the AHR based on the Markov assumption setting leads to an overestimated value of AHR slightly greater than 1 probably due to misspecification. Consequently, there is no rejection of the null hypothesis implying no treatment effect of OS, whereas the true value of 0.9484 indicates a slight treatment effect. The power of the Cox model is quite low despite the almost unbiased estimate of the AHR. Hence, the corresponding standard error is relatively high, therefore the confidence interval covers often the value 1 in the analysis of the 1000 simulated data. As expected the semi-Markov model is most

powerful assuming the correct treatment effect assumption. The power drops strongly, if the semi-Markov model-based method presumes that assumption C involves independent treatment effect on all three transitions.

The power investigation of the remaining scenarios is shown in Appendix C.1. Regarding power results for the other scenarios, sometimes the model-based methods are quite robust to misspecification, such that the power is not adversely affected. However, it depends on the scenario and the treatment effect situation of the observed data, how the power of the model-based method performs in case of any misspecification. The power of the Cox model/log-rank test depends on the scenario. In the scenarios, where we assume a treatment effect on transitions $0 \to 1$ and $1 \to 2$, the log-rank test has a relatively high power, but mostly smaller than the power for the model-based methods. For the scenario where we assume only treatment effect on transition $0 \to 1$, the log-rank test was not able to detect the treatment effect whereas the semi-Markov model based methods did.

5.4 Application

The above methods were applied to the colon cancer data[55] in order to illustrate their performance in terms of analyzing OS. In fact, this dataset is the same used for the application in Section 3.4. Recalling the description of the clinical trial, its objective of interest was to assess the performance of two adjuvant therapies in improving surgical cure rate in stage III colon cancer. Patients were assigned to observation, the treatment levamisole alone or a combination of levamisole plus fluorouracil. The measure outcomes were time-to-progression and time-to-death. The observed survival data results from 929 patients who were followed up for 5 years or more (median follow-up, 6.5 years). During the trial, 425 individuals died, 54 were censored after progression and 423 were censored

before progression. The maximum follow-up time was 9.1 years, by which point based on the Kaplan-Meier estimates.

It is not obvious what assumptions about the treatment effects and the type of illness-death model apply to the colon dataset. In order to get a more detailed impression of the dataset, we investigate whether the true scenario in the data is more based on a Markov or semi-Markov illness-death model and which transition is affected by the treatment. Figure 5.5 shows the estimated survival curves based on Kaplan-Meier and the model-based methods among the control group and for treatment group levamisole plus fluorouracil.

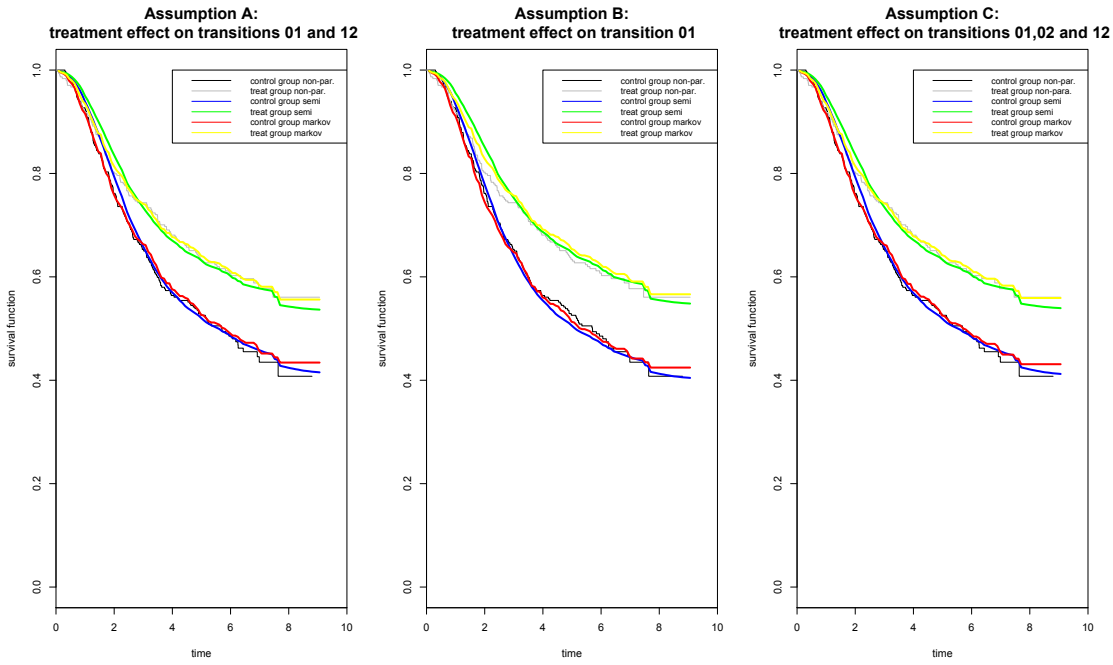

Figure 5.5: The estimated survival curves of OS among treatment group levamisolePlus and control group based on Kaplan-Meier and various semi-parametric model-based methods

Assuming a treatment effect on transitions $0 \to 1$ and $0 \to 2$ (assumption A), the estimated survival curves of the model-based methods are very close to the non-parametric survival curves for the treatment and control group, respectively. Comparing the Markov model-based method with the semi-Markov

model-based method, the estimated survival curves are very similar. Although there seems to be a little benefit for the method with the Markov assumption for the treatment group, as the corresponding survival curves seem to be closer to the non-parametric survival curve. Furthermore, the model-based methods with the assumption B result in similar estimated survival curves, whereas there seems no clear advantage for either the semi-Markov or Markov assumption.

Table 5.6 shows the AHR estimates between the control group and treatment group levamisole plus fluorouracil, the respective standard errors for the semi-Markov model-based method, Markov model-based method and the Cox model. As it is not clearly known what the true underlying treatment effect assumption of the observed data might be, the results for the model-based methods based on 4 different treatment assumptions A, B, C and D are considered. The results for the both model-based methods with assumption B involving treatment effect on transition $0 \rightarrow 1$ show lower standard errors compared to those for the models with assumption A, C and D. Hence, a benefit exist for assumption B. However, the estimated treatment effects for each transition, modelled by the Cox regression model, case A seems to be the most plausible assumption. As shown in Table C.9 in Appendix C.2, the treatment effect for transition $0 \rightarrow 1$ is highly significant while the $1 \rightarrow 2$ is mildly significant in the counter direction. Hence, assumption B is probably not reasonable despite the benefit for the assumption according to Table 5.6. Moreover, the standard error for the model-based method with assumptions and A and B is lower than the standard error for the Cox model. However, the AHR estimates of the model-based methods for assumption C are the closest to the AHR estimate of the Cox model.

Table 5.6: Investigation of performance of the model-based method for Colon data.
assumption A: treatment effect on transitions $0 \to 1$ and $1 \to 2$;
assumption B: treatment effect on transition $0 \to 1$;
assumption C: independent treatment effect on all transitions;
assumption D: same treatment effect on all transitions.

Colon data(maximum follow up=9.1), comparison of treatment treatment effect on OS between control group and arm levamisole plus fluorouracil)				
Results	Semi-Markov model-based method with different treatment effect assumptions			
	A	B	C	D
AHR	0.7215	0.6509	0.7120	0.8274
SE_{AHR}	0.0803	0.0655	0.0860	0.0985
$P-value$	0.0039	3.07E-05	0.0058	0.1181
Results	Markov model-based method with different treatment effect assumptions			
	A	B	C	D
AHR	0.7081	0.6496	0.6986	0.8005
SE_{AHR}	0.0765	0.0661	0.0836	0.0923
$P-value$	0.0017	3.46E-05	0.0033	0.0574
Results	Cox Model			
HR_{COX}	0.6888			
$SE_{HR_{COX}}$	0.0818			
$P-value$	0.0017			

5.5 Discussion

We investigated how information about the time-to-progression can improve the efficiency in terms of estimating the treatment effect on OS under different assumptions regarding which transitions are affected by treatment. Our main measure outcome to express the group difference is the average hazard ratio, as it is still meaningful without proportional hazards of overall survival among the treatment group hazard ratio.

Based on the simulation results, the multi-state model based method performs well in some settings. In particular, a gain in efficiency of the model-based methods in contrast to the Cox regression can be achieved, provided it is pos-

sible to assume no treatment effect on at least one of the transition intensities. However, the benefit depends on the type of data, as the bias of the hazard ratio based on the Cox regression was quite often lower than the AHR based on the multi-state model-based method.

Comparing the Markov multi-state model-based method with the semi-Markov multi-state model-based method in terms of efficiency gains, there is only a slight difference and it depends on the data which type of the model-based method shows benefits. The results indicated that there was little or no difference in efficiency between the semi-Markov and Markov models when estimating AHR. The Markov model-based method was quite robust to misspecification, when we assume a true scenario based on an non-homogeneous semi-Markov model, and the semi-Markov model seems to be more sensitive to misspecification than the Markov model. It is somewhat surprising that the AHR can smooth biased survival estimates out in our simulation quite well in case of misspecification, such that the AHR estimates seem to be much less biased than expected.

Using the multi-state model-based method in practice requires making assumptions about the analysis before collecting the data. These decisions involve the choice of the Markov or semi-Markov assumption and which treatment effects could be excluded. It might be likely to incorrectly assume how the treatment affects the course of disease of the data to be analyzed. We considered the power in our simulation, in order to control and to explore to what extent the power suffers from the misspecification. Due to the results the misspecification of the treatment effect can have quite an effect to the power. Furthermore, the procedure can only partially control the Type I error rate error. Essentially if the hypothesis H_0 is that there is absolutely no effect of treatment for any transition then Type I error is controlled. However, the Type I error is not controlled with respect to the set of all processes for which AHR=1. An AHR with that

value can imply a case where an effect of treatment exists for the transitions, but calculating the AHR can cancel out all the individual effects.

A further issue to be considered is the use of the average hazard ratio. We don't assume proportional hazards of overall survival among the control and treatment group. Hence, hazards are expected to change over time and the hazard ratio. The average hazard ratio is able to cope with non-proportional hazards well as it considers group difference depending on time. The value of AHR describes the average of the hazard ratio over time through a estimated weight function. The AHR depends on the way of defining the weight function. In this chapter, our weight function is defined such that the AHR represents the concordance odds definition of AHR. As we assume non-parametric survival curves, we modified the definition of the AHR[68] by restricting it up to the maximum follow-up time.

A commonly used alternative to get the treatment effect estimates is based on the Cox regression model. The hazard ratio based on this standard model is not optimal under proportional hazards of overall survival between the control and the treatment group. Neverthless, the hazard ratio estimates of the Cox model worked quite well in the simulation. However, there is no guarantee the Cox model estimate will be close to the concordance odds definition of the AHR. Struthers and Kalbfleisch[73] investigated the form of a Cox model under non-proportional hazards. The HR estimate based on a Cox model is heavily dependent on the censoring distribution. It is expected that the simulation results would provide much more biased results of the hazard ratio when exponential rather than uniform distributed censoring were used. However, the scenarios in the simulation were chosen to ensure the Cox estimate is reasonably close to the AHR.

The standard method to test for treatment effects is the log-rank test. It is purely non-parametric and equivalent to performing a score test in a Cox

model with one binary covariate. However, the log-rank test is most powerful under proportional hazards of overall survival between the treatment groups. Therefore, the test might not be appropriate for many applications, as proportional hazards are not always realistic. An alternative of the log-rank test is the weighted log-rank test which can better accommodate situations where the treatment effect on OS is expected to be more pronounced at either the beginning, middle or end of a clinical study. A weighted log-rank test can be represented by Flemington-Harrington test,[44] which is seen as a variation of the standard log-rank test. In a Flemington-Harrington test it is allowed to choose the weights, whereas one corresponds to the weights for earlier survival times and the other one corresponds to the more weights for later survival time. We applied the weighted log-rank-test to all our scenarios. However, there was not a real benefit using a weighted log-rank test in our simulation cases compared to the standard log-rank test.

Further work might refer to the improvement of design in oncology trials by using semi-parametric multi-state model-based methods. More details will be discussed in the subsequent Chapter 6.

CHAPTER 6

Thesis Conclusions and Further Work

6.1 Overview

This thesis has concerned topics relating to time-to-event analysis in oncology trials. We explored the performance of various statistical models for quantifying the association between progression-free and overall survival, and exploiting the relationship to estimate treatment effects. This chapter outlines conclusions, limitations and further work arising from the research-related three previous chapters.

6.1.1 Conclusions

The research in Chapter 3 does not result in one preference of the various methods we considered for quantifying association between PFS and OS. The simulations suggest that the illness-death model–based method provides good estimates of Kendall's τ across several scenarios. In some situations, copula-based methods perform well however their performance is sensitive to the choice of copula. The Clayton copula is most appropriate in scenarios, which might realistically reflect an oncology trial, but the use of copula models in practice is questionable.

Chapter 4 explored various methods to jointly model PFS and OS within a Bayesian framework and test for the group difference in terms of the average hazard ratio. According to the simulations results the Gaussian copula-based model performed the best overall, but the illness-death model-based approach showed a good performance as well. However, in contrast to the good performance of the Clayton copula based approach in Chapter 3, it does not show a good behaviour due to overestimated average hazard ratios.

Chapter 5 explored various methods for gaining efficiency in testing for, and estimating the treatment effects on, overall survival in oncology trials compared to the standard methods based on Cox regression or the log-rank test.

The performance was investigated by simulation for a wide range of scenarios and illustrated using data from a clinical trial of treatments for colon cancer. In most of the situations, the semi-parametric multi-state model-based method performs better than the Cox-based approach. In the scenarios considered there was little, and not significant difference in the performance under Markov and semi-Markov.

6.1.2 Limitations and Further Work

We expressed the association for quantifying the relationship between PFS and OS by the use of Kendall's τ. In the discussion of Chapter 3, we concluded that a fully non-parametric approach, based on the copula-based approach and illness-death model-based approach, for obtaining Kendall's τ is not possible when the support of the censoring distribution does not contain the support of the PFS and OS distributions, due to the inability to non-parametrically estimate the hazard functions beyond the maximum follow-up times. An alternative non-parametric approach would be to aim to estimate a restricted version of Kendall's τ. For instance, rather than compute the unconditional probability of concordance and discordance, one could replace the probability terms in (3.1) with ones conditional on $S_1 \wedge T_1 < t^*, S_2 \wedge T_2 < t^*$ for some choice of t^*, where (S_1, T_1) and (S_2, T_2) are two independent realizations of (S, T). A similar approach was employed in the context of estimating concordance odds in an Aalen additive hazards model[52] and we also use a similar idea in Chapter 5 to define an estimable version of the average hazard ratio. Provided t^* is less than or equal to the maximum follow-up time, a consistent model-based estimate of the restricted Kendall's τ can be obtained from, for instance, the Nelson-Aalen estimates of the transition intensities under a Markov assumption. Note, however, that restricting the definition of Kendall's τ would also allow consistent estimation using IPCW method with the added advantage of not requiring a

particular Markov or semi-Markov assumption.

Furthermore, in a situation with two or treatment groups, the Kendall's τ can be modified to a weighted Kendall's τ for measuring a weighted rank correlation. In Chapter 4, we calculated a weighted Kendall's τ rather than using the Kendall's τ, at which the treatment groups are pooled. However, it is doubtful how meaningful the weighted Kendall's τ would be in a situation with two or more treatment groups.

Throughout the thesis it has been assumed that both time-of-death and time-to-progression can be continuously observed up to right-censoring. While such an assumption is commonly applied, in practice assessments of progression are intermittent resulting in different right-censoring times for progression and death and interval-censored progression times.[75] For instance, for the multi-state model-based approach as in Chapter 3, provided the assessment times for progression are known, the model can be fitted using a likelihood that properly accounts for the intermittent observation.[17,42] However, adaptation of other approaches (such as the copula-based method and the IPCW-method in Chapter 3) where PFS is a composite measure of progression, which may be interval-censored, and of OS, which is right censored, is less straightforward.

When estimating treatment effects on overall survival in Chapters 4 and 5, we used the average hazard ratio. An illness-death model induces non-proportional hazards of overall survival among the control and treatment group. The average hazard ratio is able to cope with non-proportional hazards well as it considers group difference depending on time. The AHR weights the hazard ratio over time via a specified weight function. In Chapter 5 we used a truncated version of the weight function that gives a concordance odds interpretation to the AHR. Alternative outcome measures instead of the AHR include the difference of survival estimates between the treatment groups at a prespecified time point[19] and the difference in restricted mean survival time.[65] The semi-

parametric multi-state models in Chapter 5 can be the basis for calculating the estimated difference which can be calculated at a specific time. A non-parametric approach would consist of the estimation of separate Kaplan-Meier survival estimates for each group and calculation of the difference in survival at a given point. It would be interesting to see whether there are similar gains in efficiency using the Cox multi-state model with these treatment effect measures.

Moreover, the decision how to measure progression in cancer trials is of relevance. Throughout this thesis, we have assumed a binary assessment of progression. However, progression is often defined in terms of tumour size or tumour growth. As such progression could be viewed as a continuum rather than a binary event which suggests survival methods directly using tumour measurements could be used. In theory, methods like that such as a joint longitudinal-survival modelling have the potential to provide more information than just a survival model with binary measurements for progression. For instance, there might be situations where progression is defined in terms of tumour growth and the measurements result in being close to criterion for progression, but not all requirements are met. It would be reasonable to expect the overall survival of such patients is more similar to those who did progress than patients with little or no tumour growth. As such the survival model utilizing the tumour measurements provides more information than just a binary survival model for progression.

While this thesis has focussed on the analysis of oncology trial data, related future work could consider how the methods impact the design of trials. In particular, one could consider how to design a phase III trial in oncology that incorporates the semi-parametric multi-state model-based method described in Chapter 5. The necessity to control the Type I error rate in a phase III trial potentially precludes directly using the method in Chapter 5, since Type I error will be inflated if either the model is misspecified (i.e. treatment effects on transi-

tion intensities are not proportional) or the set of intensities without a treatment difference is misidentified. Instead, the idea would be to keep the primary endpoint as overall survival, assessed using a non-parametric method such as the log-rank test, but to use the semi-parametric multi-state model-based method for interim analyses of the trial for the basis of stopping for futility. If the assumptions of the semi-parametric multi-state model-based method are correct it would lead to making better decisions at interim analyses, thus shortening the average trial length or providing greater power for the same average number of patients. Whereas, if the model is incorrect there is no impact on Type I error since any decision regarding a significant difference would be on the basis of a log-rank test. Please refer to Muller and Schafer[56] for full details of the general general principle of adaptive group sequential designs. They presented a general approach combining the concept of adaptive interim analyses and classical group sequential testing.

Such an approach would perhaps be most useful for multi-arm trials assessing two or more treatments. In such a situation, the semi-parametric multi-state model, if extended to the case of three or more groups, could be used to decide which, is any, treatments to proceed with at interim analyses. As before, the final analysis of the treatment effect on overall survival could still be based on log-rank tests.

A further extension of the model in Chapter 4 would be to include a frailty term between time-to-progression and time-to-death after progression. This is an alternative way of allowing non-Markov dependence between time-to-progression and time-to-death after progression which should be relatively easy to implement within a Bayesian framework. However, the inclusion of a frailty term would further complicate the expression for the average hazard ratio.

Finally, a possible extension of the work from Chapter 4 would be to extend the approach for a single trial to one which pools information across several tri-

als. According to Buyse,[11] in order to establish PFS as a surrogate endpoint, a meta-analytic approach is needed. Applying the approach to one trial is not as meaningful as applying it to more trials for testing PFS as a surrogate endpoint for OS. However, previous work in this vein has used a copula approach, therefore extending the multi-state modelling approach to the analysis of several trials, allowing the baseline hazards and treatment effects to vary across trials as random effects, would be potential work for the future. The data we used in the application in Chapter 4 involve the aggregation of 4 trials for ovarian cancer and as such would be a good example to work with. However, the dataset provided in **R** does not specify which patient belongs to which trial.

Bibliography

[1] A. Alonso and G. Molenberghs, *E*valuating time to cancer recurrence as a surrogate marker for survival from an information theory perspective, Statistical Methods in Medical Research, 17 (2008), pp. 497–504.

[2] P. K. Andersen, S. Z. Abildstrom, and S. Rosthoj, *C*ompeting risks as a multistate model, Statistical Methods in Medical Research, 11 (2002), pp. 203–215.

[3] P. K. Andersen and N. Keiding, *M*ulti-state models for event history analysis, Statistical Methods in Medical Research, 11 (2002), pp. 91–115.

[4] P. K. Andersen and M. P. Perme, *I*nference for outcome probabilities in multistate models, Lifetime Data Analysis, 14 (2008), p. 405.

[5] T. Bayes, *A*n essay towards solving a problem in the doctrine of chances., Philosophical transactions of the Royal Society of London, (1763), pp. 370–418.

[6] S. M. Berry, B. P. Carlin, J. J. Lee, and P. Muller, *B*ayesian adaptive methods for clinical trials, CRC press, 2010.

[7] J. Beyersmann, A. Latouche, A. Buchholz, and M. Schumacher, *S*imulating competing risks data in survival analysis, Statistics in Medicine, 28 (2009), pp. 956–971.

[8] K. R. Broglio and D. A. Berry, *D*etecting an overall survival benefit that is derived from progression-free survival, JNCI: Journal of the National Cancer Institute, 101 (2009), pp. 1642–1649.

[9] T. Burzykowski, M. Buyse, M. J. Piccart-Gebhart, G. Sledge, J. Carmichael, H.-J. Lück, J. R. Mackey, J.-M. Nabholtz, R. Paridaens, L. Biganzoli, et al., *E*valuation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer, Journal of Clinical Oncology, 26 (2008), pp. 1987–1992.

[10] T. Burzykowski, G. Molenberghs, M. Buyse, H. Geys, and D. Renard, *V*alidation of surrogate end points in multiple randomized clinical trials with failure time end points, Journal of the Royal Statistical Society: Series C (Applied Statistics), 50 (2001), pp. 405–422.

[11] M. Buyse, *U*se of meta-analysis for the validation of surrogate endpoints and biomarkers in cancer trials, The Cancer Journal, 15 (2009), pp. 421–425.

[12] M. Buyse, T. Burzykowski, K. Carroll, S. Michiels, D. J. Sargent, L. L. Miller, G. L. Elfring, J.-P. Pignon, and P. Piedbois, *P*rogression-free survival is a surrogate for survival in advanced colorectal cancer, Journal of Clinical Oncology, 25 (2007), pp. 5218–5224.

[13] M. Buyse and G. Molenberghs, *C*riteria for the validation of surrogate endpoints in randomized experiments, Biometrics, (1998), pp. 1014–1029.

[14] M. Buyse, G. Molenberghs, T. Burzykowski, D. Renard, and H. Geys, *T*he validation of surrogate endpoints in meta-analyses of randomized experiments, Biostatistics, 1 (2000), pp. 49–67.

[15] C. L. Chiang et al., *I*ntroduction to stochastic processes in biostatistics, Wiley, 1968.

[16] D. G. Clayton, *A* model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, Biometrika, 65 (1978), pp. 141–151.

[17] D. Commenges and A. Gégout-Petit, *L*ikelihood for generally coarsened observations from multistate or counting process models, Scandinavian Journal of Statistics, 34 (2007), pp. 432–450.

[18] A. S. Conlon, J. M. Taylor, D. J. Sargent, and G. Yothers, *U*sing cure models and multiple imputation to utilize recurrence as an auxiliary variable for overall survival, Clinical Trials, 8 (2011), pp. 581–590.

[19] A. S. Conlon, J. M. Taylor, and J. Sargent, *I*mproving efficiency in clinical trials using auxiliary information: Application of a multi-state cure model, Biometrics, 71 (2015), pp. 460–468.

[20] R. Cook and J. Lawless, *S*ome comments on efficiency gains from auxiliary information for right-censored data, Journal of Statistical Planning and Inference, 96 (2001), pp. 191–202.

[21] D. Cox and H. Miller, *T*he Theory of Stochastic Processes, Methuen's monographs on applied probability and statistics, Taylor & Francis, 1977.

[22] M. J. Daniels and M. D. Hughes, *M*eta-analysis for the evaluation of potential surrogate markers, Statistics in Medicine, 16 (1997), pp. 1965–1982.

[23] D. Dejardin, E. Lesaffre, and G. Verbeke, *J*oint modeling of progression-free survival and death in advanced cancer clinical trials, Statistics in Medicine, 29 (2010), pp. 1724–1734.

[24] F. A. Eskens and J. Verweij, *C*linical studies in the development of new anticancer agents exhibiting growth inhibition in models: facing the challenge of a proper study design, Critical Reviews in Oncology/Hematology, 34 (2000), pp. 83–88.

[25] C. L. Faucett, N. Schenker, and J. M. Taylor, *S*urvival analysis using auxiliary variables via multiple imputation, with application to aids clinical trial data, Biometrics, 58 (2002), pp. 37–47.

[26] J. P. Fine, H. Jiang, and R. Chappell, *O*n semi-competing risks data, Biometrika, 88 (2001), pp. 907–919.

[27] D. M. Finkelstein and D. A. Schoenfeld, *A*nalysing survival in the presence of an auxiliary variable, Statistics in Medicine, 13 (1994), pp. 1747–1754.

[28] F. Fleischer, B. Gaschler-Markefski, and E. Bluhmki, *A* statistical model for the dependence between progression-free survival and overall survival, Statistics in Medicine, 28 (2009), pp. 2669–2686.

[29] T. R. Fleming, R. L. Prentice, M. S. Pepe, and D. Glidden, *S*urrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and aids research, Statistics in Medicine, 13 (1994), pp. 955–968.

[30] M. J. Frank, *O*n the simultaneous associativity of f(x, y) and x+y- f (x, y), Aequationes Mathematicae, 19 (1979), pp. 194–226.

[31] L. S. Freedman, B. I. Graubard, and A. Schatzkin, *S*tatistical validation of intermediate endpoints for chronic diseases, Statistics in Medicine, 11 (1992), pp. 167–178.

[32] H. Fu, Y. Wang, J. Liu, P. M. Kulkarni, and A. S. Melemed, *J*oint modeling of progression-free survival and overall survival by a bayesian normal induced copula estimation model, Statistics in Medicine, 32 (2013), pp. 240–254.

[33] M. H. Gail, R. Pfeiffer, H. C. Van Houwelingen, and R. J. Carroll, *O*n meta-analytic assessment of surrogate outcomes, Biostatistics, 1 (2000), pp. 231–246.

[34] A. E. Gelfand and A. F. Smith, *S*ampling-based approaches to calculating marginal densities, Journal of the American statistical association, 85 (1990), pp. 398–409.

[35] S. Geman and D. Geman, *S*tochastic relaxation, gibbs distributions, and the bayesian restoration of images, IEEE Transactions on pattern analysis and machine intelligence, (1984), pp. 721–741.

[36] C. Genest and R. J. Mackay, *C*opules archimédiennes et families de lois bidimensionnelles dont les marges sont données, Canadian Journal of Statistics, 14 (1986), pp. 145–159.

[37] A. Hastings, *D*isturbance, coexistence, history, and competition for space, Theoretical population biology, 18 (1980), pp. 363–373.

[38] P. Hougaard, *S*urvival models for heterogeneous populations derived from stable distributions, Biometrika, 73 (1986), pp. 387–396.

[39] K. Jachno, S. Heritier, and R. Wolfe, *A*re non-constant rates and non-proportional treatment effects accounted for in the design and analysis of randomised controlled trials? a review of current practice, BMC Medical Research Methodology, 19 (2019), p. 103.

[40] C. Jackson, *f*lexsurv: A platform for parametric survival modeling in R, Journal of Statistical Software, 70 (2016), pp. 1–33.

[41] H. Joe, *D*ependence modeling with copulas, Chapman & Hall, CRC Press, 2014.

[42] J. D. Kalbfleisch and J. F. Lawless, *T*he analysis of panel data under a Markov assumption, Journal of the American Statistical Association, 80 (1985), pp. 863–871.

[43] J. D. Kalbfleisch and R. L. Prentice, *E*stimation of the average hazard ratio, Biometrika, 68 (1981), pp. 105–112.

[44] D. G. Kleinbaum and M. Klein, *S*urvival analysis, vol. 3, Springer, 2010.

[45] M. R. Kosorok and T. R. Fleming, *U*sing surrogate failure time data to increase cost effectiveness in clinical trials, Biometrika, 80 (1993), pp. 823–833.

[46] S. W. Lagakos, *U*sing auxiliary variables for improved estimates of survival time, Biometrics, (1977), pp. 399–404.

[47] L. Lakhal, L.-P. Rivest, and D. Beaudoin, *I*PCW estimator for Kendall's tau under bivariate censoring, The International Journal of Biostatistics, 5 (2009).

[48] Y. Li and Q. Zhang, *A* Weibull multi-state model for the dependence of progression-free survival and overall survival, Statistics in Medicine, 34 (2015), pp. 2497–2513.

[49] R. S. Lin, J. Lin, S. Roychoudhury, K. M. Anderson, T. Hu, B. Huang, L. F. Leon, J. J. Liao, R. Liu, X. Luo, et al., *A*lternative analysis methods for time to event endpoints under non-proportional hazards: A comparative analysis, arXiv preprint arXiv:1909.09467, (2019).

[50] Y. Luke, *T*he special functions and their approximations, Vol. i. 1969.

[51] M. Mandel, *S*imulation-based confidence intervals for functions with complicated derivatives, The American Statistician, 67 (2013), pp. 76–81.

[52] T. Martinussen and C. B. Pipper, *E*stimation of odds of concordance based on the Aalen additive model, Lifetime Data Analysis, 19 (2013), pp. 100–116.

[53] L. Meira-Machado, J. de Uña-Álvarez, C. Cadarso-Suárez, and P. K. Andersen, *M*ulti-state models for the analysis of time-to-event data, Statistical Methods in Medical Research, 18 (2009), pp. 195–222.

[54] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *E*quation of state calculations by fast computing machines, The journal of chemical physics, 21 (1953), pp. 1087–1092.

[55] C. G. Moertel, T. R. Fleming, J. S. Macdonald, D. G. Haller, J. A. Laurie, C. M. Tangen, J. S. Ungerleider, W. A. Emerson, D. C. Tormey, J. H. Glick, et al., *F*luorouracil plus levamisole as effective adjuvant therapy after resection of stage iii colon carcinoma: a final report, Annals of Internal Medicine, 122 (1995), pp. 321–326.

[56] H.-H. Müller and H. Schäfer, *A*daptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches, Biometrics, 57 (2001), pp. 886–891.

[57] D. Oakes, *A* concordance test for independence in the presence of censoring, Biometrics, 38 (1982), pp. 451–455.

[58] D. Oakes, *B*ivariate survival models induced by frailties, Journal of the American Statistical Association, 84 (1989), pp. 487–493.

[59] G. Omura, M. Buyse, S. Marsoni, K. Bertelsen, P. Conte, A. Jakobsen, and J. Vermorken, *C*yclophosphamide plus cisplatin versus cyclophosphomide, doxorubicin, and cisplatin chemotherapy of ovarian carcinoma: a meta-analysis, Journal of Clinical Oncology, 9 (1991), pp. 1668–1674.

[60] X. Paoletti, K. Oba, Y.-J. Bang, H. Bleiberg, N. Boku, O. Bouché, P. Catalano, N. Fuse, S. Michiels, M. Moehler, et al., *P*rogression-free survival as a surrogate for overall survival in advanced/recurrent gastric cancer trials: a meta-analysis, Journal of the National Cancer Institute, 105 (2013), pp. 1667–1670.

[61] R. L. Prentice, *S*urrogate endpoints in clinical trials: definition and operational criteria, Statistics in Medicine, 8 (1989), pp. 431–440.

[62] H. Putter, M. Fiocco, and R. B. Geskus, *T*utorial in biostatistics: competing risks and multi-state models, Statistics in Medicine, 26 (2007), pp. 2389–2430.

[63] G. Rauch, W. Brannath, M. Brückner, and M. Kieser, *T*he average hazard ratio–a good effect measure for time-to-event endpoints when the proportional hazard assumption is violated?, Methods of Information in Medicine, 57 (2018), pp. 089–100.

[64] P. Royston and M. K. Parmar, *F*lexible 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 (2002), pp. 2175–2197.

[65] P. Royston and M. K. Parmar, *R*estricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome, BMC medical research methodology, 13 (2013), p. 152.

[66] P. Royston and M. K. Parmar, *A*n approach to trial design and analysis in the era of non-proportional hazards of the treatment effect, Trials, 15 (2014), p. 314.

[67] E. Saad, A. Katz, P. Hoff, and M. Buyse, *P*rogression-free survival as surrogate and as true end point: insights from the breast and colorectal cancer literature, Annals of Oncology, 21 (2009), pp. 7–12.

[68] M. Schemper, S. Wakounig, and G. Heinze, *T*he estimation of average hazard ratios by weighted Cox regression, Statistics in Medicine, 28 (2009), pp. 2473–2489.

[69] J. H. Shih and T. A. Louis, *I*nferences on the association parameter in copula models for bivariate survival data, Biometrics, 51 (1995), pp. 1384–1399.

[70] Y. Shu, J. P. Klein, and M.-J. Zhang, *A*symptotic theory for the Cox semi-Markov illness-death model, Lifetime Data Analysis, 13 (2007), pp. 91–117.

[71] M. Sklar, *F*onctions de répartition à n dimensions et leurs marges, Université Paris 8, 1959.

[72] A. Stone, C. Wheeler, and A. Barge, *I*mproving the design of phase ii trials of cytostatic anticancer agents, Contemporary Clinical Trials, 28 (2007), pp. 138–145.

[73] C. A. Struthers and J. D. Kalbfleisch, *M*isspecified proportional hazard models, Biometrika, 73 (1986), pp. 363–369.

[74] J. Xu, J. D. Kalbfleisch, and B. Tai, *S*tatistical analysis of illness–death processes and semicompeting risks data, Biometrics, 66 (2010), pp. 716–725.

[75] L. Zeng, R. Cook, L. Wen, and A. Boruka, *B*ias in progression-free survival analysis due to intermittent assessment of progression, Statistics in Medicine, 34 (2015), pp. 3181–3193.

Supplementary Material

APPENDIX A

Chapter 3: Quantifying the association between progression-free survival and overall survival in oncology trials using Kendall's τ

A.1 Derivation of the formula for model-based Kendall's τ

Under an assumption that patients are independent and identically distributed and that S and T are continuous random variables, from (3.11),

$$\tau_{\text{mod}} = 4P(S_1 > S_2, T_1 > T_2) - 1,$$

where (S_1, T_1) and (S_2, T_2) are two independent pairs of realisations of (S, T).

The probability $P(S_1 > S_2, T_1 > T_2)$ can be expanded into three cases defined by which of patient 1 or patient 2 dies before progression. Specifically we may write

$$\begin{aligned} P(S_1 > S_2, T_1 > T_2) =& P(S_2 = T_2 < S_1) \\ &+ P(S_1 > S_2, T_1 > T_2, T_2 > S_2, T_1 > S_1) \\ &+ P(S_1 = T_1, S_2 < S_1, S_2 < T_2 < T_1). \end{aligned} \tag{A.1}$$

Note that S_j can also be interpreted as the sojourn time in state 0 and that

$$T_j = \begin{cases} S_j & \text{if } \Delta_j = 0 \\ S_j + V_j & \text{if } \Delta_j = 1, \end{cases}$$

where Δ_j is an indicator of whether progression occurs before death and V_j is the sojourn time in the progression state. Let $f_{S1}(t) = \pi_{01}(t) \exp(-\Pi_0(t))$ and $f_{S2}(t) = \pi_{02}(t) \exp(-\Pi_0(t))$ be the cause-specific densities for progression and death before progression, respectively. Also, let $S_S(t) = \exp(-\Pi_0(t))$ be the survivor function of the sojourn distribution in state 0 and $S_{V|S}(s; u) = \exp(-\Pi_{12}(s; u))$ and $f_{V|S}(s; u) = \pi_{12}(s; u) \exp(-\Pi_{12}(s; u))$ be the survivor function and density of the conditional sojourn distribution in state 1 given a sojourn of time u in state

0. Then

$$\begin{aligned} P(S_2 = T_2 < S_1) &= \int_0^\infty P(S_2 = s, \Delta_2 = 0) P(S_1 > s) ds \\ &= \int_0^\infty f_{S2}(s) S_S(s) ds \\ &= \int_0^\infty \pi_{02}(s) \exp\{-2\Pi_0(s)\} ds, \end{aligned}$$

$$\begin{aligned} & P(S_1 > S_2, T_1 > T_2, T_2 > S_2, T_1 > S_1) \\ = & \int_0^\infty \int_0^{s_1} P(S_1 = s_1, \Delta_1 = 1) P(S_2 = s_2, \Delta_2 = 1) \\ & \qquad \times P(T_1 > T_2 | S_1 = s_1, S_2 = s_2, \Delta_1 = \Delta_2 = 1) ds_2 ds_1 \\ = & \int_0^\infty \int_0^{s_1} \int_0^\infty P(S_2 = s_2, \Delta_2 = 1) P(S_1 = s_1, \Delta_1 = 1) P(V_1 = s_3 | S_1 = s_1) \\ & \qquad \times P(V_2 < s_1 + s_3 - s_2 | S_2 = s_2) ds_3 ds_2 ds_1 \\ = & \int_0^\infty \int_0^{s_1} \int_0^\infty f_{S1}(s_1) f_{S1}(s_2) f_{V|S}(s_3; s_1) (1 - S_{V|S}(s_1 + s_3 - s_2; s_2)) ds_3 ds_2 ds_1 \\ = & \int_0^\infty \int_0^{s_1} \int_0^\infty \pi_{01}(s_1) \pi_{01}(s_2) \exp\{-\Pi_0(s_1) - \Pi_0(s_2)\} \pi_{12}(s_3; s_1) \\ & \qquad \times \exp\{-\Pi_{12}(s_3; s_1)\} [1 - \exp\{-\Pi_{12}(s_1 + s_3 - s_2; s_2)\}] ds_3 ds_2 ds_1 \end{aligned}$$

and

$$\begin{aligned} & P(S_1 = T_1, S_2 < S_1, S_2 < T_2 < T_1) \\ & = \int_0^\infty P(S_1 = s_1, \Delta_1 = 0) P(S_2 < T_2 < s_1) ds_1 \\ & = \int_0^\infty \int_0^{s_1} P(S_1 = s_1, \Delta_1 = 0) P(S_2 = s_2, \Delta_2 = 1) \\ & \qquad \times P(V_2 < s_1 - s_2 | S_2 = s_2) ds_2 ds_1 \\ & = \int_0^\infty \int_0^{s_1} f_{S2}(s_1) f_{S1}(s_2) (1 - S_{V|S}(s_1 - s_2)) ds_2 ds_1 \\ & = \int_0^\infty \int_0^{s_1} \pi_{02}(s_1) \pi_{01}(s_2) \exp\{-\Pi_0(s_1) - \Pi_0(s_2)\} \\ & \qquad \times (1 - \exp\{-\Pi_{12}(s_1 - s_2; s_2)\}) ds_2 ds_1. \end{aligned}$$

Substituting these terms into (A.1) and then into (3.11) produces the expression given in (3.10).

A.2 Additional Simulation Results

Tables A.1, A.2 and A.3 present the bias and standard deviation (SD) of the estimators of Kendall's τ in the realistic, unrealistic and non-Markov scenarios, respectively.

Table A.1: Scenario A (realistic scenario): simulation results including the Kendall's τ, Bias, SD and MSE for each model.

			Censoring cases			
	Model	Summary	1	2	3	4
1	model-based method	Bias	0.0003	0.0003	0.0004	0.0003
		SD	0.0076	0.0083	0.0093	0.0097
		MSE	0.00006	0.00007	0.00009	0.00010
2	IPCW estimator	Bias	-0.0013	-0.0011	-0.0030	-0.0059
		SD	0.0092	0.0087	0.0101	0.0095
		MSE	0.00009	0.00008	0.00011	0.00012
3	one-stage fully parametric model (Clayton)	Bias	0.0024	0.0023	-0.0023	-0.0068
		SD	0.0083	0.0084	0.0092	0.0098
		MSE	0.00008	0.00008	0.00009	0.00014
4	one-stage fully parametric model (Hougaard)	Bias	-0.1197	-0.1158	-0.1663	-0.1617
		SD	0.0140	0.0122	0.0158	0.0126
		MSE	0.01452	0.01356	0.02792	0.02631
5	one-stage fully parametric model (Frank)	Bias	0.0026	-0.0052	-0.0324	-0.0563
		SD	0.0146	0.0075	0.0104	0.0099
		MSE	0.00022	0.00008	0.00116	0.00327
6	two-stage semi-parametric model (Clayton)	Bias	0.0015	-0.0004	-0.0082	-0.0110
		SD	0.0089	0.0082	0.0107	0.0098
		MSE	0.00008	0.00007	0.00018	0.00022
7	two-stage semi-parametric model (Hougaard)	Bias	-0.1408	-0.1411	-0.1956	-0.1932
		SD	0.0136	0.0130	0.0163	0.0135
		MSE	0.02001	0.02007	0.03854	0.03750
8	two-stage semi-parametric model (Frank)	Bias	-0.0301	-0.0330	-0.0551	-0.0696
		SD	0.0108	0.0101	0.0124	0.0113
		MSE	0.00102	0.00119	0.00319	0.00498

Setting: sample size in each simulation is 1000, 1000 simulations.
True Kendall's τ: 0.8348.
Censoring case 1: 20% exponential distributed censoring
Censoring case 2: 20% uniform distributed censoring
Censoring case 3:45% exponential distributed censoring
Censoringcase 4: 45% uniform distributed censoring

Table A.2: Scenario B (unrealistic scenario): simulation results including the Kendall's τ, Bias, SD and MSE for each model.

			Censoring cases			
	Model	Summary	1	2	3	4
1	model-based method	Bias	0.0006	0.0004	0.0006	0.0009
		SD	0.0166	0.0166	0.0174	0.0168
		MSE	0.0003	0.0003	0.0003	0.0003
2	IPCW estimator	Bias	0.0150	0.0116	0.0526	0.0483
		SD	0.0222	0.0222	0.0253	0.0244
		MSE	0.0007	0.0006	0.0034	0.0029
3	one-stage fully parametric model (Clayton)	Bias	-0.0457	-0.0486	-0.0073	-0.0018
		SD	0.0203	0.0208	0.0373	0.0207
		MSE	0.0025	0.0028	0.0014	0.0004
4	one-stage fully parametric model (Hougaard)	Bias	-0.0417	-0.0259	-0.0226	-0.0259
		SD	0.0430	0.0304	0.0256	0.0242
		MSE	0.0036	0.0025	0.0012	0.0013
5	one-stage fully parametric model (Frank)	Bias	-0.0296	-0.0352	-0.0084	-0.0952
		SD	0.0472	0.0205	0.0231	0.0458
		MSE	0.0031	0.0017	0.0006	0.0112
6	two-stage semi-parametric model (Clayton)	Bias	-0.0509	-0.0519	-0.0090	-0.0026
		SD	0.0174	0.0181	0.0232	0.0250
		MSE	0.0029	0.0030	0.0006	0.0006
7	two-stage semi-parametric model (Hougaard)	Bias	-0.0348	-0.0362	-0.0220	-0.0264
		SD	0.0183	0.0193	0.0194	0.0196
		MSE	0.0015	0.0017	0.0009	0.0011
8	two-stage semi-parametric model (Frank)	Bias	-0.0337	-0.0360	-0.0094	-0.0135
		SD	0.0209	0.0196	0.0218	0.0209
		MSE	0.0016	0.0017	0.0006	0.0006

Setting: sample size in each simulation is 1000, 1000.
True Kendall's τ: 0.1201.
Censoring case 1: 20% exponential distributed censoring
Censoring case 2: 20% uniform distributed censoring
Censoring case 3: 45% exponential distributed censoring
Censoring case 4: 45%uniform distributed censoring

Table A.3: Scenario C (non-Markov scenario): simulation results including the Kendall's τ, Bias, SD and MSE for each model.

			Censoring cases			
	Model	Summary	1	2	3	4
1	model-based method	Bias	-0.0183	-0.0169	-0.0118	-0.0061
		SD	0.0093	0.0093	0.0113	0.0106
		MSE	0.00042	0.00037	0.00027	0.00015
2	IPCW estimator	Bias	0.0011	0.0003	0.0047	0.0041
		SD	0.0089	0.0087	0.0099	0.0096
		MSE	0.00008	0.00008	0.00012	0.00011
3	one-stage fully parametric model (Clayton)	Bias	-0.0053	-0.0043	-0.0009	-0.0006
		SD	0.0087	0.0090	0.0093	0.0099
		MSE	0.00010	0.00010	0.00009	0.00010
4	one-stage fully parametric model (Hougaard)	Bias	-0.1312	-0.1246	-0.1598	-0.1687
		SD	0.0106	0.0106	0.0129	0.0121
		MSE	0.01731	0.01564	0.02572	0.02859
5	one-stage fully parametric model (Frank)	Bias	-0.0007	-0.0079	-0.0245	-0.0486
		SD	0.0086	0.0076	0.0096	0.0096
		MSE	0.00008	0.00012	0.00069	0.00245
6	two-stage semi-parametric model (Clayton)	Bias	-0.0094	-0.0085	-0.0089	-0.0035
		SD	0.0101	0.0090	0.0115	0.0093
		MSE	0.00019	0.00015	0.00021	0.00010
7	two-stage semi-parametric model (Hougaard)	Bias	-0.1279	-0.1313	-0.1748	-0.1794
		SD	0.0133	0.0127	0.0151	0.0135
		MSE	0.01653	0.01739	0.03078	0.03237
8	two-stage semi-parametric model (Frank)	Bias	-0.0266	-0.0295	-0.0435	-0.0584
		SD	0.0099	0.0088	0.0112	0.0101
		MSE	0.00080	0.00095	0.00202	0.00352

Setting: sample size in each simulation=1000, 1000 simulations.
True Kendall's τ: 0.8155.
Censoring case 1: 20% exponential distributed censoring
Censoring case 2: 20% uniform distributed censoring
Censoring case 3: 45% exponential distributed censoring
Censoring case 4: 45% uniform distributed censoring

Box plots of the estimators under exponentially distributed censoring are shown.

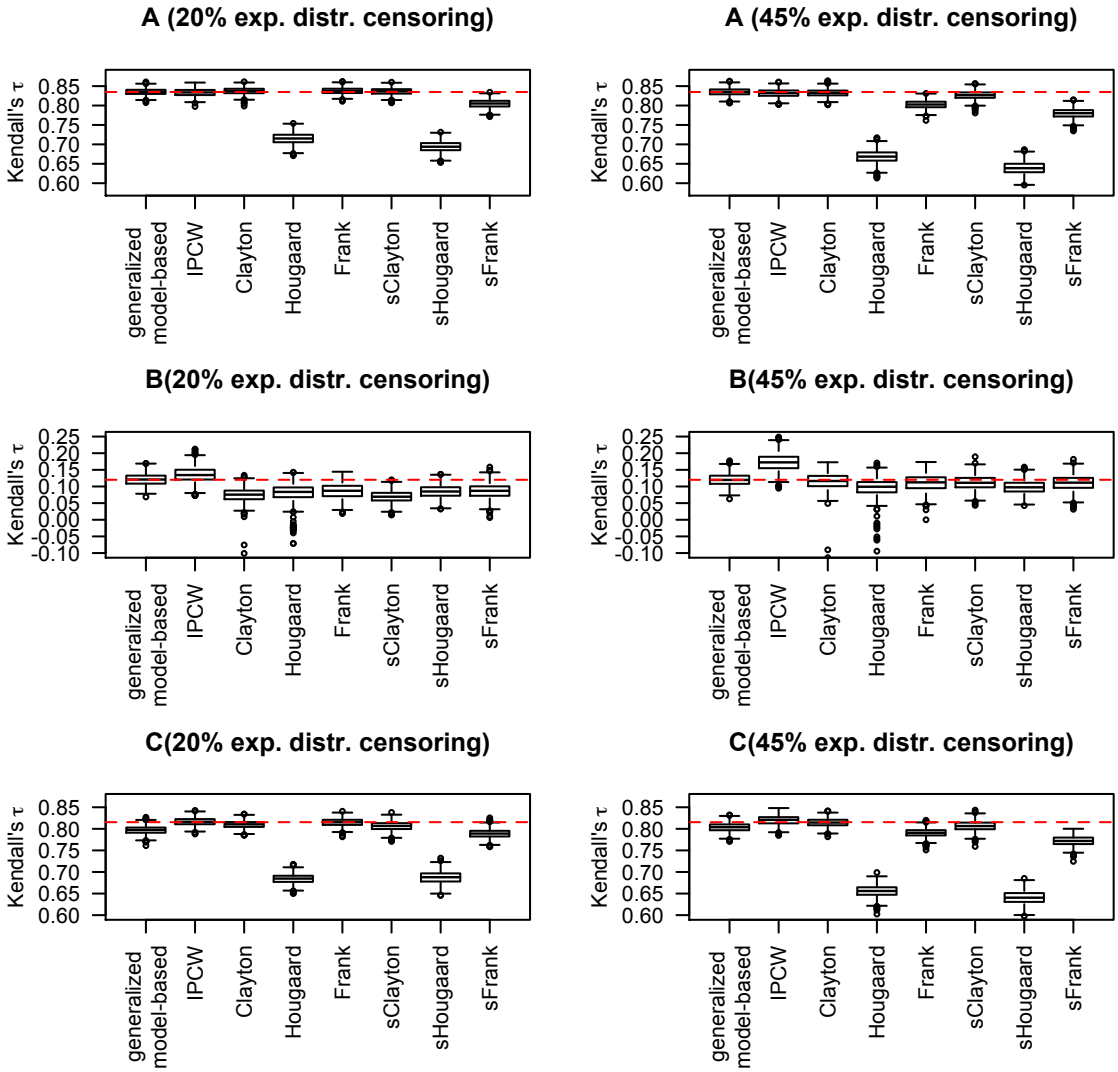

Figure A.1: The three-state illness death model under exponential censoring.
A: normal scenario, where the parameters are used from an external dataset from a trial of treatments for colon cancer
B: unrealistic scenario
C: non-Markov scenario
sClayton:two stage semi parametric Clayton model; sHougaard: two stage semi parametric Hougaard model; sFrank: two stage semi parametric Frank model

Contour plots of the density function of the non-Markov scenario

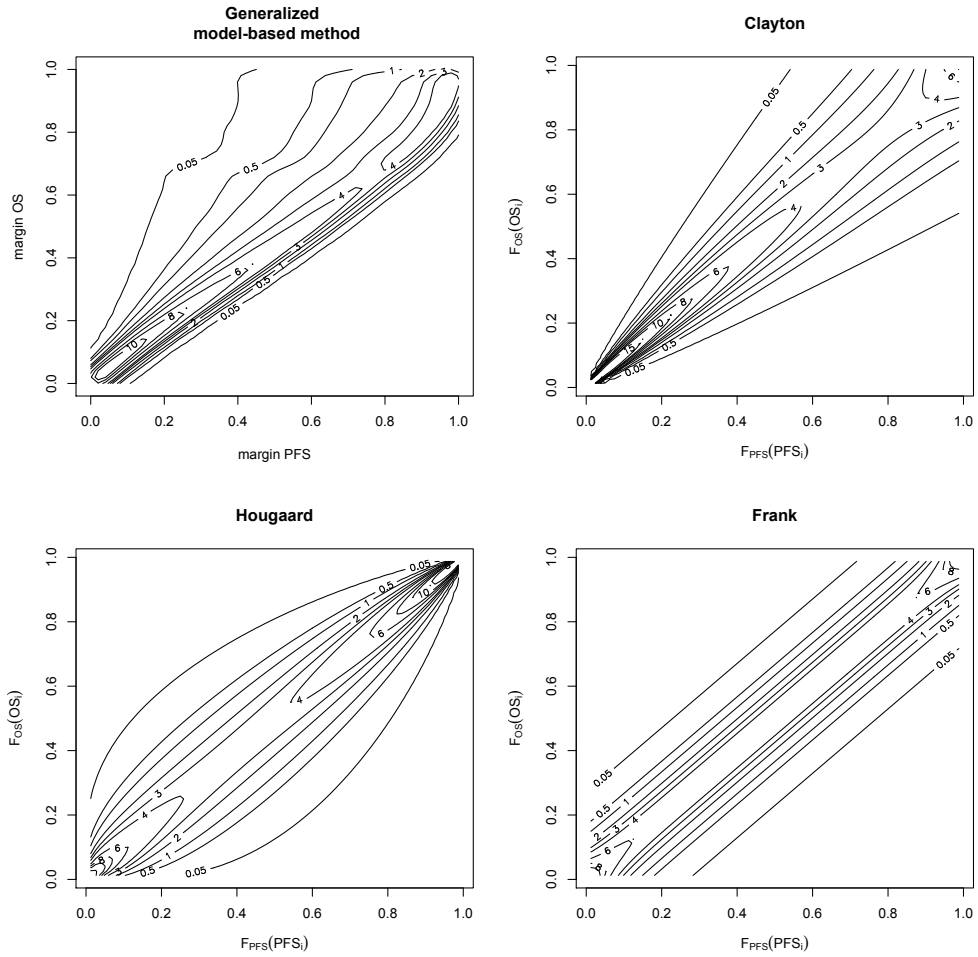

Figure A.2: Simulation scenario C (non-Markov case): contour plots for the bivariate density function based on the model-based method and the survivor joint Clayton's copula model, Hougaard's copula model and Frank's copula model. Kendall's τ is 0.815.

Proportion of the lowest AIC of each full parametric and semi-parametric copula model

Table A.4: The proportion of each full parametric copula regarding the lowest AIC based on 1000 simulation for each scenario (20% uniform distributed censoring).

	AIC		
Scenario	parametric Clayton	parametric Hougaard	parametric Frank
A (20%)	91.6	0.0	8.4
B (20%)	99.8	0.0	0.2
C (20%)	79.9	0.0	20.1

Table A.5: The proportion of each semi-parametric copula model regarding the lowest AIC based on 1000 simulation for each scenario (20 % uniform distributed censoring).

	AIC		
Scenario	semi-parametric Clayton	semi-parametric Hougaard	semi-parametric Frank
A (20%)	99.7	0.0	0.3
B (20%)	100.0	0.0	0.0
C (20%)	98.1	0.0	1.9

A.3 Additional Application Results

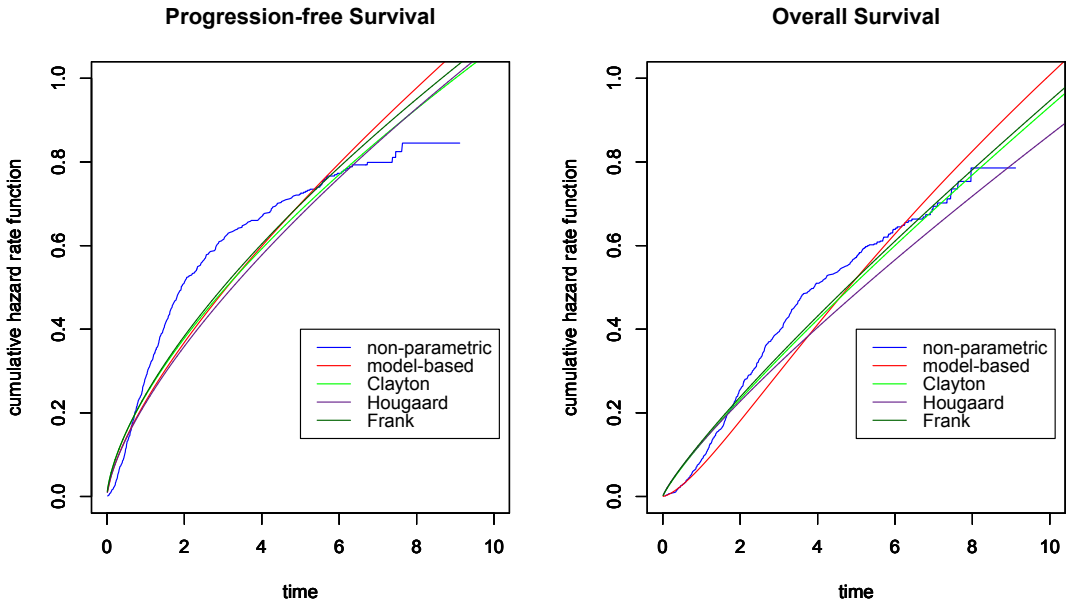

Figure A.3: Non-parametric cumulative hazard rate functions for both PFS and OS and the cumulative hazard functions for both PFS and OS based on the generalized model-based methods and the three copula approaches assuming Weibull hazards.

Table A.6: Estimate of the RP copula models and the se of each parameter estimation

	RP Clayton model		RP Hougaard model		RP Frank model	
Model parameter	estimates	se	estimates	se	estimates	se
γ_{PFS_0}	0.1675	0.2778	-1.5865	0.2215	-0.0765	0.2778
γ_{PFS_1}	2.0621	0.1998	0.8622	0.1348	1.8935	0.1998
γ_{PFS_2}	0.0405	0.0321	-0.1125	0.0303	0.0376	0.0321
γ_{PFS_3}	0.0379	0.0335	0.1647	0.0348	0.0308	0.0335
γ_{OS_0}	-1.6132	0.2064	-2.4361	0.1448	-1.6625	0.2064
γ_{OS_1}	2.2008	0.2862	0.7460	0.1261	2.0947	0.2862
γ_{OS_2}	-0.0202	0.0859	-0.2404	0.0615	0.0120	0.0859
γ_{OS_3}	0.1797	0.1018	0.3403	0.0824	0.1209	0.1018

A.4 Derivation of the log-likelihood function for the model estimation based on the generalized multi-state model

In Section 3.2.5 we consider an illness-death model characterised by transition intensities (based on Weibull distribution) given by

$$\pi_{01}(t) = \alpha_{01}\Big(\frac{1}{\lambda_{01}}\Big)^{\alpha_{01}} t^{\alpha_{01}-1},$$
$$\pi_{02}(t) = \alpha_{02}\Big(\frac{1}{\lambda_{02}}\Big)^{\alpha_{02}} t^{\alpha_{02}-1},$$
$$\pi_{12}(s) = \alpha_{12}\Big(\frac{1}{\lambda_{12}}\Big)^{\alpha_{12}} s^{\alpha_{12}-1},$$

where t and s refer to time since randomization and to time since progression, respectively. The estimation of the parameter vector $\theta = (\alpha_{01}, \alpha_{02}, \alpha_{12}, \lambda_{01}, \lambda_{02}, \lambda_{12})$ for the illness-death model is estimated via maximum-likelihood.

In order to estimate the parameters, a log-likelihood function has to be constructed. Using the model estimation as in Li and Zhang,[48] we first describe the individual survival experiences by four possible different cases: Patients can (1) progress and then censor, (2) progress and then die, (3) die before progression or (4) get censored without experiencing progression and death. In terms of failure times, every individual $i \in \{1, \ldots, n\}, t_{i1}$ represents the time from the initial state to the state of progression or death, t_{i2} exists given progression and describes the time from progression to death. The likelihood of the estimates can be derived from the single likelihood $L_i^{(k)}(\theta), (k = 1, 2, 3, 4)$ for each experiences of patient:

if case 1 occurs:

$$L_i^{(1)}(\theta) = f_1(t_{i1})S_2(t_{i1})S_3(t_{i2}),$$

if case 2 occurs:

$$L_i^{(2)}(\theta) = f_1(t_{i1})S_2(t_{i1})f_3(t_{i2}),$$

if case 3 occurs:

$$L_i^{(3)}(\theta) = S_1(t_{i1}) f_2(t_{i1}),$$

if case 4 occurs:

$$L_i^{(3)}(\theta) = S_1(t_{i1}) S_2(t_{i1}),$$

where

$f_1(\cdot)$ and $S_1(\cdot)$: density and survival function for time-to-progression,

$f_2(\cdot)$ and $S_2(\cdot)$: density and survival function for time-to-death without progression,

$f_3(\cdot)$ and $S_3(\cdot)$: density and survival function for time-to-death given progression.

Then, the overall log-likelihood for all subjects is given by:

$$\begin{aligned} log(L(\theta)) = \sum_{i=1}^{n} & (d_1)(1-d_2)(1-d_3) L_i^{(1)}(\theta) \\ & + (d_1)(1-d_2)(d_3) L_i^{(2)}(\theta) \\ & + (1-d_1)(d_2) L_i^{(3)}(\theta) \\ & + (1-d_1)(1-d_2) L_i^{(4)}(\theta), \end{aligned}$$

where d_1 describes the censoring indicator for progression, d_2 is the indicator for death without progression, and d_3 is the indicator for death given progression. The censoring indicator is zero for censoring, otherwise 1.

APPENDIX B

Joint modelling of PFS and OS in a Bayesian framework

B.1 Estimation of weighted Kendall's tau for 2-arm trial

Monte-Carlo methods provide a particularly convenient way of evaluating the model based Kendall's τ. Here we denote PFS by S and OS by T. We can use the fact that for a model where S and T are continuous,

$$\begin{aligned}
\tau = \frac{1}{4}\{ \quad & 2P(S_{1_c} > S_{2_c}, T_{1_c} > T_{2_c}) - 2P(S_{1_c} < S_{2_c}, T_{1_c} > T_{2_c}) \\
& +2P(S_{1_t} > S_{2_t}, T_{1_t} > T_{2_t}) - 2P(S_{1_t} < S_{2_t}, T_{1_t} > T_{2_t}) \\
& +2P(S_{1_c} > S_{2_t}, T_{1_c} > T_{2_t}) - 2P(S_{1_c} < S_{2_t}, T_{1_c} > T_{2_t}) \\
& +2P(S_{1_t} > S_{2_c}, T_{1_t} > T_{2_c}) - 2P(S_{1_t} < S_{2_c}, T_{1_t} > T_{2_c})\} \\
= \frac{1}{4}\{ \quad & 2P(S_{1_c} > S_{2_c}, T_{1_c} > T_{2_c}) + \{1 - 2P(S_{1_c} > S_{2_c}, T_{1_c} > T_{2_c})\} \\
& +2P(S_{1_t} > S_{2_t}, T_{1_t} > T_{2_t}) + \{1 - 2P(S_{1_t} > S_{2_t}, T_{1_t} > T_{2_t})\} \\
& +2P(S_{1_c} > S_{2_t}, T_{1_c} > T_{2_t}) + \{1 - 2P(S_{1_c} > S_{2_t}, T_{1_c} > T_{2_t})\} \\
& +2P(S_{1_t} > S_{2_c}, T_{1_t} > T_{2_c}) + \{1 - 2P(S_{1_t} > S_{2_c}, T_{1_t} > T_{2_c})\}\} \\
= 4\{ & P(S_{1_c} > S_{2_c}, T_{1_c} > T_{2_c}) + (P(S_{1_t} > S_{2_t}, T_{1_t} > T_{2_t}) \\
& + P(S_{1_c} > S_{2_t}, T_{1_c} > T_{2_t}) - P(S_{1_t} > S_{2_c}, T_{1_t} > T_{2_c})\} - 1. \qquad \text{(B.1)}
\end{aligned}$$

It is therefore only necessary to evaluate $P(S_1 > S_2, T_1 > T_2)$ which can be achieved by simulating $2M$ pairs of (S_i, T_i) and then taking

$$\hat{P}(S_1 > S_2, T_1 > T_2) = M^{-1} \sum_{i=1}^{M} I(S_i > S_{i+M}, T_i > T_{i+M}). \qquad \text{(B.2)}$$

B.1.1 Estimation of standard error of the weighted Kendall's tau based on the multi state model

Simulation for general illness-death models can be achieved using the methods in Beyersmann *et al*.[7] The Monte-Carlo standard error associated with the approximation is at most $1/2\sqrt{M}$. Typically, $M = 1 \times 10^6$ or 1×10^7 samples can be generated using very little computation time meaning the Monte-Carlo standard error is negligible. A point estimate for 3.12 can be obtained by simulating $2M$ independent pairs of PFS and OS times from the illness-death model with parameter estimates $\hat{\boldsymbol{\theta}} := (\hat{\lambda_{01}}, \hat{\lambda_{02}}, \hat{\lambda_{12}}, \hat{\alpha_{01}}, \hat{\alpha_{02}}, \hat{\alpha_{12}})$. The posterior sample for the parameters of the parametric illness-death model are estimated via maximum likelihood within the Bayesian framework and then the mean of the parameters of the posterior samples. A simulation based approach may also be used to obtain confidence intervals for τ_{mod} using a variant of the *simulation delta method*.[51] This involves firstly generating B samples

$$\boldsymbol{\theta}_1^* \dots \boldsymbol{\theta}_B^* \sim N(\hat{\boldsymbol{\theta}}, I(\hat{\boldsymbol{\theta}})^{-1}),$$

where $I(\hat{\boldsymbol{\theta}})$ is the observed Fisher information of the log-likelihood. For each of the B samples, a pair of (S, T) from the illness-death process with parameters $\boldsymbol{\theta}_b^*$ are simulated $2M$ times. The next step is to estimate τ_{mod} denoted by τ_{mod}^{*b} for every $b \in [1, B]$ using (3.12). Confidence intervals can then be constructed based either upon the sample standard deviation or sample quantiles of $\tau_{\text{mod}}^{*1}, \dots, \tau_{\text{mod}}^{*B}$. A non-parametric bootstrap variant of this algorithm is also possible where B bootstrap samples are generated by repeatedly resampling from the original data and the maximum likelihood estimates are recomputed to generate each $\boldsymbol{\theta}_b^*$.

B.1.2 Additional Simulation Results

Table B.1: Simulation results for all 4 scenarios, low censoring

	Model	Summary	Scenario 1	2	3	4
		max. follow-up time	10	10	10	10
		true AHR	0.7542	0.8087	0.7419	0.4323
1	model-based method	AHR_{OS}	0.7389	0.7960	0.7365	0.4233
		SE_{AHR}	0.0528	0.0568	0.0535	0.0313
		$BIAS_{AHR}$	-0.0153	-0.0127	-0.0054	-0.009
2	Clayton-copula based methods	AHR	0.8373	0.9218	0.7666	0.4475
	(non-proportional hazards assumption)	SE_{AHR}	0.0571	0.0629	0.0537	0.0341
		$BIAS_{AHR}$	0.0831	0.1131	0.0247	0.0152
3	Clayton-copula based methods	AHR	0.8392	0.9334	0.7880	0.4080
	(proportional hazards assumption)	SE_{AHR}	0.0500	0.0559	0.0486	0.0274
		$BIAS_{AHR}$	0.085	0.1247	0.0461	-0.0243
4	Gaussian-copula based approach	AHR	0.7540	0.8110	0.7431	0.4371
		SE_{AHR}	0.0346	0.0369	0.0341	0.0239
		$BIAS_{AHR}$	-0.0002	0.0023	0.0012	0.0048
5	Cox Model	HR_{COX}	0.7404	0.7971	0.7508	0.4059
		$SE_{HR_{COX}}$	0.0526	0.0566	0.0536	0.0300
		$BIAS_{HR_{COX}}$	-0.0138	-0.0116	0.0089	-0.0264

Table B.2: Weighted Kendall's tau for the model-based method and the two versions of the Clayton copula method, high censoring

	Model	Summary	Scenario 1	2	3	4
		max. follow-up time	7	5	6	7
		true Kendall's τ	0.5538	0.5447	0.5101	0.4228
1	model-based method	Kendall's τ	0.5514	0.5425	0.5090	0.4197
		SE_τ	0.0166	0.0170	0.0173	0.0178
		$BIAS_\tau$	-0.0024	-0.0052	-0.0057	-0.0011
2	Clayton-copula based methods	Kendall's τ	0.5600	0.5573	0.5270	0.4251
	(non-proportional hazards assumption)	SE_τ	0.0171	0.0172	0.0179	0.0199
		$BIAS_\tau$	0.0062	0.0126	0.0169	0.0023
3	Clayton-copula based methods	Kendall's τ	0.5596	0.5568	0.5267	0.4252
	(proportional hazards assumption)	SE_τ	0.0170	0.0172	0.0179	0.0198
		$BIAS_\tau$	0.0058	0.0121	0.0166	0.0024

Table B.3: Weighted Kendall's tau for the model-based method and the two versions of the Clayton copula method, low censoring

	Model	Summary	Scenario 1	2	3	4
		max. follow-up time	10	10	10	10
		true Kendall's τ	0.5533	0.5457	0.5102	0.4224
1	model-based method	Kendall's τ	0.5515	0.5434	0.5092	0.4213
		SE_τ	0.0144	0.0145	0.0150	0.0152
		$BIAS_\tau$	-0.0018	-0.0023	-0.001	-0.0011
2	Clayton-copula based methods (non-proportional hazards assumption)	Kendall's τ	0.5359	0.5320	0.4802	0.3886
		SE_τ	0.0160	0.0161	0.0173	0.0183
		$BIAS_\tau$	-0.0174	-0.0137	-0.03	-0.0338
3	Clayton-copula based methods (proportional hazards assumption)	Kendall's τ	0.5352	0.5311	0.4799	0.3882
		SE_τ	0.0160	0.0161	0.0173	0.0182
		$BIAS_\tau$	-0.0181	-0.0146	-0.0303	-0.0342

APPENDIX C

Gaining efficiency in oncology trials using multi-state model based method for modelling the survival times

C.1 Additional Simulation Results

Illustration of the hazard function and survival function and cumulative hazard function of OS for scenarios 2-8

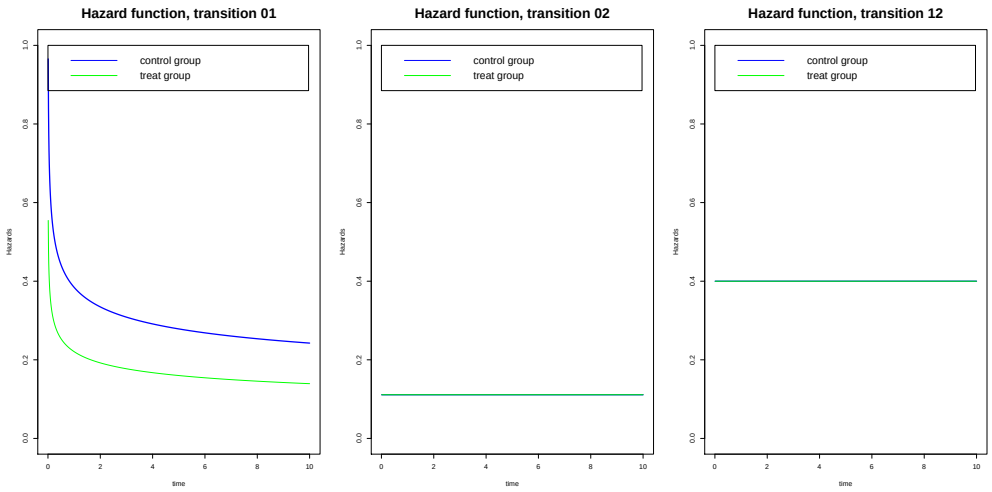

Figure C.1: Hazard functions for simulation scenario 2

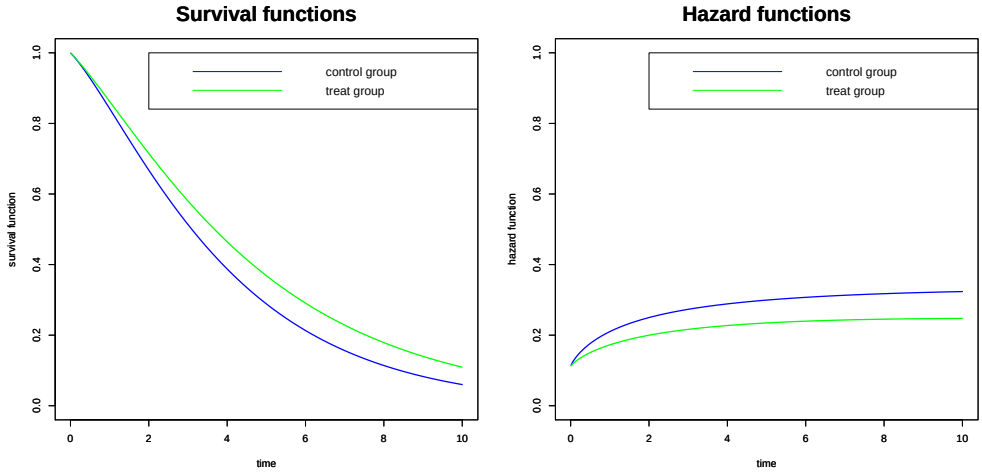

Figure C.2: Survival function of OS and cumulative hazard function for simulation scenario 2

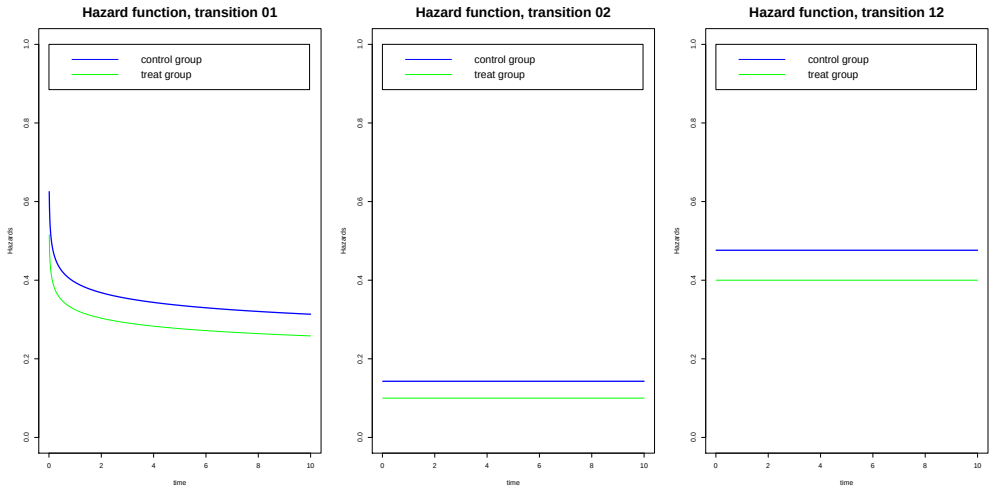

Figure C.3: Hazard functions for simulation scenario 3

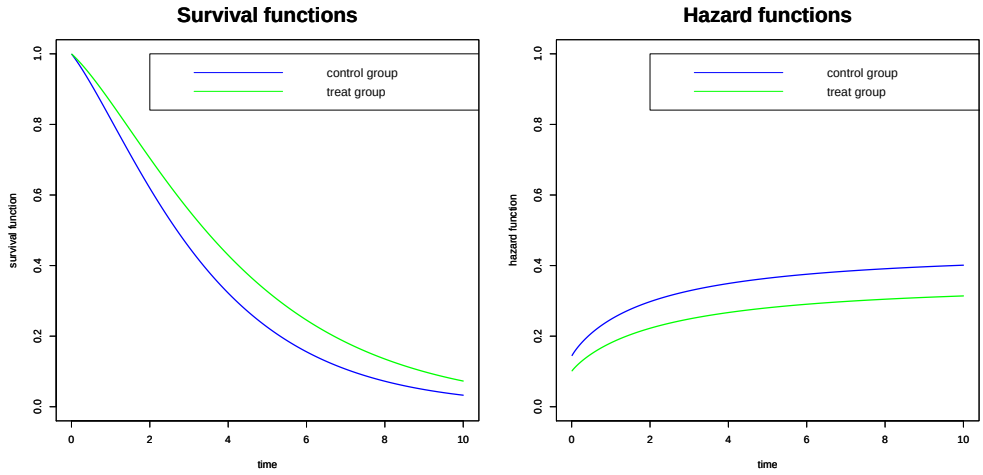

Figure C.4: Survival function of OS and cumulative hazard function for simulation scenario 3

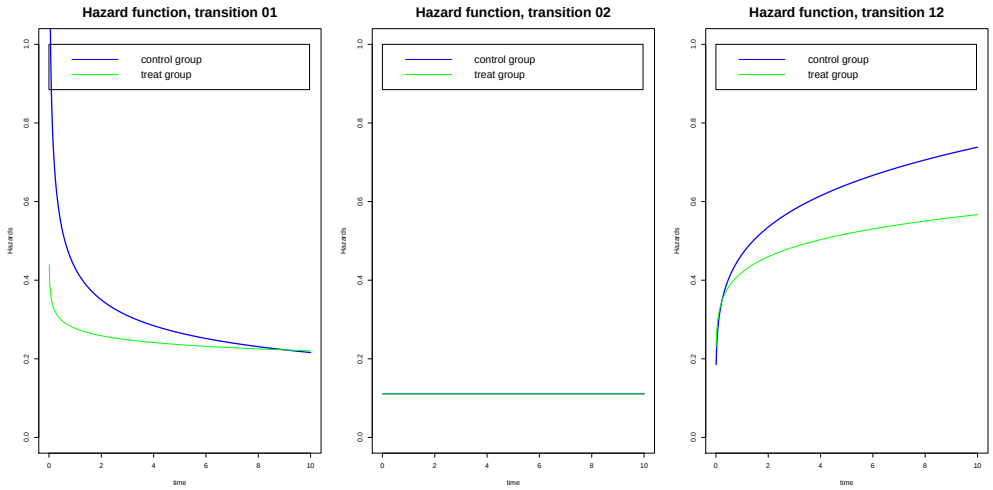

Figure C.5: Hazard functions for simulation scenario 4

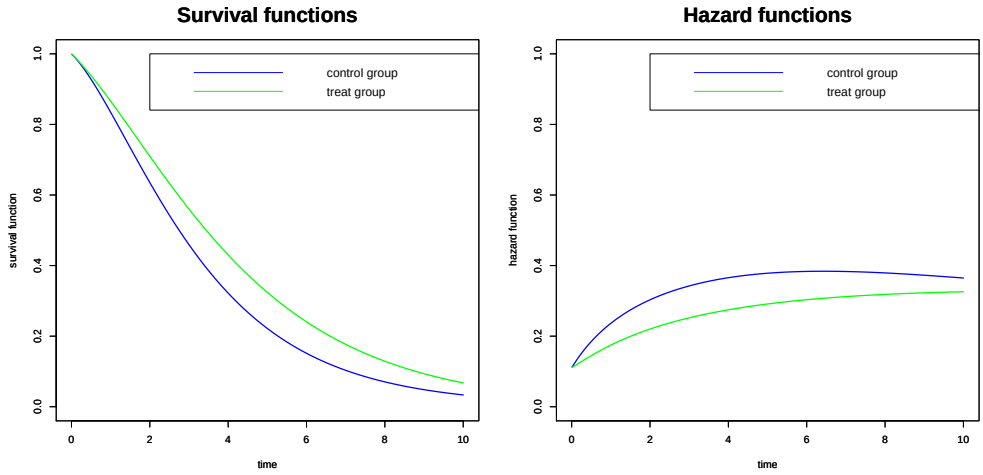

Figure C.6: Survival function of OS and cumulative hazard function for simulation scenario 4

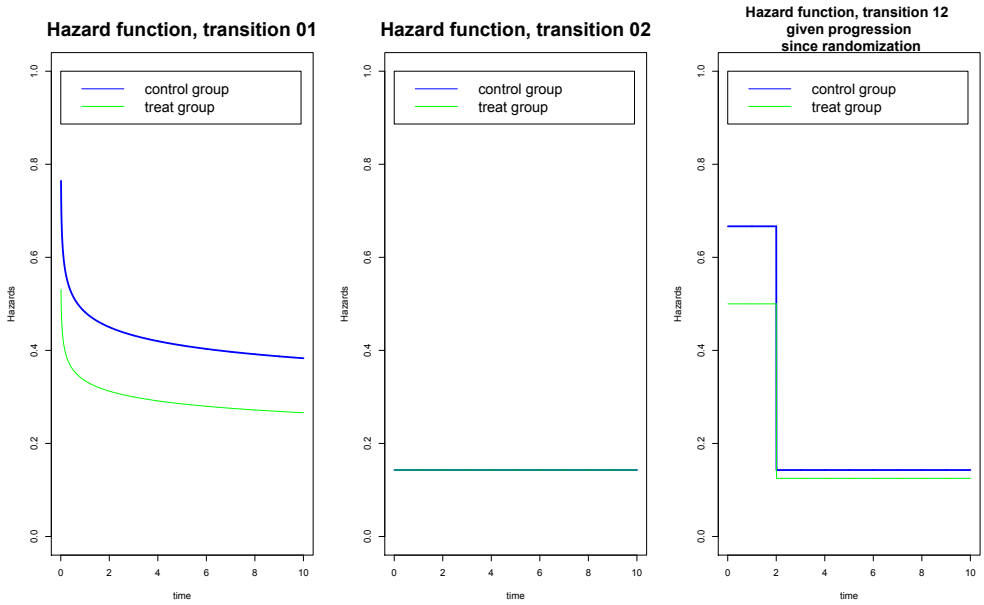

Figure C.7: Hazard functions for simulation scenario 5

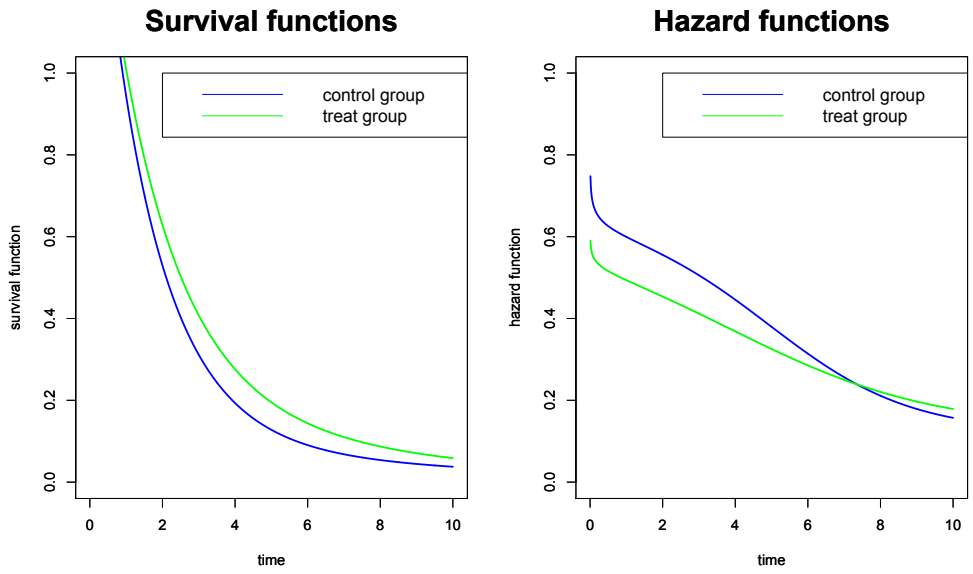

Figure C.8: Survival function of OS and cumulative hazard function for simulation scenario 5

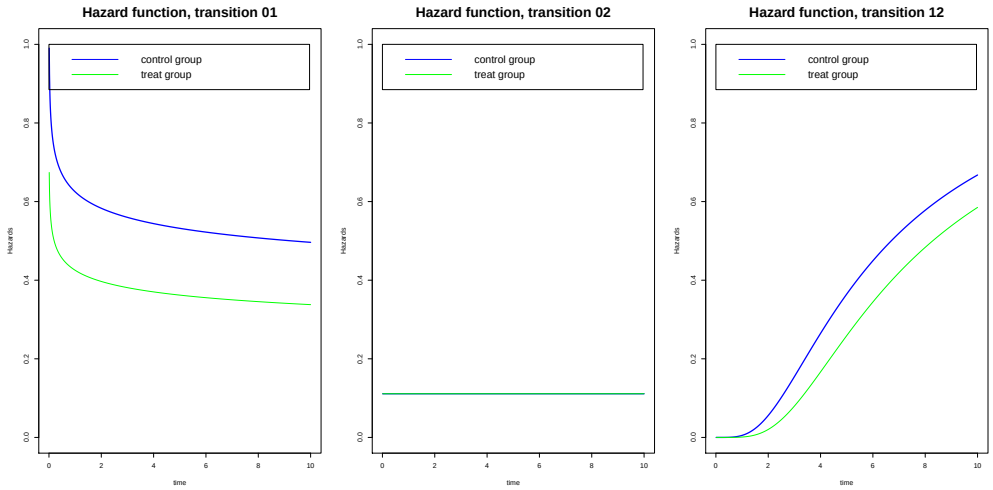

Figure C.9: Hazard functions for simulation scenario 6

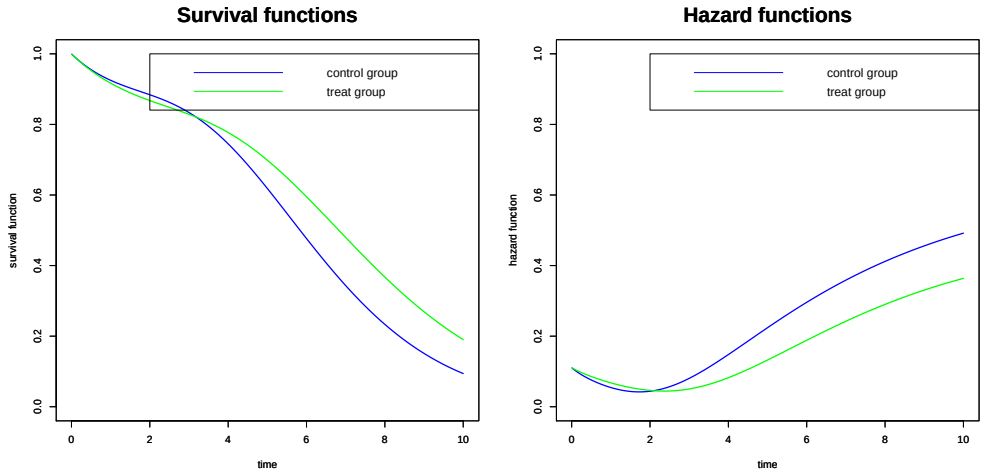

Figure C.10: Survival function of OS and cumulative hazard function for simulation scenario 6

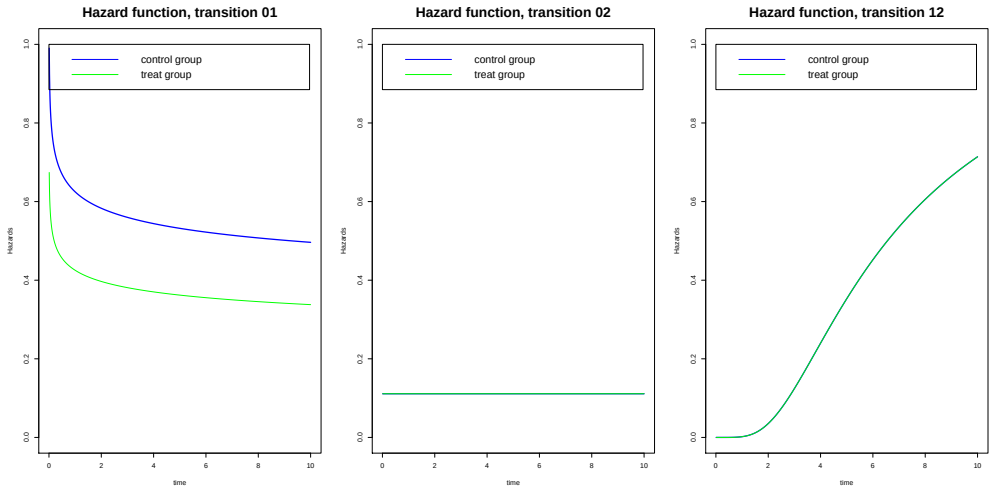

Figure C.11: Hazard functions for simulation scenario 7

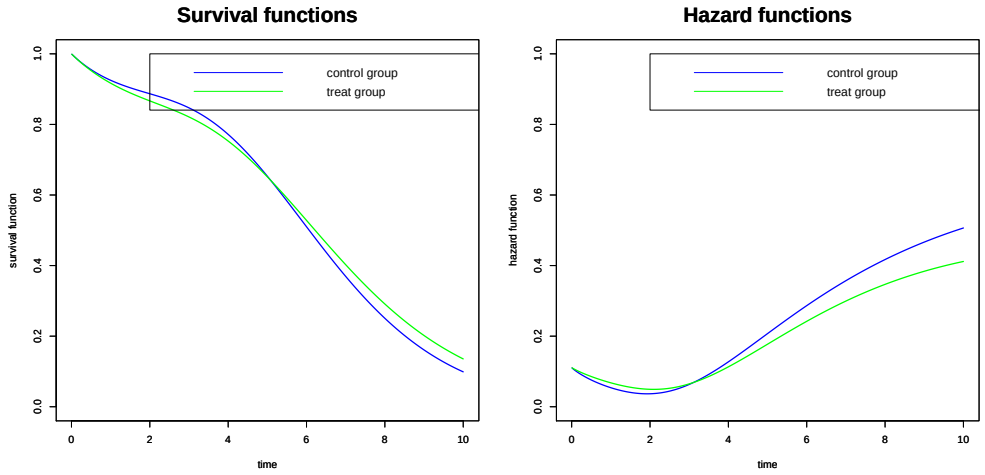

Figure C.12: Survival function of OS and cumulative hazard function for simulation scenario 7

Table C.1: Results with the estimates of the AHR, SE and BIAS from each method: Results for all scenarios, low censoring

Scenario	True HR	maximum follow-up	Model								
			semi-Markov model-based method			Markov model-based method			Cox model		
true scenario: treatment effect on transitions $0 \to 1$ and $1 \to 2$:											
1. scenario	0.7050	12.5	0.7019	0.0366	-0.0031	0.7016	0.0367	-0.0034	0.6992	0.0498	-0.0058
4. scenario	0.7571	10	0.7635	0.0408	0.0064	0.7550	0.0405	-0.0021	0.7596	0.0541	0.0025
5. scenario	0.8200	16	0.8124	0.0436	-0.0076	0.8118	0.0416	-0.0082	0.8283	0.0588	0.0084
7. scenario	0.7359	14	0.7559	0.0355	0.0199	0.7546	0.0368	0.0187	0.7125	0.0506	-0.0235
true scenario: treatment effect on transition $0 \to 1$:											
2. scenario	0.8138	13	0.8048	0.0264	-0.0090	0.8043	0.0266	-0.0095	0.8096	0.0573	-0.0042
6. scenario	0.8471	10	0.8062	0.0327	-0.0409	0.8091	0.0317	-0.0380	0.8298	0.0581	-0.0173
8. scenario	0.9482	13	0.9265	0.0173	-0.0217	1.0049	0.0100	0.0568	0.9150	0.0645	-0.0332
true scenario: treatment effect on transitions $0 \to 1$, $1 \to 2$ and $0 \to 2$:											
3. scenario	0.7419	10	0.7417	0.0548	-0.0002	0.7415	0.0549	-0.0004	0.7514	0.0536	0.0096

Investigation of performance of the model-based method for scenario 1-7, when assumption about treatment effect is misspecified

Table C.2: assumption A: treatment effect on transitions $0 \to 1$ and $1 \to 2$;
assumption B: treatment effect on transition $0 \to 1$;
assumption C: independent treatment effect on all transitions;
assumption D: same treatment effect on all transitions.

Scenario 1(maximum follow up= 6.7, true HR=0.7059)				
Results	Semi-Markov model-based method with different treatment effect assumptions			
	A (true assumption)	B	C	D
AHR	0.7062	0.8058	0.7071	0.6284
SE_{AHR}	0.0422	0.0288	0.0583	0.0479
Power	100.0000	100.0000	99.5000	100.0000
Results	Markov model-based method with different treatment effect assumptions			
	A (true assumption)	B	C	D
AHR	0.7058	0.8049	0.7066	0.6278
SE_{AHR}	0.0422	0.0289	0.0583	0.0481
Power	100.0000	100.0000	99.5000	100.0000
Results	Cox Model			
HR_{COX}	0.7058			
$SE_{HR_{COX}}$	0.0582			
Power	98.8000			

Table C.3: assumption A: treatment effect on transitions $0 \to 1$ and $1 \to 2$;
assumption B: treatment effect on transition $0 \to 1$;
assumption C: independent treatment effect on all transitions;
assumption D: same treatment effect on all transitions.

Scenario 2(maximum follow up=5, true HR= 0.8182)				
Results	Semi-Markov model-based method with different treatment effect assumptions			
	A	B (true assumption)	C	D
AHR	0.8180	0.8152	0.8196	0.7057
SE_{AHR}	0.0485	0.0273	0.0678	0.0538
Power	94.1000	100.0000	74.0000	99.5000
Results	Markov model-based method with different treatment effect assumptions			
	A	B (true assumption)	C	D
AHR	0.8176	0.8148	0.8192	0.7054
SE_{AHR}	0.0486	0.0274	0.0677	0.0540
Power	94.0000	100.0000	74.0000	99.5000
Results	Cox Model			
HR_{COX}	0.8198			
$SE_{HR_{COX}}$	0.0673			
Power	68.7000			

Table C.4: assumption A: treatment effect on transitions $0 \to 1$ and $1 \to 2$;
assumption B: treatment effect on transition $0 \to 1$;
assumption C: independent treatment effect on all transitions;
assumption D: same treatment effect on all transitions.

Scenario 3(maximum follow up=6, true HR= 0.7406)				
Results	Semi-Markov model-based method with different treatment effect assumptions			
	A	B	C (true assumption)	D
AHR	0.8520	0.9301	0.7416	0.7626
SE_{AHR}	0.0513	0.0294	0.0607	0.0564
Power	79.5000	66.0000	97.1000	96.2000
Results	Markov model-based method with different treatment effect assumptions			
	A	B	C (true assumption)	D
AHR	0.8517	0.9298	0.7410	0.7621
SE_{AHR}	0.0514	0.0295	0.0607	0.0567
Power	79.4000	65.6000	97.1000	96.2000
Results	Cox Model			
HR_{COX}	0.7447			
$SE_{HR_{COX}}$	0.0609			
Power	95.2000			

Table C.5: assumption A: treatment effect on transitions $0 \to 1$ and $1 \to 2$;
assumption B: treatment effect on transition $0 \to 1$;
assumption C: independent treatment effect on all transitions;
assumption D: same treatment effect on all transitions.

Scenario 4(maximum follow up=5, true HR=0.7541)				
Results	Semi-Markov model-based method with different treatment effect assumptions			
	A (true assumption)	B	C	D
AHR	0.7676	0.8025	0.7694	0.6780
SE_{AHR}	0.0460	0.0285	0.0624	0.0509
Power	99.3000	100.0000	93.4000	100.0000
Results	Markov model-based method with different treatment effect assumptions			
	A (true assumption)	B	C	D
AHR	0.7587	0.8082	0.7605	0.6678
SE_AHR	0.0459	0.0279	0.0619	0.0505
Power	99.5000	100.0000	93.4000	100.0000
Results	Cox Model			
HR_{COX}	0.7545			
$SE_{HR_{COX}}$	0.0610			
Power	94.0000			

Table C.6: assumption A: treatment effect on transitions $0 \to 1$ and $1 \to 2$;
assumption B: treatment effect on transition $0 \to 1$;
assumption C: independent treatment effect on all transitions;
assumption D: same treatment effect on all transitions.

Scenario 5(maximum follow up=8, true HR= 0.81966)				
Results	Semi-Markov model-based method with different treatment effect assumptions			
	A (true assumption)	B	C	D
AHR	0.79695	0.93269	0.79854	0.73030
SE_{AHR}	0.05056	0.01665	0.06815	0.05099
Power	96.40000	99.80000	79.60000	99.70000
Results	Markov model-based method with different treatment effect assumptions			
	A (true assumption)	B	C	D
AHR	0.79940	0.90444	0.80094	0.72962
SE_{AHR}	0.04736	0.02158	0.06462	0.05013
Power	97.70000	100.00000	79.60000	99.70000
Results	Cox Model			
HR_{COX}	0.80621			
$SE_{HR_{COX}}$	0.06591			
Power	75.30000			

Table C.7: assumption A: treatment effect on transitions $0 \to 1$ and $1 \to 2$;
assumption B: treatment effect on transition $0 \to 1$;
assumption C: independent treatment effect on all transitions;
assumption D: same treatment effect on all transitions.

Scenario 6(maximum follow up=5, true HR=0.8503)				
Results	Semi-Markov model-based method with different treatment effect assumptions			
	A	B (true assumption)	C	D
AHR	0.8310	0.8232	0.8313	0.7809
SE_{AHR}	0.0515	0.0335	0.0658	0.0588
Power	90.4000	100.0000	74.2000	95.8000
Results	Markov model-based method with different treatment effect assumptions			
	A	B (true assumption)	C	D
AHR	0.8250	0.8239	0.8254	0.7726
SE_{AHR}	0.0523	0.0328	0.0669	0.0596
Power	91.9000	100.0000	75.5000	96.2000
Results	Cox Model			
HR_{COX}	0.8395			
$SE_{HR_{COX}}$	0.0666			
Power	61.5000			

Table C.8: assumption A: treatment effect on transitions $0 \to 1$ and $1 \to 2$;
assumption B: treatment effect on transition $0 \to 1$;
assumption C: independent treatment effect on all transitions;
assumption D: same treatment effect on all transitions.

Scenario 7(maximum follow up=12, true HR=0.7359)				
Results	Semi-Markov model-based method with different treatment effect assumptions			
	A(true assumption)	B	C	D
AHR	0.7436	0.9360	0.7423	0.6869
SE_{AHR}	0.0412	0.0169	0.0578	0.0398
Power	100.0000	100.0000	97.8000	100.0000
Results	Markov model-based method with different treatment effect assumptions			
	A(true assumption)	B	C	D
AHR	0.7409	1.0097	0.7399	0.7128
SE_{AHR}	0.0437	0.0107	0.0594	0.0389
Power	100.0000	13.7000	97.8000	100.0000
Results	Cox Model			
HR_{COX}	0.7280			
$SE_{HR_{COX}}$	0.0597			
Power	98.0000			

C.2 Additional Application Results

Table C.9: Results of the treatment effects for each transition in the illness-death model. Cox regression model was used to model the treatment effects.

	log(hazard ratio)	hazard ratio	se(log(hazard ratio))	p_value
transition $0 \to 1$	-0.5125	0.5990	0.1187	1.57E-05
transition $0 \to 2$	-0.1069	0.8986	0.3802	0.7785
transition $1 \to 2$	0.3053	1.3570	0.1261	0.0155