Identifiability of cure models in a competing risk framework

Jacobs, T.

Tijn Jacobs

Identifiability of cure models in a competing risk framework

MSc Applied Mathematics

16 August 2023

Supervisors Prof. dr. M. Fiocco
dr. E. Musta (UvA)

Leiden University
Mathematical Institute

Abstract

In survival analysis, a competing risk model is a statistical method used to analyze time-to-event data in situations where multiple events of interest may occur and compete for occurrence. The events are considered ‘competing’ because the occurrence of one event prevents the occurrence of other events. Traditional survival analysis focuses on a single event of interest, such as death due to a particular cause. However, in real-world scenarios, there can be multiple events that individuals in a study population might experience. These events can have different causes. For example, in a study involving cancer patients, the events of interest could be death from cancer, death from other causes, and disease recurrence.

In survival analysis, a cure model is a statistical model used when a proportion of the study population is considered ‘cured’, meaning they will never experience the event of interest. This concept is particularly important when studying diseases with a good prognosis. A notable example is paediatric oncology, where patients may be considered cured if they experience long event-free survival.

Despite the growing recognition of the significance of considering cured fractions in statistical analysis, there remains limited research on the theoretical aspects of combining competing risks and cure models. The integration of these two approaches has not been extensively studied until now.

This research aims to fill the existing gap in the field by focusing on the concept of identifiability. First, a general model that involves two competing events and cause-specific cure for both events is considered. The main objective is to identify the model parameters, particularly the dependence relationship between the two cure status indicators. A logistic model to estimate cure probabilities and a semi-parametric Cox model to assess cause-specific hazards (or subdistribution hazards) are employed. The results demonstrated that, under appropriate assumptions, certain parameters can be effectively identified. However, it is also revealed that the model becomes unidentifiable without these specific assumptions. It is further shown that the models previously proposed in the literature can be seen as special cases of this general model.

The thesis presents a novel estimation procedure for the general model, utilizing the EM (Expectation-Maximization) algorithm. The flexibility of this procedure allows it to be applied to special cases of the model. Two simulation studies were conducted to investigate the performance of the estimation procedure and to study the practical identifiability properties of the model for cure and competing risks. The results showed good performance for most parameters of the model.

In conclusion, this thesis provides valuable insights into the practical identifiability of parameters

through both theoretical and simulation-based analyses. This research significantly contributes to a better understanding of competing risks and cure models. The understanding of these statistical methods enables more accurate analysis of patient outcomes and treatment effects in diverse clinical and non-clinical contexts. Ultimately, this research positively impacts the field, facilitating better decision-making and improving overall outcomes for patients and individuals in various settings.

Contents

Chapter 1

Introduction

The treatment of cancer has improved greatly over the past decades. This results in more and more patients who experience long relapse-free survival. Some patients will never experience the event of interest during their lifetime, i.e. they can be considered ‘cured’. Cure is nowadays identified in several cancers, for example, breast cancer (Rutqvist et al., 1984), colon cancer (Sargent et al., 2009) and childhood leukaemia (Bleyer, 1990). This highlights the importance of considering a fraction of cured patients when analysing and interpreting clinical trial results and raises the need for more sophisticated statistical methods: the cure model. Cure models were developed to incorporate a cured fraction of patients in the traditional survival models. These models give insight into, not only the life-prolonging effects of treatment but also the – possibly – curative effects of treatment. Therefore allowing for a disentangled interpretation of the effect of the treatment under study (Paoletti and Asselain, 2010; Yilmaz et al., 2013).

Competing risk models are crucial in clinical studies for a comprehensive analysis of patient outcomes. Competing risks arise when individuals may experience multiple potential outcomes or events, and the occurrence of one event prevents the occurrence of other events of interest. This phenomenon is particularly relevant in the field of cancer research, where patients may experience various competing risks such as disease recurrence, development of secondary malignancies, and death from unrelated causes (Koller et al., 2012). Failing to account for competing risks can lead to biased estimates of event probabilities and hinder the accurate evaluation of treatment effects. For example, if a patient dies from a related cause before experiencing disease recurrence, ignoring competing risks may overestimate the probability of recurrence. In clinical trials, this can have significant implications for assessing the efficacy of interventions and making informed treatment decisions (Van Walraven and McAlister, 2016).

This establishes the relevance of considering both the possibility of cure and several competing risks when performing a statistical analysis. However, the extension of competing risks to cure models, or vice versa, has not been thoroughly studied from a theoretical perspective. Several researchers have worked in this particular domain (Chen et al., 2020; Nicolaie et al., 2019; Zhang et al., 2019; Choi et al., 2015, 2017). Nonetheless, a significant gap exists in terms of a solid theoretical foundation supporting the work. Furthermore, both frameworks present fundamental problems with identifiability. This emphasizes the relevance of theoretical research into the identifiability problems presented by these two frameworks.

Identifiability refers to the ability to uniquely determine the parameters of a model based on the observed data. It implies that the data provide enough information to distinguish one set of parameter values from another. In other words, an identifiable model ensures that different parameter values lead to different distributions or patterns of data, allowing us to estimate the true underlying parameters accurately.

Moreover, it is a fundamental property of a statistical model and is essential for valid statistical inference and interpretation. If a model is not identifiable, it means that multiple sets of parameter values can produce the same observed data patterns, making it impossible to determine the true parameter values solely based on the data.

In a more mathematical manner, identifiability can be defined as follows. Let $\mathcal{F} = \{f_\theta \mid \theta \in \Theta\}$ denote a statistical model where Θ is a (possibly infinite dimensional) parameter space. The model is called *identifiable* if two almost everywhere equal elements from the model have the same parameters. This can be formulated as:

$$f_\theta = f_{\theta'} \quad \mathbb{P} - a.e. \implies \theta = \tilde{\theta} \quad \text{for all } \theta, \tilde{\theta} \in \Theta.$$

As mentioned before, taking both cure and competing risks into account is highly relevant in certain – clinical and non-clinical – settings. This thesis aims to delve into the extension of the cure model to incorporate competing risks, with a special focus on the identifiability aspect. We will define a generalized model with a cause-specific notion of cure. Several models discussed in the methodological literature will become special cases of this generalized model.

The thesis is structured as follows. First, in Chapter 2, the basic tools from survival analysis are introduced and a background is given to the identifiability problems arising in competing risks analysis and cure analysis. In Chapter 3 several extensions of the cure model to incorporate competing risks are given. It is investigated whether these extended models are identifiable or not. After having studied the identifiability problems from a theoretical perspective, the practical identifiability problems are investigated. This starts with an estimation procedure of these different models which is given in Chapter 4. In Chapter 5 a simulation study is performed to investigate the finite sample performance of the proposed methodology. The thesis ends with a discussion in Chapter 6.

Chapter 2

Basics of Survival Analysis

Survival analysis is the statistical field concerned with the analysis of time-to event-data. Since time is needed to observe the event of interest, survival data is characterized by the presence of censored observations. *Censoring* occurs when the exact time to event for an individual is not observed. In this project, solely right censoring is considered. *Right censoring* occurs when the exact time to the event of interest is not observed up to the end of the study period. If the event of interest is the recurrence of a tumour, for example, some patients may still be under observation at the end of the study without experiencing the event of interest or they may have left the study before. It is then only known that up to that time the tumour has not yet recurred, but the exact time of recurrence is unknown. This can, for example, happen when a patient moves away and drops out of a clinical trial or because the study ends. In this thesis, it is assumed that censoring is not related to the occurrence of the event of interest, i.e. the censoring mechanism is non-informative. Informative censoring will not be further discussed.

Let C and T denote the random variables indicating the censoring time and the lifetime (or time-to-event of interest) respectively. The follow-up time is then given by $T^* = \min(T, C)$. Censoring occurs thus when $C < T$, i.e. the time until censoring was smaller than the time-to-event. In addition to the follow-up time T^*, we observe the status $\delta := \mathbf{1}(T^* = T)$. The random variable δ indicates whether the event was observed ($\delta = 1$) or censored ($\delta = 0$). The data structure for the i-th subject is thus given by (T_i, δ_i). If we also consider a covariate vector $\mathbf{Z}$, the data can be represented as follows: $(T_i, \delta_i, \mathbf{Z}_i)$.

The other types of censoring data are left-censored data (it occurs when the event of interest has occurred or started before the data collection began) and interval-censored data (it only specifies the time within a certain interval where the event occurred). These other types of censoring will not be discussed in this thesis.

In this chapter, we will first introduce the basic functions used in survival analysis. These include the survival function, hazard rate and cumulative hazard function. Then it will be shown how the likelihood is constructed for survival data. The non-parametric Kaplan-Meier estimator for the survival function and the semi-parametric Cox regression proportional hazard model for regression of survival data are introduced. Furthermore, the concept of competing risks and the standard cure model will be presented.

2.1 The basic functions

In this section, the basic functions from survival analysis are introduced. We will present standard identities and introduce the intuition behind the concepts.

The *survival function* $S(t)$ represents the probability of surviving up to a certain time $t \geq 0$. It is defined as:

$$S(t) := \mathbb{P}(T > t) = 1 - \mathbb{P}(T \leq t) = 1 - F(t) = 1 - \int_0^t f(u)du = \int_t^\infty f(u)du, \tag{2.1}$$

where F denotes the cumulative distribution function of T. We assume throughout this thesis that T is a continuous random variable with a probability density function f. This in turn implies that:

$$f(t) = -\frac{\partial S(t)}{\partial t}. \tag{2.2}$$

If T is a continuous random variable, the survival function is a continuous and decreasing function with the property that $S(0) = 1$ and $S(t) \geq 0$ for all $t \geq 0$. That is, the probability of being alive at the beginning of the study is one and the probability of surviving up to any time is at least zero. A proper survival function has the property that the probability of surviving forever is zero, i.e. $\lim_{t\to\infty} S(t) = 0$. Indicating that all individuals under study will eventually experience the event of interest.

The *hazard rate* $\lambda(t)$ is the next fundamental quantity in survival analysis. It is defined as:

$$\lambda(t) := \lim_{dt\to 0} \frac{\mathbb{P}(t \leq T < t + dt \mid T \geq t)}{dt}, \tag{2.3}$$

and expresses the instantaneous rate of occurrence of the event, given that the individual has survived up to time t. Furthermore, if T is a continuous random variable, the hazard rate can be written in terms of the density and survival function in the following manner:

$$\lambda(t) = \frac{f(t)}{S(t)}. \tag{2.4}$$

Closely linked to the hazard rate is the *cumulative hazard function*. It is defined as:

$$\Lambda(t) := \int_0^t \lambda(u)du. \tag{2.5}$$

Combining this with formula (2.2), it can be seen that:

$$S(t) = \exp\{-\Lambda(t)\}. \tag{2.6}$$

Note that the cumulative hazard is not a probability. It is merely a measure of the accumulated risk of experiencing the event. A high cumulative hazard indicates a low probability of survival and vice versa.

2.2 The likelihood for survival data

In the presence of right-censoring, the likelihood function needs to be constructed with a bit more care, since we have two types of observations: $(T^* = t, \delta = 1)$ and $(T^* = t, \delta = 1)$. These correspond, respectively, to the events: $\{T = t, C > t\}$ and $\{C = t, T > t\}$. These two different types of observations also have different contributions to the likelihood.

Throughout this thesis, we will sometimes write $\mathbb{P}(T = t)$ for a continuous random variable T, indicating the density evaluated at a specific point t. Although this is an abuse of notation, it eases interpretation and ought not to lead to confusion.

Now assume that the data is right-censored and that the censoring is independent of the survival time T and non-informative, that is, the distribution of the censoring times C does not depend on the parameters of the lifetime distribution. Independence between T and C will be assumed throughout the thesis. The likelihood for a censored and uncensored observation are, respectively, given by:

$$L_0(\theta) = f(t) \quad \text{and} \quad L_1(\theta) = S(t),$$

where θ denotes the set of all relevant parameters. Note that we are ignoring multiplicative terms that depend on the censoring distribution ($\mathbb{P}(C > T)$ and $\mathbb{P}(C = t)$ respectively) because they do not provide any information and do not affect the maximization of the likelihood. An uncensored observation thus contributes to the likelihood by means of the density, while a censored observation contributes through the survival function. This coincides with the intuition that at the moment of censoring all we know is that the event time is larger than the observed censoring time. Then, for the independent pairs (T_i, δ_i) of random variables, the likelihood can be written as:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} = \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S(t_i) = \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} \exp\{-\Lambda(t_i)\}, \tag{2.7}$$

where the identities given in (2.4) and (2.6) were used. These likelihoods will be further developed when considering competing risks and cure models.

2.3 Kaplan-Meier estimator of the survival function

The functions presented in Section 2.1 can be estimated using parametric or non-parametric methods. For example, a flexible parametric model for the estimation of the survival function is given by the Weibull model:

$$S(t) = \exp\big\{-(\lambda t)^k\big\},$$

where $\lambda, k > 0$ are, respectively, the scale and shape parameters. Note that when $k = 1$ the Weibull distribution reduces to an exponential distribution. The Weibull model for survival is one of the most used parametric models. Although the model is flexible, misspecification of the statistical model is a fatal pitfall. Therefore, we will focus on non-parametric and semi-parametric methods in this thesis.

2.3.1 Non-parametric estimation of survival functions

Let $t_1 < t_2 < ... < t_m$ denote the time points at which an event was observed. Furthermore, define d_j to be the number of observed events on t_j and n_j the number of subjects at risk on t_j. Then the Kaplan-Meier estimator (Kaplan and Meier, 1958) of the survival function given by:

$$\widehat{S}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j}\right). \tag{2.8}$$

This estimator is also known as the *product-limit* estimator and is widely used in practice. It is a stepwise function with jumps at the time points at which we observe an event. In the absence of censored observations, this estimator reduces to the complement of the empirical distribution function. The variance of the estimator for a given time point t can be estimated by Greenwood's formula:

$$\widehat{\tau}(t) = \widehat{S}^2(t) \sum_{j:t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}. \tag{2.9}$$

2.4 Semi-parametric Proportional Hazards regression Model

To consider the effect of covariates, some form of regression model is required. In this thesis, we will focus on Cox proportional hazards regression model (Cox, 1972). This method is widely used in survival analysis and medical applications to model the effects of covariates on survival probabilities. It gives a straightforward interpretation of the relative risk of covariates and is easily implemented using the `survival` library (Therneau et al., 2021) in `R` software environment (R Core Team, 2022).

The data consists of the triples $(T_i^*, \delta_i, \mathbf{Z}_i)$ for $i \in \{1, 2, ..., n\}$. Here T_i^*, δ_i and $\mathbf{Z}_i \in \mathcal{X}$ are, respectively, the follow-up time, the status indicator and the vector of covariates of the i-th subject, where $\mathcal{X}$ the covariate space. The effect of the covariates is modelled as follows:

$$\lambda(t \mid \mathbf{Z}) = \lambda_0(t) \exp\{\beta^\top \mathbf{Z}\}. \tag{2.10}$$

Here $\lambda_0(t)$ denotes the baseline hazard and β is a vector of regression parameters. The baseline hazard corresponds to the hazard rate of subjects with all covariates equal to zero and is left unspecified. An expression for the survival function in the Cox model can be derived:

$$\begin{aligned} S(t \mid \mathbf{Z}) &= \exp\left\{-\int_0^t \lambda(u \mid \mathbf{Z})\, du\right\} \\ &= \exp\left\{-\int_0^t \lambda_0(t) \exp\{\beta^\top \mathbf{Z}\}\, du\right\} \\ &= \left(\exp\left\{-\int_0^t \lambda_0(t)\, du\right\}\right)^{\exp\{\beta^\top \mathbf{Z}\}} \\ &= S_0(t)^{\exp\{\beta^\top \mathbf{Z}\}}, \end{aligned} \tag{2.11}$$

where $S_0(t)$ denotes the baseline survival function. It can be interpreted as the survival function for those individuals with covariates all equal to zero. Moreover, an expression for the density in the Cox model can be derived. It is given by:

$$f(t \mid \mathbf{Z}) = \lambda(t \mid \mathbf{Z}) S(t \mid \mathbf{Z}). \tag{2.12}$$

Note that for two subjects with covariates $\mathbf{Z}_1$ and $\mathbf{Z}_2$ the ratio of hazard rates is given by:

$$\frac{\lambda(t \mid \mathbf{Z}_1)}{\lambda(t \mid \mathbf{Z}_2)} = \frac{\lambda_0(t) \exp\{\beta^\top \mathbf{Z}_1\}}{\lambda_0(t) \exp\{\beta^\top \mathbf{Z}_2\}} = \exp\{\beta^\top (\mathbf{Z}_1 - \mathbf{Z}_2)\}. \tag{2.13}$$

This is a constant with respect to time. More specifically, it is the relative risk of an individual with covariates $\mathbf{Z}_1$ experiencing the event compared to an individual with covariates $\mathbf{Z}_2$. For this reason, the Cox model is often called the *proportional hazards (PH) model* or *Cox proportional hazards model.* So the relative risk is constant, i.e. it does not change over time. This is a crucial assumption of the model. A violation can lead to misleading conclusions. The proportional hazards assumption can be assessed through visual assessment of the Kaplan-Meier curves, $\log\{-\log\}$ plots and testing of scaled Schoenfeld residuals. Literature on the PH assumption and on methods for assessing is vast – see for example Barlow and Prentice (1988), Therneau et al. (1990) or Schoenfeld (1982).

2.5 Competing risks

Competing risks data consists of subjects who are at risk for multiple types of events, denoted by $k = 1, 2, ..., K$. Competing events are characterized by the fact that their occurrence precludes any other event. Let D denote the random variable indicating which of the competing events occurred first. The term 'competing risks indicator' will be used to denote D, contrasting with the previously introduced status indicator. For example, in the case of a cancer study, one can consider local recurrence of the tumour and distant metastasis as competing events. A concise treatment of the theory and applications of competing risks is given by Putter et al. (2007) and a more extensive treatment can be found in Crowder (2001) and Crowder (2012).

In this section, the historic approach of potential survival times to competing risks is explained. This approach has a fatal identifiability problem as shown by Tsiatis (1975). Afterwards, two different hazard functions for competing risks are introduced: the cause-specific hazard and the subdistribution hazard. Both are observable from the data. These hazard functions have different interpretations and, therefore, serve different purposes. This aspect is discussed in the last subsection.

2.5.1 Historic approach

Historic approaches considered competing risks as a multivariate failure time model. Any individual would have a failure time distribution for each of the competing events. The first event that occurs is observed, and all others are latent variables. That is, if we have an uncensored observation, we only observe the *actual survival time* $T = \min\{T_1, T_2, ..., T_K\}$, $D = k$ if $T = T_k$ indicating which event was observed with $D = 0$ if the observation was censored. The unobserved T_i can be referred to as *potential survival times*, as they would have potentially occurred. Censoring is assumed to be independent of the competing events.

The joint distribution is given by:

$$\bar{S}(t_1, t_2, ..., t_K) = \mathbb{P}(T_1 > t_1, T_2 > t_2, ..., T_K > t_k) \tag{2.14}$$

The marginal survival function is given by $S_k(t) = \mathbb{P}(T_k > t) = \bar{S}(0, ..., 0, t, 0, ..., 0)$. A major problem with this approach is identifiability. It was shown by Cox (1959) and Tsiatis (1975) that without extra assumptions on the dependence structure, the joint distribution is not uniquely identifiable. Tsiatis (1975) showed that for any joint distribution of competing events, one can find a joint distribution function with independently distributed competing events such that they both lead to the same cumulative incidence function, i.e. are indistinguishable just based on the observed data. As one is generally interested in unravelling the distribution of events which are dependently distributed. This causes identifiability problems. The joint distribution and marginal distributions of the competing events are therefore not identifiable. In addition, the independence of survival times cannot be tested. In order to overcome this problem, the focus shifted to estimating two alternative hazard functions: the cause-specific hazard and the subdistribution hazard which will be introduced in the next section.

2.5.2 Two different hazard functions

For any of the competing events we can define the *cause-specific hazard*:

$$\lambda_k(t) := \lim_{dt \to 0} \frac{\mathbb{P}(t \leq T < t + dt, D = k \mid T \geq t)}{dt}. \tag{2.15}$$

It expresses the instantaneous rate of occurrence of a particular event in individuals who have not experienced any event. The cause-specific hazard can be uniquely determined from the data. Moreover, any quantity based on the cause-specific hazard can be uniquely determined from the data. We can therefore define our functions in a competing risk setting based on the cause-specific hazard. First, the *cause-specific cumulative hazard* is given by:

$$\Lambda_k(t) = \int_0^t \lambda_k(u) \; du. \tag{2.16}$$

Next, define:

$$S_k(t) = \exp\{-\Lambda_k(t)\}. \tag{2.17}$$

Since the S_k's are based on the cause-specific hazard, they can be estimated, but we cannot interpret them as marginal survival functions. The quantity S_k can be interpreted as a marginal survival function if the competing event time distributions and the censoring distribution are independent. Next, define:

$$S(t) = \exp\left\{-\sum_{k=1}^{K} \Lambda_k(t)\right\}. \tag{2.18}$$

This function has a survival probability interpretation. It is the probability of not having experienced any of the K events at time t. With these functions, we can define the *cumulative incidence* *function (CIF)*:

$$I_k(t) := \mathbb{P}(T \leq t,\ D = k) = \int_0^t \lambda_k(u) S(u) du. \tag{2.19}$$

The cumulative incidence function expresses the probability of having experienced an event of cause k before time t. We note that $\lim_{t\to\infty} I_k(t) = \mathbb{P}(D = k) \leq 1$. This indicates that $I_k(t)$ is not a proper probability distribution. In the literature, the cumulative incidence function is therefore often called the *sub-distribution function.* It can be estimated non-parametrically. Define the following quantities:

$$\widehat{\lambda}_k(t_j) = \frac{d_{kj}}{n_j} \quad \text{and} \quad \widehat{S}(t) = \prod_{j:t_j \leq t} \left(1 - \sum_{k=1}^{K} \widehat{\lambda}_k(t_j)\right), \tag{2.20}$$

where d_{kj} is the number of observed events of type k at time t_j. Then the estimator for the cumulative incidence function of cause k at time t is given by:

$$\widehat{I}_k(t) = \sum_{j:t_j \leq t} \widehat{\lambda}_k(t_j) \widehat{S}(t_{j-1}). \tag{2.21}$$

The hazard rate that exhibits a one-to-one relationship with the cause-specific cumulative incidence is known as the *subdistribution hazard* introduced by Fine and Gray (1999). It is defined by:

$$\lambda_k^{sd}(t) := \lim_{dt\to 0} \frac{\mathbb{P}(t \leq T < t + dt, D = k \mid T \geq t \cup (T < t \cap D \neq k))}{dt}. \tag{2.22}$$

It expresses the instantaneous rate of occurrence of a particular event in individuals who have not experienced an event of that type. So we are considering the rate of the event in individuals who have either not experienced any event or have experienced any of the other competing events.

This approach considers individuals still at risk for an event of cause k after they experience the competing event j. A possible explanation is that the subdistribution only considers risk k and does not want any information about the occurrence of other competing events. Unlike the cause-specific hazard, the risk set in the subdistribution hazard decreases at each time point when there is an occurrence of failure from any other cause.

The subdistribution hazard can be written as:

$$\lambda_k^{sd}(t) = -\frac{\partial \log(1 - I_k(t))}{\partial t}. \tag{2.23}$$

2.5.3 Two different modelling approaches

In the previous section, two hazard functions were proposed with each a different interpretation. Dependent on the goal of the study one of the two ought to be chosen wisely. A different modelling approach can be attributed to each hazard. The direct modelling of the subdistribution hazard was first proposed by Fine and Gray (1999). Interchanging the different modelling purposes and different interpretations of the hazard is a common pitfall among applied researchers. For an accurate and comprehensible discussion of the differences, the reader is referred to Austin and Fine (2017)

When the cause-specific hazards are modelled, each hazard is analyzed separately by treating individuals failing from other causes as censored observations. The cause-specific hazard ratio represents the relative change in the rate of occurrence of the event of interest in subjects who have not experienced any events yet. This rate is a measure of the frequency with which events happen and not a measure of the incidence of the event.

On the other hand, the subdistribution hazard function accounts for the competing risks and estimates the cumulative incidence function of the event of interest, taking into account the presence of competing risks. This analysis does not treat individuals failing from other causes as censored observations. The subdistribution hazard is used to determine factors associated with the incidence of a given event, and it assumes that the occurrence of competing events affects the hazard of the event of interest. The subdistribution hazard can also be interpreted as the hazard rate of the event of interest among a hypothetical population in which the competing risks have been eliminated. It is thus a measure of the incidence of the particular event (Austin et al., 2021; Putter et al., 2020).

2.6 Introduction to cure models

In this section, we will briefly introduce the standard cure model, i.e. the cure model without the presence of competing risks. In Section 2.1 the survival function was introduced, and it was mentioned that it is proper if $\lim_{t\to\infty} S(t) = 0$. Nonetheless, if certain patients are not susceptible to the event of interest, they will never experience it. This could, for example, happen when a patient is immune to a certain disease under study. That patient will never experience the disease. The survival time is therefore infinite, resulting in a survival function that is not proper. The limiting value of the survival function of the whole population is then given by: $\lim_{t\to\infty} S(t) = \alpha > 0$, where $\alpha \in (0, 1]$ is the fraction of 'cured' individuals.

The cure model starts from the assumption that *at baseline* any individual is either cured or susceptible to the event of interest. An individual is cured if he is immune to the event of interest and will never experience the event ($T = \infty$). It is thus from onset determined whether one can or cannot experience the event. The population is therefore classified into two groups: cured and susceptible, i.e. a mixture with a relative size equal to α and $1 - \alpha$. The model following this approach, *the mixture cure model*, has been introduced by Boag (1949) and Berkson and Gage (1952). The works of Farewel (1977, 1982) have further developed these models in a parametric fashion. Later, extensions to semi-parametric and non-parametric models were provided. A comprehensible review of the cure model can be found in Amico and Van Keilegom (2018). Legrand and Betrand (2019) give an overview of the model with a focus on the application to oncology. At last, a general and comprehensive overview of cure models and their extensions is given by Peng and Yu (2021).

2.6.1 The mixture cure model

Let $\mathbf{X}$, $\mathbf{Z}$ and B denote two covariate vectors and the cure status (i.e. $B = 1$ indicating cure and $B = 0$ indicating susceptible) respectively. If we observe an uncensored event, we know that that person is uncured. This is not the case for a censored observation. If we observe a censored event, we know that that individual has not experienced the event yet, but we do not know whether he will eventually experience the event. That individual can thus either be cured or still susceptible.

This can be put in a more mathematical formulation as follows:

$$\{\delta = 1\} \implies \{B = 0\} \quad \text{and} \quad \{\delta = 0\} \implies \{B = 0\} \vee \{B = 1\}. \tag{2.24}$$

This implies that cure status is only partially observed as we cannot observe the cure status for censored observations. We can therefore construct a probability distribution for the cure status indicator B. Let $\pi(\mathbf{X}) = \mathbb{P}(B = 1 \mid \mathbf{X})$ denote the probability of being cured at baseline given the covariates $\mathbf{X}$. Since cured individuals will never experience the event of interest, their survival time is infinite, i.e. $T = \infty$. This implies that their probability to survive up to and including a time $t \geq 0$ equals one, that is $\mathbb{P}(T \geq t \mid B = 1, \mathbf{Z}) = 1$. The survival function of the mixture population $S(t \mid \mathbf{X}, \mathbf{Z}) = \mathbb{P}(T > t \mid \mathbf{X}, \mathbf{Z})$, can then be written as follows:

$$S(t \mid \mathbf{X}, \mathbf{Z}) = \pi(\mathbf{X}) + (1 - \pi(\mathbf{X}))S_u(t \mid \mathbf{Z}). \tag{2.25}$$

Here $S_u(t \mid \mathbf{Z}) = \mathbb{P}(T \geq t \mid B = 0, \mathbf{Z})$ denotes the survival functions of the uncured patients. As discussed before, in the presence of a cure fraction, the traditional survival function of the whole population is not a proper survival function. This can be seen from the fact that $\lim_{t\to\infty} S(t \mid \mathbf{X}, \mathbf{Z}) = \pi(\mathbf{X})$. In other words, if we could wait for an infinite amount of time, only the cured fraction of the population would have survived. Contrary, the survival function for the susceptible patients is a proper survival function, that is, $\lim_{t\to\infty} S_u(t) = 0$. If an individual is uncured, he will eventually experience the event of interest.

Furthermore, note that the sets of covariates $\mathbf{X}$ and $\mathbf{Z}$ can be different. This coincides with the intuition that the risk factors associated with the long- and short-term effects do not have to be the same. The incidence and latency are, respectively, modelled with a logistic regression model and a proportional hazards model. In this thesis, we focus exclusively on these models due to their common practical use. However, one has the option to select other models for the incidence and latency in general. For simplicity, in the rest of this work, it is assumed that these sets of covariates coincide.

2.6.2 Identifiability of the cure model

The identifiability of cure models was systematically studied by Li et al. (2001) and later by Hanin and Huang (2014). An important condition for the model to be identifiable is the existence of a *cure threshold* $\tau > 0$ defined as an upper bound of survival time of uncured individuals:

$$\tau := \inf\{s > 0 \ : \ \mathbb{P}(s < T < \infty) = 0\}. \tag{2.26}$$

This definition implies that:

$$\{T > \tau\} \implies \{T = \infty\}. \tag{2.27}$$

This means that if somebody survives up to the cure threshold, it is, almost surely, known that he is cured. To ensure identifiability, the mere presence of a cure threshold is not enough. It is also necessary to observe this cure threshold in the data. In other words, the duration of the study follow-up period must be bigger than the cure threshold, i.e. $\mathbb{P}(C > \tau)$. After the cure threshold τ we almost surely do not observe any event.

This assumption can be checked from the data – or at least made plausible – by looking at the Kaplan-Meier estimate of the survival probability. The estimate is characterized by a plateau in the survival probability which contains a significant amount of censored observations. This plateau indicates that no events are occurring after a certain point in time, and suggests the existence of a cure threshold. The practical implication: clinical trials investigating diseases with good prognoses ought to have sufficient follow-up to observe the cure threshold. Additionally, you need some practical e.g. medical knowledge that supports such an assumption. Some statistical tests have been developed for testing for sufficient follow-up (Maller and Zhou, 1996), the presence of a cure fraction (Zhao et al., 2009; Hsu and Todem, 2016) and the proportional hazards assumption for the uncured individuals (Peng and Taylor, 2017; Wileyto et al., 2013). For identifiability results for the standard cure model, the reader is referred to Hanin and Huang (2014) and Parsa and Van Keilegom (2023).

Remark. *Hanin and Huang (2014) use a different definition of identifiability. It contrasts with our notion of 'almost sure' identifiability as introduced in Chapter 1. Therefore some explanations are provided. Let $S(t \mid \mathbf{X}, \mathbf{Z})$ as defined in (2.25) and define $\tilde{S}(t \mid \mathbf{X}, \mathbf{Z})$ in a similar way. Then the mixture cure model is identifiable (Hanin and Huang, 2014) if for all $\mathbf{X}, \mathbf{Z} \in \mathcal{X}$ and $t > 0$:*

$$S(t \mid \mathbf{X}, \mathbf{Z}) = \tilde{S}(t \mid \mathbf{X}, \mathbf{Z}) \implies \pi(\mathbf{X}) = \tilde{\pi}(\mathbf{X}) \quad \textit{and} \quad S_u(t \mid \mathbf{Z}) = \tilde{S}_u(t \mid \mathbf{Z}). \tag{2.28}$$

Note that our definition of 'almost sure' identifiability is stronger. The definition in Hanin and Huang (2014) relies on the identifiability of the models chosen for π and S_u to uniquely identify the parameters. If a model is identifiable in the 'almost sure' sense, then the parameters can be uniquely determined from the data.

2.7 Cure in the presence of competing risks

The literature on cure in the presence of competing risks is scarce. Few methodological papers have been published on this subject. Each is motivated by different practical applications. Throughout these papers, different definitions of cure in the presence of competing events are used all motivated by the application at hand. In a setting where a subject can experience multiple events, it is not immediately clear what *cure* means.

In this section, we outline the different definitions of cure in a competing risk setting. The motivating applications are highlighted and – in some cases – an introduction to the induced model is given.

2.7.1 Cure as immunity to the risk of interest

Here, cure is defined as being insusceptible to only the risk of interest. So being cured means that an individual will never experience the event of interest, and can experience any of the competing events. This perspective is proposed by Basu and Tiwari (2010) and has not gained much attention in the literature. It is motivated by the application to breast cancer data from the 'Surveillance, Epidemiology, and End Results' program of the US National Cancer Institute. The data contains information on primary (and possibly secondary) cancers, as well as cause-of-death information for non-survivors. As the prognosis for breast cancer has improved greatly over the past decades a significant part of the patients are still alive at the end of the follow-up. We can consider those patients as cured. However, cured patients are still susceptible to death due to secondary cancer or death due to other reasons. Among those patients, an individual cannot be cured.

The model proposed by Basu and Tiwari (2010) models the subdistribution hazard directly and uses a Bayesian estimation procedure. As identifiability is less of a problem in the Bayesian framework, details about this model are not further provided.

2.7.2 Cure as immunity to all risks

Cure can also be defined as being not susceptible to any of the competing risks. In that case, being cured precludes the of occurrence events of all types. We will refer to this perspective as *complete cure*. If we, for example, consider the case of osteosarcoma, cure entails being insusceptible to local recurrence, distant metastasis and death due to cancer. It truly means that the patient will never experience anything related to the original sarcoma. Several methodological articles (Choi et al., 2015, 2017; Chen et al., 2020) adhere to this view on cure in a competing risk setting.

2.7.3 Cure as immunity to a subset of the risks

In a more general view, cure can be defined as being insusceptible to a subset of the competing risks. This means that a cured individual can experience some of the events, while it is immune to others. This approach is discussed by Zhang et al. (2019) and motivated by credit scoring of online consumer loans. The mixture cure model without competing was already applied to credit scoring loans by Tong et al. (2014). For credit scoring purposes, the time-to-default is measured and ongoing loans can be considered as right-censored. Since most people will not default during the loans' lifetime, it is appropriate to use a cure mode, where being cured means not going into default.

The extension to competing risks is motivated by prepayments. Prepayment is a different endpoint of the study and precludes defaulting. According to the authors, there may exist a group of people who will never default nor prepay and a sub-population of the loans who are immune to defaulting but can prepay. Here the structure of the cure mixture model is not always evident and thus a more flexible interpretation of cure is needed. The authors deal with this problem by estimating four models. These are given by:

(Model A) All individuals are susceptible to both competing events. This coincides with the classical competing risk model without cure.

(Model B) A sub-population is cured of event 1, and all individuals are susceptible to event 2.

(Model C) A sub-population is cured of event 2, and all individuals are susceptible to event 1.

(Model D) A sub-population is cured of both risks and the others are not cured of any of the risks. This coincides with the complete cure model introduced above.

The authors only consider these fixed cases but do not consider the general case in which the population consists of a subpopulation that is cured for event 1 but not for event 2, a subpopulation that is cured for event 2 but not for event 1, a subpopulation that is cured for both and a last one who is cured for none.

The authors used model-selection based on the AIC scores (where the model with the lowest AIC was selected) to find the cure structure. This approach does not provide a measure of certainty about the different cure models, i.e. it merely gives an indication which of the models fits the data better. It is illustrative for the question at hand: can we recover the subset of risks for which one can be cured from the data? We will return to this question – and to this specific example – later.

Chapter 3

Identifiability of the cure model with competing risks

In the previous chapter, several definitions of cure in a competing risk framework were introduced. These different definitions yield a different mathematical structure of the cure model. The main goal of this chapter is to investigate in which settings and under which assumptions the cure model in the presence of competing risks is identifiable. The identification of the cure fractions is here of particular interest.

First, the concept of cure structure will be discussed. This concept captures the specification of cure in a setting where competing risks are present. Then, it will be shown that – under the assumption of independent potential survival times – both the distribution of the survival times and the cure fractions are identifiable. Several particular cases of cure structures are highlighted under independence. Next, the assumption of independence will be dropped and it will be shown that neither the distribution of the potential survival times nor the cure fractions are identifiable under a general cure structure. Finally, it is proven that if the cure occurs simultaneously for all competing events, the cause-specific hazards, the sub-distribution hazards and the Vertical model are identifiable.

3.1 The cure structure

The cure structure was already discussed in the previous chapters, although a formal definition was omitted. It refers to the division of competing risks into two subsets: those for who one can be cured and those for who one is always susceptible. Within this framework, if an individual is cured of any of the events for which that individual can be cured, it will be cured of all the events in that subset. For example, for the definition of *complete cure* presented in Section 2.7.2, this classification is clear: an individual can be cured of all competing events simultaneously, while there is no event for which a subject remains susceptible once cured.

In this thesis, we will focus on a model with two competing risks, i.e. $K = 2$. This implies that there are four, $2^K = 4$, possible cure structures for an individual. Those are given by:

1. An individual is cured of all competing events simultaneously, i.e. *complete cure*.

2. An individual is only cured of competing event one.
3. An individual is only cured of competing event two.
4. An individual is not cured at all.

A priori it is not always evident for which risks an individual is cured and for which ones it is not. This can be illustrated by the breast cancer example Basu and Tiwari (2010) presented in Section 2.7.1. The authors consider death due to three competing risks: breast cancer-related, another type of cancer-related and death not cancer-related. The authors define that a patient is cured if one does not die due to breast cancer. In this context, cure is regarded as immunity to the competing event of interest. Consequently, a patient cannot be cured of secondary cancers that arise due to primary cancer. Secondary malignancies are also often considered as part of the events from which one is cured if a patient is cured of the original cancer. Therefore, we could consider a cure structure where one is cured of death due to cancer and death due to secondary cancers simultaneously. This highlights the ambiguity of the definition of cure in the presence of competing risks. We will therefore allow for an individual to belong to one of these four categories and not make any assumptions about the cure structure. At the population level, the cure structure can therefore be a mixture of the four discussed before. This extends the model presented by Zhang et al. (2019).

In this chapter, we will investigate whether it is possible to recover the cure structure from the data. We introduce random variables related to the cure structure. The cure status for each competing risk is denoted by Bernoulli random variables B_i $(i = 1, 2)$ where $B_i = 1$ indicates that a patient is cured of risk i. We will be using a logistic model to estimate the cure probabilities for each competing event and do not assume independence of B_1 and B_2. For a fixed covariate value $x \in \mathcal{X}$, the joint distribution of the two variables B_1 and B_2 can be characterized by the following :

$$\begin{aligned} p_{0,0}(x) &= \mathbb{P}(B_1 = 0, B_2 = 0 \mid X = x), \\ p_{0,1}(x) &= \mathbb{P}(B_1 = 0, B_2 = 1 \mid X = x), \\ p_{1,0}(x) &= \mathbb{P}(B_1 = 1, B_2 = 0 \mid X = x), \\ p_{1,1}(x) &= \mathbb{P}(B_1 = 1, B_2 = 1 \mid X = x), \end{aligned} \tag{3.1}$$

with the property that $p_{0,0}(x) + p_{0,1}(x) + p_{1,0}(x) + p_{1,1}(x) = 1$ for all $x \in \mathcal{X}$. The quantities defined in (3.1) provide the cure chances given some covariates x. These can be modelled using logistic regression. To enhance the interpretability, we will model the cure status of the second risk conditional on the cure status of the first competing event. In this framework, it will be easier to see the dependence between two cure status random variables, e.g. if there is no dependence then $\pi_2^1 = \pi_2^0$. Define the following (conditional) cure probabilities:

$$\begin{aligned} \pi_1(x) &:= \mathbb{P}(B_1 = 1 \mid x) = \frac{e^{\gamma_1 x}}{1 + e^{\gamma_1 x}}, \\ \pi_2^1(x) &:= \mathbb{P}(B_2 = 1 \mid B_1 = 1, x) = \frac{e^{\gamma_2^1 x}}{1 + e^{\gamma_2^1 x}}, \\ \pi_2^0(x) &:= \mathbb{P}(B_2 = 1 \mid B_1 = 0, x) = \frac{e^{\gamma_2^0 x}}{1 + e^{\gamma_2^0 x}}, \end{aligned} \tag{3.2}$$

where x denotes the vector of covariates related to the incidence and the γ's the vectors of logistic regression coefficient – including an intercept. From (3.2) one can easily see how being cured of risk 1 affects being cured of risk 2. Hereinafter, the dependence of the probabilities defined in (3.2) on the covariates x will be omitted from the notation. The different cure structures that were introduced earlier in Section 2.7, can all be captured in terms of π defined in (3.2). In the case of *complete cure*, we have that case we have that $\pi_2^1 = 1$ and $\pi_2^0 = 0$. Since an individual is either cured of all risks or susceptible to all risks. If we consider the cure structure where one is cured of only the event of interest (Section 2.7.1), we have that $\pi_2^1 = \pi_2^0 = 0$. The question of recovering the cure structure thus boils down to identifying the (conditional) cure status probabilities. If the cure status probabilities are not identifiable, the cure structure cannot be recovered from the data.

The latency submodel will be modelled by a Cox proportional hazard (PH) model on the cause-specific hazards or subdistribution hazards. In the case of independent survival times, we will model the cause-specific hazards. Here the marginal survival and hazard functions coincide with the cause-specific counterparts. The survival time for event k can then be written as:

$$S_k(t \mid B_k = 0, x) = \exp\left\{-\int_0^t \lambda_k^0(u) \exp\{\beta_k^\top x\}\, du\right\}, \tag{3.3}$$

where β_k is the vector with regression coefficients for risk k of the Cox model and λ_k^0 is the baseline cause-specific hazard. This is the probability of surviving up to time t for event k for all individuals uncured of the respective event. The likelihood for this model consists of the product of the following quantities: $L_0(\theta), L_1(\theta)$ and $L_2(\theta)$ where θ denotes the set of all relevant parameters. $L_0(\theta), L_1(\theta)$ and $L_2(\theta)$ provide the contribution of a censored observation, an uncensored observation of type 1 and an uncensored observation of type 2 respectively.

The general likelihood contributions can be computed by conditioning on the cure status. The contribution of a censored observation is given by the probability that an individual survives up to the particular time t:

$$\begin{aligned}
L_0(\theta) &= \mathbb{P}(T > t) \\
&= \mathbb{P}(T > t \mid B_1 = 0, B_2 = 0, \theta)\mathbb{P}(B_1 = 0, B_2 = 0 \mid \theta) \\
&\quad + \mathbb{P}(T_1 > t \mid B_1 = 0, B_2 = 1, \theta)\mathbb{P}(B_1 = 0, B_2 = 1 \mid \theta) \\
&\quad + \mathbb{P}(T_2 > t \mid B_1 = 1, B_2 = 0, \theta)\mathbb{P}(B_1 = 1, B_2 = 0 \mid \theta) \\
&\quad + \mathbb{P}(B_1 = 1, B_2 = 1 \mid \theta) \qquad\qquad (3.4)\\
&= (1 - \pi_1)(1 - \pi_2^0) \cdot \mathbb{P}(T > t \mid B_1 = 0, B_2 = 0, \theta) \\
&\quad + (1 - \pi_1)\pi_2^0 \cdot \mathbb{P}(T_1 > t \mid B_1 = 0, B_2 = 1, \theta) \\
&\quad + \pi_1(1 - \pi_2^1) \cdot \mathbb{P}(T_2 > t \mid B_1 = 1, B_2 = 0, \theta) \\
&\quad + \pi_1\pi_2^1,
\end{aligned}$$

where T_1 and T_2 denote the potential survival times for, respectively, competing events 1 and 2.

The likelihood that an individual experiences an event at time t from event 1 is given by:

$$\begin{aligned}L_1(\theta) &= \mathbb{P}(T_1 = t \mid T_2 > t, B_1 = 0, B_2 = 0, \theta)\mathbb{P}(T_2 > t \mid B_1 = 0, B_2 = 0, \theta)\mathbb{P}(B_1 = 0, B_2 = 0 \mid \theta) \\ &\quad + \mathbb{P}(T_1 = t \mid B_1 = 0, B_2 = 1, \theta)\mathbb{P}(B_1 = 0, B_2 = 1 \mid \theta) \\ &= (1-\pi_1)(1-\pi_2^0) \cdot \mathbb{P}(T_1 = t \mid B_1 = 0, B_2 = 0, \theta)\mathbb{P}(T_2 > t \mid B_1 = 0, B_2 = 0, \theta) \\ &\quad + (1-\pi_1)\pi_2^0 \cdot \mathbb{P}(T_1 = t \mid B_1 = 0, B_2 = 1, \theta).\end{aligned} \tag{3.5}$$

A similar expression can be derived for the likelihood that an individual experiences an event of type 2 at time t. These are the general likelihood contributions. Under more specific models for the latency – as presented later in this chapter – these expressions will be specified. In the next section, we will consider a model where the potential survival times of the uncured individuals are assumed to be independently distributed.

3.2 Identifiability for independent survival times

In this section, we will conjecture – and partially prove – that under the assumption of independence of the potential survival times for the uncured, the cure structure is identifiable. The independence assumption means that if $B_1 = B_2 = 0$, then T_1 and T_2 are independent. As the proof is incomplete, part of the theorem is presented as a conjecture. The model introduced in this section will be coined the *competing risks cure* model. The next chapter will be devoted to the estimation of this model.

Consider a setting with $K = 2$ competing risks where the cure structure is a priori unknown. Furthermore, assume that the following holds:

(A1) (i) $\beta_1^\top x$ and $\beta_2^\top x$ do not contain an intercept.

(ii) The matrix $\mathrm{Var}(X)$ has full rank.

(A2) (i) The potential survival times for the uncured individuals T_1 and T_2 are independent.

(ii) Two different cure thresholds τ_1 and τ_2 exist for event 1 and 2 respectively such that $\tau_1 < \tau_2$ and:

$$\mathbb{P}(\tau_1 < T_1 < \infty) = \mathbb{P}(\tau_2 < T_2 < \infty) = 0 \quad \text{and} \quad \mathbb{P}(C > \tau_2) > 0.$$

Here **(A1)** ensures the identifiability of the Cox model while **(A2)** ensures the identifiability of the cure structure and the survival functions.

In Section 2.6.2 the existence of a cure threshold in the absence of competing risks was discussed. In the presence of competing risks, there are multiple cure thresholds whose existence needs to be validated from the data to ensure identifiability. In addition, the cure thresholds must be apart as stated in the assumption, that is $\tau_1 < \tau_2$. This implies the presence of some uncensored events between the cure thresholds τ_1 and τ_2.

For simplicity, we consider only one set of covariates $x \in \mathcal{X}$ for both the incidence and latency submodels. The proof can be easily extended to the general case where the covariates for the incidence and latency submodels differ.

For this model, there are again three different types of contributions to the likelihood:

$$\begin{aligned} L_0(\theta) &= \pi_1\pi_2^1 + \pi_1(1-\pi_2^1)S_2(t) + (1-\pi_1)\pi_2^0 S_1(t) + (1-\pi_1)(1-\pi_2^0)S(t), \\ L_1(\theta) &= f_1(t)\left\{(1-\pi_1)\pi_2^0 + (1-\pi_1)(1-\pi_2^0)S_2(t)\right\}, \\ L_2(\theta) &= f_2(t)\left\{\pi_1(1-\pi_2^1) + (1-\pi_1)(1-\pi_2^0)S_1(t)\right\}, \end{aligned} \tag{3.6}$$

where $S(t) = S_1(t)S_2(t)$.

As mentioned, the full statement is divided into two parts, as it is only partially proven. We start with a theorem for the identifiability of a subset of the parameters.

Theorem 1. *Under assumptions **(A1)-(A2)** the parameters of the logistic model related to π_1 and π_2^1, the coefficients of the Cox model related to the second event, i.e. β_2 and the baseline hazard of the second event on $(\tau_1, \tau_2]$ can be identified.*

Proof. Suppose that $\mathcal{L}(\theta) = \mathcal{L}(\tilde{\theta})$ almost surely and consider the contribution of the censored observations: $L_0(\theta) = L_0(\tilde{\theta})$. The contribution under this model is given in (3.6). Let $t > \tau_2$, then we only need to consider $\pi_1, \pi_2^1 > 0$. Since, if $\pi_1\pi_2^1 = 0$ holds almost surely, then for the different likelihoods to be equal almost everywhere, it must hold that $\tilde{\pi}_1\tilde{\pi}_2^1 = 0$. So we just need to consider $\pi_1, \pi_2^1 > 0$. Now, by assumption, the probability of surviving after t for any of the susceptible (to any competing event) individuals is zero. Hence, the contribution given in (3.6) reduces to $\pi_1\pi_2^1$. This implies that we can identify the fraction of individuals who are insusceptible to all events. Since $L_0(\theta) = L_0(\tilde{\theta})$ reduces to $\pi_1\pi_2^1 = \tilde{\pi}_1\tilde{\pi}_2^1$, where $\tilde{\pi}_1$ and $\tilde{\pi}_2^1$ refer to the respective cure probabilities induced by the logistic regression parameters from $\tilde{\theta}$.

Next, if we consider $t \in (\tau_1, \tau_2]$, equality of the likelihood contributions reduces to:

$$\begin{aligned} \pi_1\pi_2^1 + \pi_1(1-\pi_2^1)S_2(t) &= \tilde{\pi}_1\tilde{\pi}_2^1 + \tilde{\pi}_1(1-\tilde{\pi}_2^1)\tilde{S}_2(t) \\ &= \pi_1\pi_2^1 + \tilde{\pi}_1(1-\tilde{\pi}_2^1)\tilde{S}_2(t). \end{aligned} \tag{3.7}$$

This implies that $\pi_1(1-\pi_2^1)S_2(t) = \tilde{\pi}_1(1-\tilde{\pi}_2^1)\tilde{S}_2(t)$ for all $t \in (\tau_1, \tau_2]$. The dependence on the covariates was omitted from the notation, but clearly, both the survival and the cure status probabilities depend on the covariates. Equation (3.7) can be rewritten introducing the covariates x:

$$\frac{S_2(t|x)}{\tilde{S}_2(t|x)} = \frac{\tilde{\pi}_1(x)(1-\tilde{\pi}_2^1(x))}{\pi_1(x)(1-\pi_2^1(x))} =: c(x). \tag{3.8}$$

Since the fraction of cure probabilities is independent of t, the fraction of survival probabilities must also be constant with respect to t. Let $S_k^0(t)$ denote the baseline survival for event $k = 1, 2$. Then Equation (3.8) can be rewritten as:

$$\frac{S_2^0(t)^{\exp\{\beta_2 x\}}}{\tilde{S}_2^0(t)^{\exp\{\tilde{\beta}_2 x\}}} = c(x). \tag{3.9}$$

This implies that:

$$\exp\{\beta_2 x\}\log S_2^0(t) = \log c(x) + \exp\left\{\tilde{\beta}_2 x\right\}\log\tilde{S}_2^0(t). \tag{3.10}$$

As this equality holds almost everywhere, we can fix a specific value x' of x for which the equality holds. This value can be plugged into the equation above and then we can divide both quantities.

This yields:

$$\begin{aligned}\frac{\exp\{\beta_2 x\}}{\exp\{\beta_2 x'\}} &= \frac{\exp\{\beta_2 x\}\log S_2^0(t)}{\exp\{\beta_2 x'\}\log S_2^0(t)} \\ &= \frac{\log c(x) + \exp\left\{\tilde{\beta}_2 x\right\}\log \tilde{S}_2^0(t)}{\log c(x') + \exp\left\{\tilde{\beta}_2 x'\right\}\log \tilde{S}_2^0(t)}.\end{aligned} \tag{3.11}$$

This equation can be solved for $\log \tilde{S}_2^0(t)$ to find that:

$$\log \tilde{S}_2^0(t) \cdot \left[\exp\left\{\tilde{\beta}_2 x\right\} - \frac{\exp\{\beta_2 x\}}{\exp\{\beta_2 x'\}}\exp\left\{\tilde{\beta}_2 x'\right\}\right] = \left[\frac{\exp\{\beta_2 x\}}{\exp\{\beta_2 x'\}}\log c(x') - \log c(x)\right]. \tag{3.12}$$

Note that $S_2^0(t) \not\equiv c$ for some constant $c > 0$ for all $t \in (\tau_1, \tau_2]$ as $\lim_{t\to\tau_2} S_2^0(t) = 0$. The same holds for $\tilde{S}_2^0(t)$. This remark together with the fact that the right-hand side of Equation (3.12) does not depend on t, we can conclude that both expressions within the square brackets are equal to zero. This implies that:

$$\frac{\exp\left\{\tilde{\beta}_2 x\right\}}{\exp\left\{\tilde{\beta}_2 x'\right\}} = \frac{\exp\{\beta_2 x\}}{\exp\{\beta_2 x'\}} \quad \text{and} \quad \frac{\exp\{\beta_2 x\}}{\exp\{\beta_2 x'\}}\log c(x') = \log c(x). \tag{3.13}$$

From the first equation, it can be derived that:

$$\exp\left\{\tilde{\beta}_2(x - x')\right\} = \exp\{\beta_2(x - x')\} \implies \tilde{\beta}_2(x - x') = \beta_2(x - x'). \tag{3.14}$$

This holds for almost every $x \in \mathcal{X}$. Together with the assumption that $\mathrm{Var}(X)$ has full rank, we can conclude that $\beta_2 = \tilde{\beta}_2$. This equality can be plugged into Equation (3.9) to find that:

$$\log \frac{S_2^0(t)}{\tilde{S}_2^0(t)} = \frac{\log c(x)}{\exp\{\beta_2 x\}}. \tag{3.15}$$

As the left-hand side of the equation depends only on the time t and the right-hand side only on the covariates x, it is implied that both are equal to some constant η. Unfortunately, it is not evident that $\eta = 0$ as this would imply that $S_2^0(t) = \tilde{S}_2^0(t)$ for all $t \in (\tau_1, \tau_2]$. It can be shown that S_2^0 and $\tilde{S}_2^0$ are the same on $(\tau_1, \tau_2]$. This yields partial identifiability of the baseline survival for the second event. Consider the likelihood contribution of an event of type 2 observed in $(\tau_1, \tau_2]$. Equality of the likelihoods on this interval yields:

$$\pi_1(1 - \pi_2^1)f_2(t) = \tilde{\pi}_1(1 - \tilde{\pi}_2^1)\tilde{f}_2(t). \tag{3.16}$$

This equation can be rewritten as:

$$\frac{f_2(t|x)}{\tilde{f}_2(t|x)} = \frac{\tilde{\pi}_1(x)(1 - \tilde{\pi}_2^1(x))}{\pi_1(x)(1 - \pi_2^1(x))} = c(x). \tag{3.17}$$

where $c(x)$ is the same as in Equation (3.8). As a consequence:

$$\frac{f_2(t|x)}{\tilde{f}_2(t|x)} = \frac{S_2(t|x)}{\tilde{S}_2(t|x)}. \tag{3.18}$$

Equation (2.12) yields the following:

$$\frac{\lambda_2(t|x)S_2(t|x)}{\tilde{\lambda}_2(t|x)\tilde{S}_2(t|x)} = \frac{S_2(t|x)}{\tilde{S}_2(t|x)} \implies \frac{\lambda_2(t|x)}{\tilde{\lambda}_2(t|x)} = 1 \quad a.e. \tag{3.19}$$

This shows that $\lambda_2(t|x) = \tilde{\lambda}_2(t|x)$ almost everywhere for $t \in (\tau_1, \tau_2]$. Since it also holds for almost every $x \in \mathcal{X}$, it identifies the baseline hazard for event 2 on the interval $(\tau_1, \tau_2]$.

Furthermore, it shows that $c(x) = 1$ almost everywhere. Equation (3.17) implies that $\tilde{\pi}_1(x)(1 - \tilde{\pi}_2^1(x)) = \tilde{\pi}_1(x)(1 - \tilde{\pi}_2^1(x))$ almost everywhere. Combining this with the fact that $\pi_1\pi_2^1 = \tilde{\pi}_1\tilde{\pi}_2^1$ a.e. shows that both π_1 and π_2^1 are identified. □

Remark. *The proof of Theorem 1 silently depends on Lemma 4 introduced in Section 3.2.1.*

From Theorem 1, it can be concluded that a subset of the parameters is identifiable. Full identifiability would e.g. follow from identification of $\lambda_2^0(t)$ on $(0, \tau_2]$.

Lemma 2. *Under the same assumptions as in Theorem 1, if λ_2^0 were identified, then all parameters of the model would be identifiable.*

Proof. Suppose that $\mathcal{L}(\theta) = \mathcal{L}(\tilde{\theta})$ almost surely. It follows (from Theorem 1 and the additional assumption) that λ_2^0, β_2, π_1 and π_2^1 have been identified. So it remains to show that we can identify λ_1^0, β_1 and π_2^0.

Contributions of an uncensored observation of event 2 on $[t, \tau_1)$ given in (3.6) can be written as:

$$\begin{aligned} f_2(t)\left\{\pi_1(1-\pi_2^1) + (1-\pi_1)(1-\pi_2^0)S_1(t)\right\} &= \tilde{f}_2(t)\left\{\tilde{\pi}_1(1-\tilde{\pi}_2^1) + (1-\tilde{\pi}_1)(1-\tilde{\pi}_2^0)\tilde{S}_1(t)\right\} \\ &= f_2(t)\left\{\pi_1(1-\pi_2^1) + (1-\pi_1)(1-\tilde{\pi}_2^0)\tilde{S}_1(t)\right\}. \end{aligned} \tag{3.20}$$

Simplification and rewriting yields the following equation:

$$\frac{S_1(t|x)}{\tilde{S}_1(t|x)} = \frac{(1-\pi_1(x))(1-\tilde{\pi}_2^0)}{(1-\pi_1(x))(1-\pi_2^0)} =: d(x). \tag{3.21}$$

Now the proof of identifiability of S_2 from the proof of Theorem 1 can be followed. These are explained in steps (3.8) – (3.15). It follows that λ_1^0 and β_1 are identified and it can be shown that $d(x) = 1$ almost surely, and thus yields identifiability of π_2^0 by Equation (3.21). □

Imposing proper parametric assumptions on the potential survival times would, according to this lemma, be sufficient to identify the survival times, and from there, the cure structure. We will not impose any parametric restriction on the baseline hazards. Therefore, for the remaining parameters, we transform our claim into a conjecture.

Conjecture 3. *Suppose that $\tau_1 < \tau_2$ and $X \in \mathcal{X}$ contains a continuous covariate, then the model is identifiable.*

To justify the use of only continuous covariates in Conjecture 3 we could construct a counterexample for a binary covariate. From (3.20) and $t \to 0$ follows:

$$c := \frac{(1-\pi_2^1(x)) \cdot \frac{\pi_1(x)}{1-\pi_1(x)} + (1-\pi_2^0(x))}{(1-\pi_2^1(x)) \cdot \frac{\pi_1(x)}{1-\pi_1(x)} + (1-\tilde{\pi}_2^0(x))} = \frac{\psi(x)+1-\pi_2^0(x)}{\psi(x)+1-\tilde{\pi}_2^0(x)} \quad a.e \tag{3.22}$$

for some constant $c \in \mathbb{R}$ and a function $\psi(x) = (1-\pi_2^1(x)) \cdot \pi_1(x)/(1-\pi_1(x))$. Recall that the cure probabilities are modelled by a logistic regression model. So we can write:

$$\psi(x) = \frac{e^{\gamma_1 x}}{1+e^{\gamma_2^1 x}}, \quad 1-\pi_2^0(x) = \frac{1}{1+e^{\gamma_2^0 x}} \quad \text{and} \quad 1-\tilde{\pi}_2^0(x) = \frac{1}{1+e^{\tilde{\gamma}_2^0 x}}. \tag{3.23}$$

Equation (3.22) holds for almost every $x \in \mathcal{X}$. For a binary covariate, we can construct a counterexample where it holds almost everywhere but $\gamma_2^0 \neq \tilde{\gamma}_2^0$. Though Equation (3.22) does not seem satisfiable if $\pi_2^0(x) \neq \tilde{\pi}_2^0(x)$ a.s. if x is a continuous covariate. By taking the derivative of the right-hand side of Equation (3.22), setting it equal to zero and solving it for π_2^0, may lead to the conclusion that $\pi_2^0(x) = \tilde{\pi}_2^0(x)$ almost surely. Unfortunately, due to the logistic form of the π's, equations are lengthy and hard to handle. Identification of π_2^0, similar to Lemma 2, leads to the identifiability of all parameters in the model. For the time being, we can only conclude that more research is necessary to round up the general proof.

In the next two subsections, identifiability is proven for two specific choices of cure structures: complete cure and cure for only the event of interest. Both cure structures have been outlined in Section 2.7.

3.2.1 The case of complete cure

First, we consider the case of complete cure. This cure structure has also been used by Zhang et al. (2019). We have discussed the models illustrated in the article in Section 2.7.3. Model D considers the case of complete cure while Model B and C coincide with the perspective that cure can only happen for the event of interest. The authors assumed independence of the potential survival times but left the problem of identifiability open. We will show that the parameters are indeed identifiable.

In case of complete cure, we only need a single cure status indicator for all competing events simultaneously:

$$\pi(x) := \mathbb{P}(B_1 = B_2 = 1 \mid x) = \mathbb{P}(B = 1 \mid x), \tag{3.24}$$

where B is the random variable denoting whether an individual is immune to all competing events and x are the covariates related to the incidence. Furthermore, we assume that the following conditions hold:

(B1) (i) $\beta_1^\top x$ and $\beta_2^\top x$ do not contain an intercept.

(ii) The matrix $\mathrm{Var}(X)$ has full rank.

(B2) (i) There exists a threshold $\tau < \infty$ such that $\mathbb{P}(\tau < T_k < \infty) = 0$ for all $t \geq \tau$ and $k = 1, 2$, and $\mathbb{P}(C > \tau \mid x) \in (0, 1)$.

(ii) There e xists a proper cure fraction $\pi(x) \in (0, 1)$ for all $x \in \mathcal{X}$.

In order to prove the next theorem, we must first prove a lemma. To show that the model with complete cure and independent survival times is identifiable, several suitably chosen subsets of the sample space are chosen. The parameters of the model can be identified over these subsets which must have positive measures.

Lemma 4. *The following subsets of the sample space have positive measure:* $(T > \tau, \delta = 0)$*,* $(T \leq \tau, \delta = 0)$ *and* $(T \leq \tau, \delta = 1)$*.*

Proof. First, consider the case where $T > \tau$ and $\delta = 0$. Note that:

$$\begin{aligned} \mathbb{P}(T > \tau, \delta = 0 \mid x) &= \mathbb{P}(C > \tau, B = 1 \mid x) \\ &= \pi(x)\mathbb{P}(C > \tau \mid x) \\ &> 0. \end{aligned} \tag{3.25}$$

Suppose that $T \leq \tau$ and $\delta = 0$. Then we have that:

$$\begin{aligned} \mathbb{P}(T \leq \tau, \delta = 0 \mid x) &= \mathbb{P}(C \leq \tau, T_1 > C, T_2 > C, ..., T_K > C \mid x) \\ &\geq \mathbb{P}(C \leq \tau, B = 1 \mid x) \\ &= \pi(x)\mathbb{P}(C \leq \tau \mid x) \\ &> 0. \end{aligned} \tag{3.26}$$

The inequality in the second line holds since:

$$\{B = 1\} = \{T_1 = T_2 = ... = T_K = \infty\} \subset \{T_1 > C, T_2 > C, ..., T_K > C\}.$$

Now consider $T \leq \tau$ and $\delta = 1$ and denote $M := \min\{T_1, T_2, ..., T_K\}$. Then we can compute:

$$\begin{aligned} \mathbb{P}(T \leq \tau, \delta = 1 \mid x) &= \mathbb{P}(M \leq \tau, M \leq C \mid x) \\ &\geq \mathbb{P}(M \leq \tau, C \geq \tau \mid x) \\ &= \mathbb{P}(M \leq \tau \mid x)\mathbb{P}(C \geq \tau \mid x) \\ &= (1 - \pi(x))\mathbb{P}(C \geq \tau \mid x) \\ &> 0. \end{aligned} \tag{3.27}$$

The last equality follows from the fact that if one of the potential survival times is finite, the individual will be cured. This can be seen here:

$$\begin{aligned} \mathbb{P}(M \leq \tau \mid x) &= \mathbb{P}(\exists\ i : T_i \leq \tau) \\ &= 1 - \mathbb{P}(\forall\ i : T_i > \tau) \\ &= 1 - \mathbb{P}(T_1 = T_2 = ... = T_K = \infty) \\ &= 1 - \pi(x). \end{aligned} \tag{3.28}$$

□

This shows that all three of the chosen subsets of the sample space have positive measures. In the proof, independence of potential survival times was not used, although it was assumed in this section. Therefore, the lemma holds in a general setting where the survival times are not independent.

For the case of complete cure, the cure statuses are given by $\pi_2^0 = 0$ and $\pi_2^1 = 1$. Later, it will be shown that in a setting of complete cure, the cause-specific hazards and cure fraction are always identifiable also without assuming independence of the potential survival times. This will be proven in Section 3.4.1. Note that the cause-specific hazards equal the normal hazard rates if the survival times are independent. Identifiability in this setting may follow as a consequence. A formal proof of the theorem will be provided. The likelihood contributions given in (3.6) reduce for this (nested) model to:

$$\begin{aligned} L_0(\theta) &= \pi + (1-\pi)S(t), \\ L_1(\theta) &= (1-\pi)f_1(t)S_2(t), \\ L_2(\theta) &= (1-\pi)f_2(t)S_1(t). \end{aligned} \tag{3.29}$$

Theorem 5. *If $\pi_2^0 = 0$ and $\pi_2^1 = 1$, the survival times are independent and a cure threshold τ for both events exists, the model is identifiable.*

Proof. Assume that $\mathcal{L}(\theta) = \mathcal{L}(\tilde{\theta})$ holds almost surely. It in particular holds that $L_0(\theta) = L_0(\tilde{\theta})$, that is:

$$\pi + (1-\pi)S(t) = \tilde{\pi} + (1-\tilde{\pi})\tilde{S}(t). \tag{3.30}$$

Since **(B2)** (i) implies that $S(t) = S_1(t)S_2(t) = 0$ for all $t \geq \tau$, equation (3.30) reduces to $\pi = \tilde{\pi}$. This subset of the sample space can be considered due to Lemma 4. By identifiability of the logistic model, this shows that the cure fraction has been identified.

Now we consider equality of the likelihood contributions $L_1(\theta) = L_1(\tilde{\theta})$ and $L_1(\theta) = L_1(\tilde{\theta})$ a.s. Since π has been identified, $1-\pi$ is known as well. The equalities of these likelihood contributions imply, respectively, that:

$$f_1(t)S_2(t) = \tilde{f}_1(t)\tilde{S}_2(t) \quad \text{and} \quad f_2(t)S_1(t) = \tilde{f}_2(t)\tilde{S}_2(t). \tag{3.31}$$

Both equality's can be rewritten and identity (2.4) applied, to find that:

$$\frac{f_1(t)}{\tilde{f}_1(t)} = \frac{\tilde{S}_2(t)}{S_2(t)} = \frac{\tilde{\lambda}_2(t)\tilde{f}_2(t)}{\lambda_2(t)f_2(t)} \quad \text{and} \quad \frac{f_2(t)}{\tilde{f}_2(t)} = \frac{\tilde{S}_1(t)}{S_1(t)} = \frac{\tilde{\lambda}_1(t)\tilde{f}_1(t)}{\lambda_1(t)f_1(t)}. \tag{3.32}$$

Mutual substitution yields:

$$\frac{\tilde{\lambda}_1(t)}{\lambda_1(t)} = 1 \quad \text{and} \quad \frac{\tilde{\lambda}_2(t)}{\lambda_2(t)} = 1. \tag{3.33}$$

It follows that both $\lambda_1(t) = \tilde{\lambda}_1(t)$ and $\lambda_2(t) = \tilde{\lambda}_2(t)$ almost surely. By identifiability of the Cox model, the hazards for both events have been identified. Therefore, the whole model is identifiable. □

Therefore, in the case of complete cure, the model is identifiable. In the next section, we discuss the case when cure happens only to the event of interest.

3.2.2 Cure only for event of interest

If we have prior knowledge that cure only happens for the event of interest, we can also identify the survival times and cure fraction. Let event 1 be the event of interest. As noted, this case induces the cure statuses to equal $\pi_2^0 = 0$ and $\pi_2^1 = 0$. The likelihood contributions are given by:

$$\begin{aligned} L_0(\theta) &= S_2(t)\big\{\pi_1 + (1-\pi_1)S_1(t)\big\}, \\ L_1(\theta) &= (1-\pi_1)f_1(t)S_2(t), \\ L_2(\theta) &= f_2(t)\big\{\pi_1 + (1-\pi_1)S_1(t)\big\}. \end{aligned} \tag{3.34}$$

Previously, we considered two cure thresholds τ_1 and τ_2. In this setup, a finite τ_2 does not exist, as one is never cured of the second competing event. We thus make the following assumptions:

(C1) (i) $\beta_1^\top x$ and $\beta_2^\top x$ do not contain an intercept.

(ii) The matrix $\text{Var}(X)$ has full rank.

(C2) (i) There exists a $\tau_1 < \infty$ such that $\mathbb{P}(\tau_1 < T_1 < \infty) = 0$ and $\mathbb{P}(C > \tau_1 \mid x) \in (0,1)$.

(ii) There exists a proper cure fraction $\pi_1(x) \in (0,1)$ for all $x \in \mathcal{X}$.

Theorem 6. *If $\pi_2^0 = 0$ and $\pi_2^1 = 0$, the survival times are independent and assumptions **(C1)** and **(C2)** hold, then the model is identifiable.*

Proof. Assume that $\mathcal{L}(\theta) = \mathcal{L}(\tilde{\theta})$ holds almost surely. It follows from the first and third equality of (3.34) that:

$$\frac{S_2(t|x)}{\tilde{S}_2(t|x)} = \frac{\tilde{\pi}_1(x) + (1-\tilde{\pi}_1(x))\tilde{S}_1(t|x)}{\pi_1(x) + (1-\pi_1(x))S_1(t|x)} \quad \text{and} \quad \frac{f_2(t|x)}{\tilde{f}_2(t|x)} = \frac{\tilde{\pi}_1(x) + (1-\tilde{\pi}_1(x))\tilde{S}_1(t|x)}{\pi_1(x) + (1-\pi_1(x))S_1(t|x)}. \tag{3.35}$$

As a consequence,

$$\begin{aligned} \frac{S_2(t|x)}{\tilde{S}_2(t|x)} &= \frac{f_2(t|x)}{\tilde{f}_2(t|x)} \\ &= \frac{\lambda_2(t|x)}{\tilde{\lambda}_2(t|x)}\frac{S_2(t|x)}{\tilde{S}_2(t|x)}. \end{aligned} \tag{3.36}$$

This shows that $\lambda(t|x) = \tilde{\lambda}(t|x)$ almost surely. Together with the identifiability of the Cox model, it identifies both $\lambda_2^0(t)$ and β_2, implying that $f_2(t|x)$ has been identified. So from the likelihood equality of an observed event of type 2, we find that:

$$\pi_1(x) + (1-\pi_1(x))S_1(t|x) = \tilde{\pi}_1(x) + (1-\tilde{\pi}_1(x))\tilde{S}_1(t|x). \tag{3.37}$$

Now let $t \to \tau_1$, then we have that $\lim_{t\to\tau_1} S_1(t|x) = \lim_{t\to\tau_1} \tilde{S}_1(t|x) = 0$. As a consequence, Equation (3.37) reduces to $\pi_1(x) = \tilde{\pi}_1(x)$ almost surely, i.e. the cure fraction has been identified.

Now it follows from:

$$(1-\pi_1(x))f_1(t|x)S_2(t|x) = (1-\pi_1(x))\tilde{f}_1(t|x)S_2(t|x) \implies f_1(t|x) = \tilde{f}_1(t|x), \tag{3.38}$$

that also $f_1(t|x)$ has been identified. This implies that $\lambda_1^0(t)$ and β_1 are also identified by the identifiability of the Cox model. □

3.3 Non-identifiability in a general setting

In this section, we will show that the cure model – including the cure structure – is not identifiable in a setting with less restrictive assumptions. Under *weaker* assumptions, we can find a proxy model for which the likelihood resembles the one from the original model. This proxy model is given by a cure model with a complete cure structure and independent survival times. As mentioned before, there is a general non-identifiability problem when competing risks are present (Tsiatis (1975)). Our results are similar to Tsiatis (1975): without cure, the proxy models coincide.

Consider two competing risks for which an individual can be cured separately. To each competing risk a potential survival time T_1 and T_2 can be associated. These are the survival times for those who are only susceptible to their respective risks. As noted in Section (2.5.1) the potential survival times T_1 and T_2 are not observed. Only the actual survival time $T = \min(T_1, T_2)$ is observed in the case of an uncensored observation.

Since the potential survival times are not independent, we cannot model the hazard directly. We will therefore model the cause-specific hazards conditional on the cure status. The cause-specific hazard (and other related functions) need to be redefined in the context of cure. For example, we can define a hazard function on the whole population, but it is more interesting to restrict it to the uncured subpopulation. These definitions are the same as in Nicolaie et al. (2019). First, define the conditional (on the cure statuses) total hazard as follows:

$$\lambda_\bullet(t) := \lim_{dt\to 0} \frac{\mathbb{P}(t \leq T < t + dt \mid T \geq t, B_1 = B_2 = 0)}{dt} \tag{3.39}$$

This is the hazard rate for all susceptible individuals. The immune individuals would never experience the event, implying a constant hazard of zero. This would not properly define a survival function. Furthermore, we introduce the conditional cause-specific hazard rate:

$$\lambda_k^*(t) := \lim_{dt\to 0} \frac{\mathbb{P}(t \leq T < t + dt, D = k \mid T \geq t, B_k = 0)}{dt}. \tag{3.40}$$

Using (3.40), we can define the function:

$$S_k^*(t) := \exp\left\{-\int_0^t \lambda_k^*(u)\ du\right\}. \tag{3.41}$$

We emphasize that this function – in general – does not define a proper survival function. It only does so under the assumption that the competing risks have independent follow-up. Furthermore, we can define a corresponding density as:

$$f_k^*(t) := \lambda_k^*(t) S_k^*(t). \tag{3.42}$$

The likelihood contributions are then given by:

$$\begin{aligned}
L_1(\theta) &= f_1^*(t)\left\{(1-\pi_1)\pi_2^0 + (1-\pi_1)(1-\pi_2^0)S_2^*(t)\right\}, \\
L_2(\theta) &= f_2^*(t)\left\{\pi_1(1-\pi_2^1) + (1-\pi_1)(1-\pi_2^0)S_1^*(t)\right\}, \\
L_0(\theta) &= \pi_1\pi_2^1 + \pi_1(1-\pi_2^1)S_2^*(t) + (1-\pi_1)\pi_2^0 S_1^*(t) + (1-\pi_1)(1-\pi_2^0)S_1^*(t)S_2^*(t).
\end{aligned} \tag{3.43}$$

These expressions for the likelihood look similar to the ones from Section 3.2, but they do not have the same interpretation. In the presence of competing risks, $S^*(t)$ does not have a survivor function interpretation. This is explained in Section 2.5. More details can be found in (Putter et al., 2007).

Furthermore, it is assumed that there exist cure thresholds $\tau_1, \tau_2 < \infty$ such that $\mathbb{P}(T_1 > \tau_1) = \mathbb{P}(T_2 > \tau_2) = 0$. This implies that:

$$\begin{aligned}
S_1^*(t) &= 0 \quad \text{for all} \quad t > \tau_1, \\
S_2^*(t) &= 0 \quad \text{for all} \quad t > \tau_2.
\end{aligned} \tag{3.44}$$

Theorem 7 (Non-identifiability of the general cure structure)**.** *If the cure thresholds coincide ($\tau_1 = \tau_2$) and both the cure probabilities and the cause-specific hazards are independent of the covariates or are left unspecified (completely non-parametric), then it is not possible to:*

(a) identify the parameters related to the potential survival times;

(b) identify the parameters of the logistic model related to the cure fractions.

Proof. To show that the model is not identifiable, we will construct a proxy model with net survival functions $\tilde{S}_1$ and $\tilde{S}_2$ and cure structure probabilities $\tilde{\pi}_1$, $\tilde{\pi}_2^0$ and $\tilde{\pi}_2^1$ for which the likelihood resembles the likelihood of the original model. In particular, these survival functions define two independent follow-up times and the cure structure probabilities are given by the complete cure structure.

As we took the potential survival times in the proxy model to be independent, we can write:

$$\tilde{S}(t) = \tilde{S}_1(t)\tilde{S}_2(t) = \exp\left\{-\int_0^t \tilde{\lambda}_1^*(u) + \tilde{\lambda}_2(u)\, du\right\}, \tag{3.45}$$

where $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ denote the cause-specific hazard rates in the proxy model. As the potential survival times are independent, the cause-specific hazards equal the regular hazards.

For the proxy model to have a complete cure structure, the cure structure variables are chosen to equal:

$$\begin{cases}
\tilde{\pi}_1 = \pi_1\pi_2^1, \\
\tilde{\pi}_2^0 = 0, \\
\tilde{\pi}_2^1 = 1.
\end{cases} \tag{3.46}$$

This ensures on the one hand that the proxy model has a complete cure structure, and on the other hand that the fraction of people never experiencing any event is equal in both models.

We will now derive the hazards for which the likelihoods in the proxy and original model resemble each other. The cure statuses in Equation (3.46) imply that the likelihood contribution of the uncensored observations in the proxy model is given by:

$$L_0(\tilde{\theta}) = \pi_1\pi_2^1 + (1-\pi_1\pi_2^1)\tilde{S}(t). \tag{3.47}$$

Assume that the likelihood contributions of the censored observations are equal, that is, $L_0(\theta) = L_0(\tilde{\theta})$. Then it holds that:

$$(1-\pi_1)(1-\pi_2^0)S_1^*(t)S_2^*(t)+(1-\pi_1)\pi_2^0S_1^*(t)+\pi_1(1-\pi_2^1)S_2^*(t)+\pi_1\pi_2^1 = \pi_1\pi_2^1+(1-\pi_1\pi_2^1)\tilde{S}(t). \tag{3.48}$$

From this equality, we find that the survival function in the proxy model must be given by:

$$\tilde{S}(t) = \frac{1}{1-\pi_1\pi_2^1}\Big((1-\pi_1)(1-\pi_2^0)S_1^*(t)S_2^*(t) + (1-\pi_1)\pi_2^0S_1^*(t) + \pi_1(1-\pi_2^1)S_2^*(t)\Big). \tag{3.49}$$

Since this survival is given by the formula in (3.45), we can take a logarithm and differentiate both sides of the equation to find that:

$$\begin{aligned}\tilde{\lambda}_1(t)+\tilde{\lambda}_2(t) &= \frac{\partial}{\partial t}\left[-\log\left\{\frac{1}{1-\pi_1\pi_2^1}\big((1-\pi_1)(1-\pi_2^0)S_1^*(t)S_2^*(t) + (1-\pi_1)\pi_2^0S_1^*(t) + \pi_1(1-\pi_2^1)S_2^*(t)\big)\right\}\right] \\ &= \frac{f_1^*(t)\left\{(1-\pi_1)\pi_2^0 + (1-\pi_1)(1-\pi_2^0)S_2^*(t)\right\} + f_2^*(t)\left\{\pi_1(1-\pi_2^1) + (1-\pi_1)(1-\pi_2^0)S_1^*(t)\right\}}{(1-\pi_1)(1-\pi_2^0)S_1^*(t)S_2^*(t) + (1-\pi_1)\pi_2^0S_1^*(t) + \pi_1(1-\pi_2^1)S_2^*(t)} \\ &= \frac{\mathbb{P}(T=t,D=1)+\mathbb{P}(T=t,D=2)}{\mathbb{P}(t<T<\infty)}.\end{aligned} \tag{3.50}$$

A natural choice for the hazards in the proxy model is as follows:

$$\tilde{\lambda}_1(t) = \frac{\mathbb{P}(T=t,D=1))}{\mathbb{P}(t\le T<\infty)} \quad \text{and} \quad \tilde{\lambda}_2(t) = \frac{\mathbb{P}(T=t,D=2)}{\mathbb{P}(t\le T<\infty)}. \tag{3.51}$$

For this choice of hazards, the likelihoods of the uncensored observation are equal as well. It can be seen that:

$$\begin{aligned}L_1(\tilde{\theta}) &= (1-\pi_1\pi_2^1)\tilde{S}(t)\cdot\tilde{\lambda}_1(t) \\ &\overset{*}{=} \mathbb{P}(t<T<\infty)\cdot\frac{\mathbb{P}(T=t,D=1)}{\mathbb{P}(t<T<\infty)} \\ &= \mathbb{P}(T=t,D=1) \\ &= L_1(\theta).\end{aligned} \tag{3.52}$$

The equality denoted with $*$ follows from Equation (3.49). Analogously, we can show that for this choice of hazards, it also holds that $L_2(\theta) = L_2(\tilde{\theta})$. □

Remark. *Although the theorem states that under general assumptions the cure structure cannot be identified, it is possible to identify the proportion of individuals insusceptible to any of the events. This fraction has the same interpretation under the different models.*

This theorem shows that for a model with competing risks and a general cure structure, we can always construct a model with independent competing risks and a complete cure structure. We will illustrate this by extending an example from Tsiatis (1975) and Crowder (2012). Consider two competing risks distributed according to Gumbel's first bivariate exponential distribution (Gumbel, 1960). The joint survivor function of T_1 and T_2 is given by:

$$S(t_1, t_2) = \exp\{-\lambda_1 t_1 - \lambda_2 t_2 - \theta t_1 t_2\}, \tag{3.53}$$

where $\lambda_1, \lambda_2 > 0$ and $0 \leq \theta \leq \lambda_1\lambda_2$. Furthermore, let the cure probabilities be equal: $\pi_1 = \pi_2^0 = \pi_2^1 = \gamma \in (0,1)$. The cause-specific hazards for this distribution are given by:

$$\lambda_1(t) = \lambda_1 + \theta t \quad \text{and} \quad \lambda_2(t) = \lambda_2 + \theta t. \tag{3.54}$$

The marginal survival function for event 1 is given by $S_1(t) = \exp\{-\lambda_1 t\}$. In the proxy model, this survival function has a different form. It is given by $\tilde{S}_1(t) = \exp\big\{-\lambda_1 t - \frac{1}{2}\theta t^2\big\}$. This function already highlights the difference between the two models. Now we also include an example which utilizes the cure probabilities. Consider the following function:

$$p(t) := \mathbb{P}(T > t, B_1 = 1). \tag{3.55}$$

In the original model, this function can be expressed as follows:

$$\begin{aligned} p(t) &= \pi_1(1-\pi_2^1)\mathbb{P}(T > t \mid B_1 = 1, B_2 = 0) + \pi_1\pi_2^1\mathbb{P}(T > t \mid B_1 = 1, B_2 = 1) \\ &= \pi_1(1-\pi_2^1)\mathbb{P}(T_2 > t) + \pi_1\pi_2^1 \\ &= \gamma(1-\gamma)S_2(t) + \gamma^2. \end{aligned} \tag{3.56}$$

Contrary, in the proxy model, this function can be expressed as follows:

$$p(t) = \pi_1\pi_2^1 = \gamma^2. \tag{3.57}$$

The differences in $S_1(t)$ and $p(t)$ in the original and proxy model are highlighted in Figure 3.1. The choices of parameters are as follows $\gamma = 1/4, \lambda_1 = \lambda_2 = 1$ and $\theta = 0.25, 0.75$.

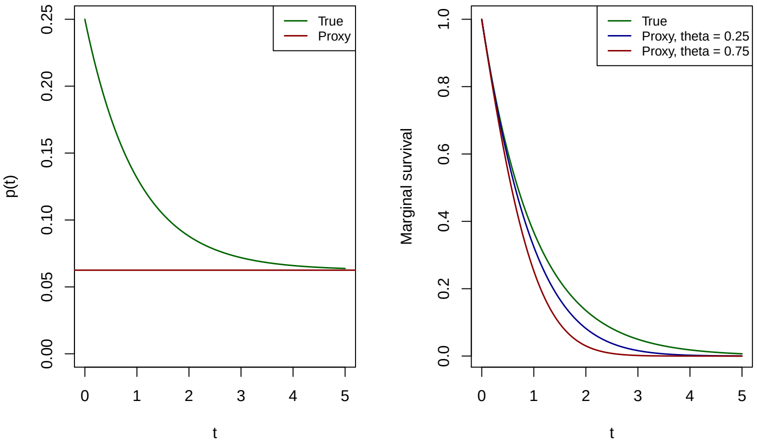

Figure 3.1: Plot of the function $p(t)$ and the different marginal survival functions ($\theta = 0.25, 0.75$).

According to Theorem 7, it is in general not possible to recover the cure structure and survival functions from the data. This has strong practical implications as we have seen through the example above. For instance, it is imperative that one has prior knowledge of the cure structure. If not, the practitioner has to make assumptions about the dependence structure of the survival times, which is not realistic in many situations. We have already seen that assuming independence of potential survival times suffices to recover the cure structure from the data. One can also make assumptions about the cure structure for the model to be identifiable. In the next section, we will treat several examples of complete cure models.

3.4 Identifiability for complete cure

In this section, we will show that the complete cure structure is identifiable without imposing independence constraints on the potential survival times. Contrary to the previous parts of this chapter, we do assume a certain form of the cure structure. This is motivated by practical reasons. For a lot of applications of the cure model in the presence of competing risks, there is sufficient prior knowledge that the cure structure is complete, i.e. cure happens for all competing events simultaneously. If we consider osteosarcoma – a malignant bone tumour – there is clinical knowledge or relevance indicating complete cure. The competing events for osteosarcoma are – among others – secondary malignancy, recurrence of the osteosarcoma and death due to the tumour. It is presumed that cure means that a patient does not experience any of these events. It, therefore, makes sense to model the cure structure as complete and not wonder about recovering the cure structure from the data.

We will show that assuming complete cure leads to identifiability if we model the cause-specific hazards and the subdistribution hazards. As mentioned before, these quantities are observable from the data and do not presume independence of the potential survival times Putter et al. (2007). Furthermore, we will show that the Vertical model (Nicolaie et al., 2019) is identifiable under suitable conditions.

Consider now the general case of K competing events. We pose no conditions on the dependence structure of the potential survival times $T_1, T_2, ..., T_K$. In case of complete cure, we only need to consider one cure threshold, as explained in Section 3.2.1. The following conditions are assumed to hold:

(D1) (i) $\beta_k^\top x$ does not contain an intercept for $k = 1, 2, ..., K$.

(ii) The matrix $\mathrm{Var}(X)$ has full rank.

(D2) (i) There exists one $\tau < \infty$ such that $\mathbb{P}(\tau < T_k < \infty) = 0$ for all $k = 1, 2, ..., K$ and $\mathbb{P}(C > \tau \mid x) \in (0, 1)$.

(ii) There exists a proper cure fraction, that is, $\pi(x) \in (0, 1)$ for all $x \in \mathcal{X}$.

The assumptions posed under **(D1)** ensure identifiability of the Cox model. Furthermore, the existence of a cure threshold τ is assumed. Similar to the standard case without competing risks, there is only one cure threshold.

3.4.1 Identifiability of the cause-specific hazard for complete cure

We consider cure in the competing risk setting and model the cause-specific hazard. The contribution of an individual who experienced an event of type k is as follows:

$$\begin{aligned} L_k(\theta) &= (1-\pi(x))\frac{\partial}{\partial u} I_k(u)|_t \\ &= (1-\pi(x))\lambda_k(t|x)S(t|x) \\ &= (1-\pi(x))\lambda_k(t\ |x))\exp\left\{-\sum_{k=1}^{K}\int_0^t \lambda_k(u|x)du\right\}. \end{aligned} \tag{3.58}$$

Since a cured individual will never experience any of the competing events, the likelihood contribution of a censored observation is given by:

$$L_0(\theta) = \mathbb{P}(T > t) = \pi(x) + (1-\pi(x))\exp\left\{-\sum_{k=1}^{K}\int_0^t \lambda_k(u|x)du\right\}. \tag{3.59}$$

Theorem 8. *Under assumptions **(D1)-(D2)** the parameters related to the cure fraction and the cause-specific hazards are identifiable.*

Proof. Assume that $\mathcal{L}(\theta) = \mathcal{L}(\tilde{\theta})$. By Lemma 4, we can consider the equality of the likelihoods over these three subsets separately. First, we consider $t > \tau$, then equality of the likelihood contribution of a censored observation L_0 is reduced to:

$$\pi(x) = \tilde{\pi}(x). \tag{3.60}$$

Since $\exp\left\{-\sum_{k=1}^{K}\Lambda_k(t|x)\right\} = 0$ by assumption **(B2)**. This implies that the cure fraction is identified as the logistic model is identifiable. Now consider the same contribution but for $T \leq \tau$:

$$\pi(x) + (1-\pi(x))S(t|x) = \tilde{\pi}(x) + (1-\tilde{\pi}(x))\tilde{S}(t|x). \tag{3.61}$$

Since π has been identified, this reduces to $S(t|x) = \tilde{S}(t|x)$, i.e. the overall survival $S(t|x)$ has also been identified.

Next, we consider equality of the likelihood contribution of an uncensored observation of type k, that is:

$$(1-\pi(x))\lambda_k(t\ |x)S(t|x) = (1-\tilde{\pi}(x))\tilde{\lambda}_k(t\ |x)\tilde{S}(t|x). \tag{3.62}$$

Since π_1 and $S(t|x)$ have been identified, this equation reduces to:

$$\lambda_k(t|x) = \tilde{\lambda}_k(t|x). \tag{3.63}$$

By identifiability of the Cox model, the cause-specific baseline hazard and Cox regression coefficients are identified. Thus the model is identifiable. □

So analogously to the standard competing risk model, the cause-specific hazards are observable from the data. Now we will turn our attention to the subdistribution hazard, which is also an observable quantity in standard competing risk modelling.

3.4.2 Identifiability of the subdistribution hazard for complete cure

We consider cure in the competing risk setting and model the subdistribution hazard. The contribution of an individual who experiences an event of type k is as follows:

$$\begin{aligned} L_k(\theta) &= \mathbb{P}(T = t, D = k, B = 0) \\ &= \mathbb{P}(B = 0)\mathbb{P}(T = t, D = k \mid B = 0) \\ &= (1 - \pi(x))\lambda_k^{sd}(t \mid \mathbf{Z})\big(1 - I_k(t)\big) \\ &= (1 - \pi(x))\lambda_k^{sd}(t \mid \mathbf{Z}) \exp\bigg\{ - \int_0^t \lambda_k^{sd}(u \mid \mathbf{Z}) du \bigg\}. \end{aligned} \tag{3.64}$$

where $\lambda_k^{sd}(t \mid \mathbf{Z})$ is the subdistribution hazard and $I_k(t)$ is the cumulative incidence function of event k. Since a cured individual will never experience any of the competing events, the likelihood contribution of a censored observation is given by:

$$\begin{aligned} L_0(\theta) &= \mathbb{P}(T > t) \\ &= \pi(x) + (1 - \pi(x))S(t \mid \mathbf{Z}) \\ &= \pi(x) + (1 - \pi(x)) \left(1 - \sum_{i=1}^{K} I_k(t) \right). \end{aligned} \tag{3.65}$$

Theorem 9. *Under assumptions **(B1)-(B2)** the cure fraction and the subdistribution hazards are identifiable.*

Proof. Assume that $\mathcal{L}(\theta) = \mathcal{L}(\tilde{\theta})$ almost surely. By Lemma 4, we can consider the equality of the likelihoods over these three subsets separately. In a similar fashion to the proof of Theorem 8, we can identify $\pi(x)$. Now consider the event $(T < \tau, D = k)$. Then we find that:

$$\lambda_k^{sd}(t \mid \mathbf{Z}) \exp\bigg\{ - \int_0^t \lambda_k^{sd}(u \mid \mathbf{Z}) du \bigg\} = \tilde{\lambda}_k^{sd}(t \mid \mathbf{Z}) \exp\bigg\{ - \int_0^t \tilde{\lambda}_k^{sd}(u \mid \mathbf{Z}) du \bigg\}. \tag{3.66}$$

Integrating both sides of the equation yields:

$$\exp\bigg\{ - \int_0^t \lambda_k^{sd}(u \mid \mathbf{Z}) du \bigg\} = \exp\bigg\{ - \int_0^t \tilde{\lambda}_k^{sd}(u \mid \mathbf{Z}) du \bigg\} + C, \tag{3.67}$$

for some constant $C \in \mathbb{R}$. Note that it must hold that $C = 0$. Since $t = 0$ results in $1 = 1 + C$. This implies that $\lambda_k^{sd}(u|x) = \tilde{\lambda}_k^{sd}(u|x)$, i.e. the subdistribution hazards are identified. It then follows from the identifiability of the Cox model that the model is identified. □

This shows that the subdistribution hazard is identifiable from the data as well. As mentioned, the cause-specific hazard and subdistribution hazard are the only observable quantities in competing risk analysis. This is thus also the case when we also take a fraction of cured patients into account and assume that the cure structure is complete.

3.4.3 Identifiability of the Vertical model

In this section, the Vertical model is introduced (Nicolaie et al., 2019). The Vertical model starts the analysis from a different decomposition of the joint probability. The joint probability can be decomposed as follows:

$$\mathbb{P}(T, D) = \mathbb{P}(D \mid T) \cdot \mathbb{P}(T),$$

where T is the survival time and D the competing risks status indicator. This is a fundamentally different decomposition compared to the classical decomposition which conditions on the event that is experienced. First, the follow-up time is modelled and then, conditional, on the follow-up time, the event is modelled. This contrasts with previous models where at onset it is determined which event will – or will not – be experienced.

The follow-up time is modelled as a whole, that is, first the time to any event is modelled. This is done using the total hazard, given in Equation (3.39), of all events. The total hazard induces a proper survival function for the uncured individuals and a corresponding density, i.e. $S_u(t|x) = \exp\Big\{-\int_0^t \lambda_\bullet(u|x)du\Big\}$ and $f_u(t|x) = -\frac{\partial}{\partial t}S_u(t|x)$. The total hazard is consequently modelled using the Cox PH model to incorporate the effect of covariates. After the time to any of the events is modelled, the incidence of the competing events is modelled. For this, the relative hazards are modelled through a multinomial model. The conditional *relative hazard* of competing event k at time t is defined as:

$$\rho_k(t) := \mathbb{P}(D = k \mid T = t, B_k = 0), \tag{3.68}$$

and equals the ratio of the hazard with respect to the total hazard, that is:

$$\rho_k(t) = \frac{\lambda_k^*(t)}{\sum_{i=1}^K \lambda_i^*(t)} = \frac{\lambda_k^*(t)}{\lambda_\bullet(t)}. \tag{3.69}$$

The relative hazard of cause k is the probability that, given that an event was experienced at time t, the event was of type k. It is modelled using a multinomial regression model, while the probability of cure $\pi(x)$ is modelled through a logistic regression model.

There are two different types of contributions to the likelihood. Individuals who experienced an event at time t of type k contributes:

$$\begin{aligned} L_k(\theta) &= \mathbb{P}(T = t, D = k, B = 0) \\ &= \mathbb{P}(B = 0)\mathbb{P}(T = t \mid B = 0)\mathbb{P}(D = k \mid T = t, B = 0) \\ &= (1 - \pi(x))f_u(t|x)\rho_k(t). \end{aligned} \tag{3.70}$$

An individual who is censored at time t contributes:

$$\begin{aligned} L_0(\theta) &= \mathbb{P}(T > t) \\ &= \mathbb{P}(B = 0)\mathbb{P}(T > t \mid B = 0) + \mathbb{P}(B = 1) \\ &= (1 - \pi(x))S_u(t|x) + \pi(x) \end{aligned} \tag{3.71}$$

As we consider an additional model for the estimation of the relative risks, we have to make further assumptions. Additionally to the conditions imposed under **(D1)-(D2)**, we assume that:

(D3) The multinomial model is identifiable.

It is well-known that the multinomial model is identifiable if one of the parameters is chosen to be equal to zero. It can be chosen arbitrarily as the estimated probabilities are not affected by it.

Theorem 10. *Under assumptions **(D1)**-**(D3)** the Vertical model is identifiable.*

Proof. Assume that $\mathcal{L}(\theta) = \mathcal{L}(\tilde{\theta})$ almost surely. By Lemma 4, we can consider the equality of the likelihoods over these three subsets separately. First, consider the case where $T > \tau$ and $\delta = 0$. On this subset, the following holds:

$$\pi(x) + (1 - \pi(x))S_u(t|x) = \tilde{\pi}(x) + (1 - \tilde{\pi}(x))\tilde{S}_u(t|x) \tag{3.72}$$

It follows from assumption **(B3)** that $S_u(t \mid \mathbf{Z}) = 0$ on this part of the sample space. Therefore, it must hold that:

$$\pi(x) = \tilde{\pi}(x). \tag{3.73}$$

By identifiability of the logistic model, the cure fraction has been identified.

Now suppose that $T \leq \tau$ and $\delta = 0$. Then Equation (3.74) still holds, but since the cure fraction has been identified, we find that:

$$\pi(x) + (1 - \pi(x))S_u(t|x) = \pi(x) + (1 - \pi(x))\tilde{S}_u(t|x), \tag{3.74}$$

which implies that $S_u(t|x) = \tilde{S}_u(t|x)$. It follows that the total hazards must be equal. By the identifiability of the Cox model, the baseline total hazard and Cox's regression coefficients are identified.

Consider $T \leq \tau$ and $D = k$. Then we have that:

$$(1 - \pi(x))f_u(t|x)\rho_k(t) = (1 - \tilde{\pi}(x))\tilde{f}_u(t|x)\tilde{\rho}_k(t). \tag{3.75}$$

Since $\pi(x)$ and $f_u(t|x)$ have been identified, this equation reduces to $\rho_k(t) = \tilde{\rho}_k(t)$, i.e. the relative hazards are identified. By identifiability of the multinomial model, the corresponding parameters are identified. □

Under the assumption of complete cure, three different models can be employed: cause-specific hazards, subdistribution hazards and a Vertical approach. All three are identifiable and are thus suited for statistical inference in practical applications.

Chapter 4

Estimation of the competing risks cure model

The estimation procedure of cure models depends on unobserved random variables known as the *cure statuses*. Therefore, the EM (Expectation-Maximization) algorithm is commonly used to estimate the parameters of the standard cure model. The EM algorithm effectively deals with latent variables by iteratively approximating the maximum likelihood estimate of the model's parameters. In this chapter we apply the EM algorithm to address our estimation problem: estimating the parameters of the competing risks cure model. This model is presented in Section 3.2. It takes into account two competing events whose potential survival times are distributed independently and no prior assumptions on the cure structure are made.

The chapter is structured as follows: first, a general introduction to the EM algorithm is given. Then details about the E-step and M-step are provided. The chapter ends with the extension of the estimation procedure to the complete cure model.

4.1 Introduction to the EM algorithm

The Expectation Maximization (EM) algorithm is applicable to maximum likelihood estimation where missing data is present. It is an iterative procedure which consists of two steps: the E-step or the *expectation step* and the M-step or the *maximization step*. The algorithm was first introduced by Dempster et al. (1977). It has been widely used in a broad range of applications since then. In a nutshell, the algorithm first approximates the missing values based on a set of parameters and the observed data, and secondly optimizes the model. These steps are repeated until convergence. To properly introduce the algorithm, a new concept needs to be introduced: the complete (log-)likelihood.

The data can be split up into two parts: the observed data $\mathcal{O}$ and the latent data W. The latent data consists of the cause-specific cure statuses $W = (B_1, B_2)$ which are defined in Section 3.1. For each subject i, the observed data consists of the follow-up time, the competing risk status indicator and the covariates, i.e. $\mathcal{O}_i = (t'_i, \delta_i, x_i)$. The *complete likelihood* is the likelihood of the observed and latent data. In case the latent data are observed, the likelihood is given by the complete likelihood. The contributions of the three types of observations (censored, event 1 and event 2) to the complete likelihood $\mathcal{L}^c(\theta; \mathcal{O}, W)$ are given by:

$$\begin{aligned}
L_0^c(\theta; \mathcal{O}, W) &= (1-\pi_1)(1-\pi_2^0)S_1(t)S_2(t)(1-B_1)(1-B_2) + (1-\pi_1)\pi_2^0 S_1(t)(1-B_1)B_2 \\
&\quad + \pi_1(1-\pi_2^1)S_2(t)B_1(1-B_2) + \pi_1\pi_2^1 B_1 B_2, \\
L_1^c(\theta; \mathcal{O}, W) &= (1-\pi_1)\pi_2^0 f_1(t)(1-B_1)B_2 + (1-\pi_1)(1-\pi_2^0)f_1(t)S_2(t)(1-B_1)(1-B_2), \\
L_2^c(\theta; \mathcal{O}, W) &= \pi_1(1-\pi_2^1)f_2(t)B_1(1-B_2) + (1-\pi_1)(1-\pi_2^0)f_2(t)S_1(t)(1-B_1)(1-B_2).
\end{aligned} \tag{4.1}$$

The dependence on $\mathcal{O}$ and W will be omitted from the notation. Furthermore, to reduce the complexity of the notation, the following indicator variables are introduced:

$$\begin{aligned}
\delta_i^0 &= \mathbf{1}(D_i = 0), \\
\delta_i^1 &= \mathbf{1}(D_i = 1), \\
\delta_i^2 &= \mathbf{1}(D_i = 2).
\end{aligned} \tag{4.2}$$

These quantities are equal to 1 if the subject is censored, experienced an event of type 1 or experienced an event of type 2, respectively. The *complete log-likelihood* is then given by:

$$\ell^c(\theta; \mathcal{O}, W) = \delta^0 \log L_0^c(\theta) + \delta^1 \log L_1^c(\theta) + \delta^2 \log L_2^c(\theta), \tag{4.3}$$

where θ is the vector of all parameters. This log-likelihood contrasts the *observed log-likelihood* which is given by taking the logarithm of the appropriate contributions given in (3.43). This yields:

$$\ell(\theta; \mathcal{O}) = \delta^0 \log L_0(\theta) + \delta^1 \log L_1(\theta) + \delta^2 \log L_2(\theta). \tag{4.4}$$

Now that the complete log-likelihood has been introduced, we can proceed to the actual EM algorithm. Let $\hat{\theta}$ denote the current estimates of the parameters. After initialization, the algorithm repeats the following two steps:

***E*-step.** Calculate the conditional expectation of the complete log-likelihood:

$$Q(\theta \mid \hat{\theta}) := \mathbb{E}_{W|\hat{\theta},\mathcal{O}}[\ell^c(\theta \mid \mathcal{O}, W)].$$

***M*-step.** Choose $\hat{\theta}_{\text{new}} \in \arg\max_{\theta\in\Omega} Q(\theta \mid \hat{\theta})$, that is

$$Q(\hat{\theta}_{\text{new}} \mid \hat{\theta}) \geq Q(\theta \mid \hat{\theta}) \tag{4.5}$$

for all $\theta \in \Omega$.

The steps are repeated until convergence. The algorithm is considered to have converged if the difference in Euclidean distance between successive parameter estimates is relatively small. It was proven by Dempster et al. (1977) that – under suitable regularity conditions – the algorithm converges to the maximum likelihood estimate.

For the estimation problem addressed here, we need to find an expression of the conditional expectation of the complete log-likelihood and maximize this expression with respect to the parameters. The coming two sections will be devoted to these two tasks respectively.

4.2 *E*-step

The E-step calculates the conditional expectation of the parameters with respect to the unobserved cure statuses given the observed data and the current estimates of the parameters. In this section, we will compute the expressions for these conditional expectations. Throughout this chapter, we will – as above – use the term *current*. This refers to the estimates or knowledge at hand in the given iteration. We will refrain from explicitly registering the number of the current iteration throughout this chapter.

This section will be split up into two subsections. First, we define the weights which are used – and updated – in each iteration. These weights coincide with the conditional expectations of the cure statuses. Then we will express the conditional expectation of the complete log-likelihood in terms of these weights and the current parameter estimates.

4.2.1 Conditional expectation of the cure statuses

Before we will delve into the conditional expectation of the complete log-likelihood, we define the following weights:

$$\begin{aligned} \phi &= \mathbb{E}_{W|\hat{\theta},\mathcal{O}}[B_1], \\ \psi^0 &= \mathbb{E}_{W|\hat{\theta},\mathcal{O}}[B_2 \ |B_1 = 0], \\ \psi^1 &= \mathbb{E}_{W|\hat{\theta},\mathcal{O}}[B_2 \ |B_1 = 1]. \end{aligned} \tag{4.6}$$

These weights are the conditional expectations of the cure statuses with respect to the observed data and current parameter estimates. They form the building blocks for the conditional expectation of the complete log-likelihood – which is computed later – and can be expressed in terms of the observed data and current parameter estimates. This section will be devoted to this task.

Quantities evaluated at the current estimates will be denoted as $\widehat{\cdot}$, to distinguish them from the ones with respect to which is maximized. This will be omitted from the notation if it is evident that the quantity belongs to either of the two groups. For example, ϕ, ψ^0 and ψ^1 are always evaluated at the current estimates. Additionally, these estimates may depend on the covariates x_i for the i-th individual in the data. Therefore, the estimates of, e.g. π_1, may be different for the different subjects in the data. Recall that the covariates, competing risks status indicator and observed follow-up time for individual i are denoted, by x_i, δ_i and t_i, respectively. So the weights can be written as follows:

$$\begin{aligned} \phi_i &= \mathbb{E}_{W|\hat{\theta},\mathcal{O}}[B_1] = \mathbb{P}(B_1 = 1 \mid x_i, t_i, \delta_i, \hat{\theta}), \\ \psi_i^0 &= \mathbb{E}_{W|\hat{\theta},\mathcal{O}}[B_2 \mid B_1 = 0] = \mathbb{P}(B_2 = 1 \mid B_1 = 0,\ x_i, t_i, \delta_i, \hat{\theta}), \\ \psi_i^1 &= \mathbb{E}_{W|\hat{\theta},\mathcal{O}}[B_2 \mid B_1 = 1] = \mathbb{P}(B_2 = 1 \mid B_1 = 1,\ x_i, t_i, \delta_i, \hat{\theta}). \end{aligned} \tag{4.7}$$

First, an expression for ϕ_i will be derived. It represents the current estimate of the probability of being cured of event one for individual i. If $\delta_i = 1$, individual i is not cured of this particular event. It therefore holds that $\phi_i = 0$. In the case that $\delta_i \neq 1$, individual i is either censored or the subject experienced an event of the other type. It can therefore be written as follows:

$$\begin{aligned}\phi_i &= \mathbb{P}(B_1 = 1 \mid x_i, t_i, \delta_i, \hat{\theta}) \\ &= \frac{\mathbb{P}(B_1 = 1, T_1 > t_i, T_2 > t_i \mid x_i, \hat{\theta})}{\mathbb{P}(T_1 > t_i, T_2 > t_i \mid x_i, \hat{\theta})} \mathbf{1}(\delta_i = 0) \\ &\quad + \frac{\mathbb{P}(B_1 = 1, T_1 > t_i, T_2 = t_i \mid x_i, \hat{\theta})}{\mathbb{P}(T_1 > t_i, T_2 = t_i \mid x_i, \hat{\theta})} \mathbf{1}(\delta_i = 2).\end{aligned} \tag{4.8}$$

In the last line, two different terms are given. These will be computed separately. The first term can be expressed as follows:

$$\begin{aligned}&\frac{\mathbb{P}(B_1 = 1, T_1 > t_i, T_2 > t_i \mid x_i, \hat{\theta})}{\mathbb{P}(T_1 > t_i, T_2 > t_i \mid x_i, \hat{\theta})} \\ &= \frac{\mathbb{P}(B_1 = 1, T_2 > t_i \mid x_i, \hat{\theta})}{\mathbb{P}(B_1 = 1, T_2 > t_i \mid x_i, \hat{\theta}) + \mathbb{P}(B_1 = 0, T_1 > t_i, T_2 > t_i \mid x_i, \hat{\theta})} \\ &= \frac{\mathbb{P}(B_1 = 1, B_2 = 1 \mid x_i, \hat{\theta}) + \mathbb{P}(B_1 = 1, B_2 = 0, T_2 > t_i \mid x_i, \hat{\theta})}{\begin{array}{l}\Big\{\mathbb{P}(B_1 = B_2 = 1 \mid x_i, \hat{\theta}) + \mathbb{P}(B_1 = 1, B_2 = 0, T_2 > t_i \mid x_i, \hat{\theta}) \\ \quad + \mathbb{P}(B_1 = 0, T_1 > t_i, B_2 = 1 \mid x_i, \hat{\theta}) + \mathbb{P}(B_1 = B_2 = 0, T_1 > t_i, T_2 > t_i \mid x_i, \hat{\theta})\Big\}\end{array}} \\ &= \frac{\widehat{\pi}_1 \widehat{\pi}_2^1 + \widehat{\pi}_1 (1 - \widehat{\pi}_2^1) S_2(t_i)}{\widehat{\pi}_1 \widehat{\pi}_2^1 + \widehat{\pi}_1 (1 - \widehat{\pi}_2^1) S_2(t_i) + (1 - \widehat{\pi}_1) \widehat{\pi}_2^0 S_1(t_i) + (1 - \widehat{\pi}_1)(1 - \widehat{\pi}_2^0) S_1(t_i) S_2(t_i)} \\ &= \left(1 + \frac{1 - \widehat{\pi}_1}{\widehat{\pi}_1} \frac{\widehat{\pi}_2^0 S_1(t_i) + (1 - \widehat{\pi}_2^0) S_1(t_i) S_2(t_i)}{\widehat{\pi}_2^1 + (1 - \widehat{\pi}_2^1) S_2(t_i)}\right)^{-1}.\end{aligned} \tag{4.9}$$

The dependence of the current estimates on x_i is omitted from the notation in the last lines. In the second last step, we expressed the conditional probabilities in terms of the (logistic) estimates of the cure probabilities. These can be computed using the following strategy:

$$\begin{aligned}\mathbb{P}(B_1 = B_2 = 0, T_1 > t_i \mid x_i, \hat{\theta}) &= \mathbb{P}(B_1 = 0 \mid x_i, \hat{\theta}) \cdot \mathbb{P}(B_2 = 0 \mid B_1 = 0, x_i, \hat{\theta}) \\ &\quad \cdot \mathbb{P}(T_1 > t_i \mid B_1 = 0, x_i, \hat{\theta}) \cdot \mathbb{P}(T_2 > t_i \mid B_2 = 0, x_i, \hat{\theta}) \\ &= \widehat{\pi}_1 \widehat{\pi}_2^0 S_1(t_i) S_2(t_i).\end{aligned} \tag{4.10}$$

The others follow similarly. This strategy can be used once more to compute the second term of the quantity in (4.8). This yields:

$$\begin{aligned}&\frac{\mathbb{P}(B_1 = 1, T_1 > t_i, T_2 = t_i \mid x_i, \hat{\theta})}{\mathbb{P}(T_1 > t_i, T_2 = t_i \mid x_i, \hat{\theta})} \\ &= \frac{\mathbb{P}(B_1 = 1, T_2 = t_i \mid x_i, \hat{\theta})}{\mathbb{P}(B_1 = 1, T_2 = t_i \mid x_i, \hat{\theta}) + \mathbb{P}(B_1 = 0, T_1 > t_i, T_2 = t_i \mid x_i, \hat{\theta})} \\ &= \frac{\mathbb{P}(B_1 = 1, B_2 = 0, T_2 = t_i \mid x_i, \hat{\theta})}{\mathbb{P}(B_1 = 1, B_2 = 0, T_2 = t_i \mid x_i, \hat{\theta}) + \mathbb{P}(B_1 = 0, B_2 = 0, T_1 > t_i, T_2 = t_i \mid x_i, \hat{\theta})} \\ &= \frac{\widehat{\pi}_1 (1 - \widehat{\pi}_2^1) f_2(t_i)}{\widehat{\pi}_1 (1 - \widehat{\pi}_2^1) f_2(t_i) + (1 - \widehat{\pi}_1)(1 - \widehat{\pi}_2^0) S_1(t_i) f_2(t_i)} \\ &= \left(1 + \frac{(1 - \widehat{\pi}_1)(1 - \widehat{\pi}_2^0)}{\widehat{\pi}_1 (1 - \widehat{\pi}_2^1)} S_1(t_i)\right)^{-1}.\end{aligned} \tag{4.11}$$

In summary, the weight ϕ_i is given by:

$$\begin{aligned}\phi_i = \mathbf{1}(\delta_i = 0)\left(1 + \frac{1-\widehat{\pi}_1}{\widehat{\pi}_1}\frac{\widehat{\pi}_2^0 S_1(t_i) + (1-\widehat{\pi}_2^0)S_1(t_i)S_2(t_i)}{\widehat{\pi}_2^1 + (1-\widehat{\pi}_2^1)S_2(t_i)}\right)^{-1} \\ + \mathbf{1}(\delta_i = 2)\left(1 + \frac{(1-\widehat{\pi}_1)(1-\widehat{\pi}_2^0)}{\widehat{\pi}_1(1-\widehat{\pi}_2^1)}S_1(t_i)\right)^{-1}.\end{aligned} \tag{4.12}$$

Now it remains to compute ψ^0 and ψ^1. Note that both are equal to 0 if $\delta = 2$. In that case, we know that the patient experienced an event of type two and – thus – was not cured of this event. In the case that $\delta \neq 2$, we find that:

$$\begin{aligned}\psi_i^0 &= \mathbf{1}(\delta_i \neq 2)\mathbb{P}(B_2 = 1 \mid B_1 = 0,\ x_i, t_i, \delta_i, \hat{\theta}) \\ &= \mathbf{1}(\delta_i \neq 2)\frac{\mathbb{P}(B_2 = 1, T_2 > t_i \mid B_1 = 0,\ x_i, \hat{\theta})}{\mathbb{P}(T_2 > t_i \mid B_1 = 0,\ x_i, \hat{\theta})} \\ &= \mathbf{1}(\delta_i \neq 2)\frac{\widehat{\pi}_2^0}{\widehat{\pi}_2^0 + (1-\widehat{\pi}_2^0)S_2(t-i)}.\end{aligned} \tag{4.13}$$

Since the potential survival times are independent, the events $\{T_1 > t_i\}$ or $\{T_1 = t_i\}$ can be ignored – i.e. their influence on the probabilities cancel out. The weight ψ^1 can be computed analogously. Note that $\mathbb{P}(B_1 = 1, \delta = 1) = 0$, i.e. it does not happen that $\{B_1 = 1\}$ and $\{\delta = 1\}$. Therefore, this yields the following expression:

$$\psi_i^1 = \mathbf{1}(\delta_i = 0)\frac{\widehat{\pi}_2^1}{\widehat{\pi}_2^1 + (1-\widehat{\pi}_2^1)S_2(t_i)}. \tag{4.14}$$

These weights form one of the building blocks of the conditional expectation of the complete log-likelihood which will be updated at each iteration of the algorithm. It now remains to find an expression for this conditional expectation. This will be given in the next subsection.

4.2.2 The conditional expectation of the complete log-likelihood

Recall from (4.3) that $\ell^c(\theta \mid \mathcal{O}, W)$ is given by three separate terms: $\log L_0^c(\theta)$, $\log L_1^c(\theta)$ and $\log L_2^c(\theta)$. The conditional expectation of these three will be computed in the respective order. For the first one, we have that:

$$\begin{aligned}\mathbb{E}_{W|\hat{\theta},\mathcal{O}}[\log L_0^c(\theta)] = \mathbb{E}_{W|\hat{\theta},\mathcal{O}}\Big[&(1-\pi_1)(1-\pi_2^0)S_1(t)S_2(t)(1-B_1)(1-B_2) \\ &+ (1-\pi_1)\pi_2^0 S_1(t)(1-B_1)B_2 \\ &+ \pi_1(1-\pi_2^1)S_2(t)B_1(1-B_2) \\ &+ \pi_1\pi_2^1 B_1 B_2\Big].\end{aligned} \tag{4.15}$$

The terms $(1-B_1)(1-B_2)$, $(1-B_1)B_2$, $B_1(1-B_2)$ and B_1B_2 are Bernoulli random variables. The expectation is therefore given by the probability of being equal to 1. The joint probability of, for example, $B_1(1-B_2)$ being equal to 1 can be computed as follows:

$$\begin{aligned}\mathbb{P}(B_1(1-B_2)=1\mid\hat{\theta},\mathcal{O}) &= \mathbb{P}(1-B_2=1\mid B_1=1,\hat{\theta},\mathcal{O})\mathbb{P}(B_1=1\mid\hat{\theta},\mathcal{O})\\ &= \big(1-\mathbb{P}(B_2=1\mid B_1=1,\hat{\theta},\mathcal{O})\big)\mathbb{P}(B_1=1\mid\hat{\theta},\mathcal{O})\\ &= (1-\psi^1)\phi.\end{aligned}\tag{4.16}$$

A similar approach can be used to compute the other probabilities. These are given by:

$$\begin{aligned}&\mathbb{P}(B_1B_2=1\mid\hat{\theta},\mathcal{O})=\phi\psi^1,\\ &\mathbb{P}((1-B_1)B_2=1\mid\hat{\theta},\mathcal{O})=(1-\phi)\psi^0,\\ &\mathbb{P}((1-B_1)(1-B_2)=1\mid\hat{\theta},\mathcal{O})=(1-\phi)(1-\psi^0).\end{aligned}\tag{4.17}$$

Plugging this result into (4.15) yields the following expression for the contribution of a censored observation to the complete log-likelihood:

$$\begin{aligned}\mathbb{E}_{W|\hat{\theta},\mathcal{O}}[\log L_0^c(\theta)] &= (1-\phi)(1-\psi^0)\log\big\{(1-\pi_1)(1-\pi_2^0)S_1(t)S_2(t)\big\}\\ &\quad+(1-\phi)\psi^0\log\big\{(1-\pi_1)\pi_2^0S_1(t)\big\}\\ &\quad+\phi(1-\psi^1)\log\big\{\pi_1(1-\pi_2^1)S_2(t)\big\}+\phi\psi^1\log\big\{\pi_1\pi_2^1\big\}\\ &=\phi\log\{\pi_1\}+(1-\phi)\log\{1-\pi_1\}\\ &\quad+(1-\phi)\psi^0\log\{\pi_2^0\}+(1-\phi)(1-\psi^0)\log\{1-\pi_2^0\}\\ &\quad+\phi\psi^1\log\{\pi_2^1\}+\phi(1-\psi^1)\log\{1-\pi_2^1\}\\ &\quad+(1-\phi)\log\{S_1(t)\}\\ &\quad+[(1-\phi)(1-\psi^0)+\phi(1-\psi^1)]\log\{S_2(t)\}.\end{aligned}\tag{4.18}$$

The other two types of contributions to the complete log-likelihood can be derived in a similar fashion. The contribution related to an observed event of type one is given can be written as:

$$\begin{aligned}\mathbb{E}_{W|\hat{\theta},\mathcal{O}}[\log L_1^c(\theta)] &= (1-\phi)\psi^0\log\big\{(1-\pi_1)\pi_2^0f_1(t)\big\}\\ &\quad+(1-\phi)(1-\psi^0)\log\big\{(1-\pi_1)(1-\pi_2^0)f_2(t)S_2(t)\big\}\\ &=(1-\phi)\log\{1-\pi_1\}\\ &\quad+(1-\phi)\psi^0\log\{\pi_2^0\}+(1-\phi)(1-\psi^0)\log\{1-\pi_2^0\}\\ &\quad+(1-\phi)\log\{f_1(t)\}\\ &\quad+(1-\phi)(1-\psi^0)\log\{S_2(t)\}.\end{aligned}\tag{4.19}$$

The conditional expectation of the complete log-likelihood contribution of an observed event of type two is given by:

$$\begin{aligned}\mathbb{E}_{W|\hat{\theta},\mathcal{O}}[\log L_2^c(\theta)] &= \phi(1-\psi^1)\log\left\{\pi_1(1-\pi_2^1)f_2(t)\right\}\\ &\quad + (1-\phi)(1-\psi^0)\log\left\{(1-\pi_1)(1-\pi_2^0)f_2(t)S_1(t)\right\}\\ &= \phi(1-\psi^1)\log\{\pi_1\} + (1-\phi)(1-\psi^1)\log\{1-\pi_1\}\\ &\quad + (1-\phi)(1-\psi^0)\log\{1-\pi_2^0\}\\ &\quad + \phi(1-\psi^1)\log\{1-\pi_2^1\}\\ &\quad + (1-\phi)(1-\psi^1)\log\{S_1(t)\}\\ &\quad + [(1-\phi)(1-\psi^0)+\phi(1-\psi^1)]\log\{f_2(t)\}.\end{aligned} \tag{4.20}$$

Now that the three types of contributions to the complete log-likelihood are expressed in terms of the current estimates and weights, they can be combined using the formula given in (4.3). This yields the following expression for the conditional expectation of the complete log-likelihood of one individual in the data:

$$\begin{aligned}&\mathbb{E}_{W|\hat{\theta},\mathcal{O}}\left[\ell^c(\theta;\mathcal{O},W)\right]\\ &\quad= \Big[\delta^1\phi + \delta^2\phi(1-\psi^1)\Big]\log\{\pi_1\} + \Big[\delta^0(1-\phi)+\delta^1(1-\phi)+\delta^2(1-\phi)(1-\psi^1)\Big]\log\{1-\pi_1\}\\ &\qquad + \Big[(\delta^0+\delta^1)(1-\phi)\psi^0\Big]\log\{\pi_2^0\} + \Big[(\delta^0+\delta^1+\delta^2)(1-\phi)(1-\psi^0)\Big]\log\{1-\pi_2^0\}\\ &\qquad + \Big[\delta^0\phi\psi^1\Big]\log\{\pi_2^1\} + \Big[\delta^0 + (\delta^1+\delta^2)\phi(1-\psi^1)\Big]\log\{1-\pi_2^1\}\\ &\qquad + \Big[\delta^0(1-\phi)+\delta^1(1-\phi)+\delta^2(1-\phi)(1-\psi^1)\Big]\log\{S_1(t)\} + \Big[\delta^1(1-\phi)\Big]\log\{\lambda_1(t)\}\\ &\qquad + \Big[(\delta^0+\delta^2)[(1-\phi)(1-\psi^0)+\phi(1-\psi^1)] + \delta^1(1-\phi)(1-\psi^0)\Big]\log\{S_2(t)\}\\ &\qquad + \Big[\delta^2[(1-\phi)(1-\psi^0)+\phi(1-\psi^1)]\Big]\log\{\lambda_2(t)\}\\ &\quad= \phi\left[\delta^1+\delta^2(1-\psi^1)\right]\log\{\pi_1\} + (1-\phi)(1-\delta^2\psi^1)\log\{1-\pi_1\}\\ &\qquad + (\delta^0+\delta^1)(1-\phi)\psi^0\log\{\pi_2^0\} + (1-\phi)(1-\psi^0)\log\{1-\pi_2^0\}\\ &\qquad + \delta^0\phi\psi^1\log\{\pi_2^1\} + \left[\delta^0+(\delta^1+\delta^2)\phi(1-\psi^1)\right]\log\{1-\pi_2^1\}\\ &\qquad + (1-\phi)(1-\delta^2\psi^1)\log\{S_1(t)\} + \delta^1(1-\phi)\log\{\lambda_1(t)\}\\ &\qquad + \left[(\delta^0+\delta^2)\phi(1-\psi^1)+(1-\phi)(1-\psi^0)\right]\log\{S_2(t)\}\\ &\qquad + \delta^2\left[(1-\phi)(1-\psi^0)+\phi(1-\psi^1)\right]\log\{\lambda_2(t)\}.\end{aligned} \tag{4.21}$$

This provides the contribution of each individual to the log-likelihood. The conditional expectation of the complete log-likelihood is given by the sum of these contributions over all individuals. Now that we have found an expression for the conditional expectation of the log-likelihood given the observed data and current estimates, the E-step is concluded. The next step is to maximize this conditional expectation with respect to the parameters. This is the M-step of the algorithm.

4.3 M-step

It remains to maximize the conditional expectation of the complete log-likelihood given in (4.21). We will now take the sum of the log-likelihood contributions over all n individuals. First note from

(4.21) that it can be written as distinct sums of functions of the parameters. This implies that these sums can be maximized separately.

The first three sums are related to the incidence part of the model. They are of the following form:

$$f(\pi) = \sum_{i=1}^{n} \omega_i \log\{\pi_i\} + \upsilon_i \log\{1 - \pi_i\}, \tag{4.22}$$

where $\pi = (\pi_1, \pi_2, ..., \pi_n)$, $\omega = (\omega_1, \omega_2, ..., \omega_n)$ and $\upsilon = (\upsilon_1, \upsilon_2, ..., \upsilon_n)$. The last two sums are related to the latency and are of the following form:

$$g(\lambda) = \sum_{i=1}^{n} \omega_i \log\{S(t_i)\} + \delta_i \upsilon_i \log \lambda(t). \tag{4.23}$$

We need to maximize these two types of functions. They seem similar to the log-likelihoods of the logistic regression and weighted Cox proportional hazards model, respectively. However, they actually differ. This difference is caused by the differing weights ω and υ and as a consequence standard optimization techniques do not suffice. The next two subsections will be devoted to the task of maximizing these two types of functions.

4.3.1 Maximization of the incidence

We need to maximize the following objective function:

$$f(\pi) = \sum_{i=1}^{n} \omega_i \log\{\pi_i\} + \upsilon_i \log\{1 - \pi_i\}, \tag{4.24}$$

where the π_i's are functions of the logistic form:

$$\pi_i = \frac{e^{\gamma^\top z_i}}{1 + e^{\gamma^\top z_i}}, \tag{4.25}$$

and γ is a $(p{+}1)$-dimensional vector of regression coefficients and z_i is the vector $(1, x_{i,1}, x_{i,2}, ..., x_{i,p})$. The maximization procedure is with respect to γ. Moreover, the weights ω and υ are constant with respect to the maximization problem. We will therefore write $f(\gamma)$ instead of $f(\pi(\gamma))$. This maximization problem is nearly equal to the maximization problem of the standard logistic model. In that case $\upsilon_i = 1 - \omega_i$. The procedure for standard logistic maximization is often called Iterative Re-weighted Least Squares. A comprehensible introduction is given in (Friedman et al., 2017). This procedure will be adapted to our context with different weights.

If we plug the logistic form of π back into f given in (4.24) and rewrite it, the objective function reduces to:

$$f(\gamma) = \sum_{i=1}^{n} \omega_i \gamma^\top z_i - (\upsilon_i + \omega_i) \log\{1 + e^{\gamma^\top z_i}\}. \tag{4.26}$$

This function will be maximized by solving for the roots of its gradient. The gradient is given by:

$$\nabla f = \sum_{i=1}^{n} z_i \left(\omega_i - (\upsilon_i + \omega_i) \frac{e^{\gamma^\top z_i}}{1 + e^{\gamma^\top z_i}} \right), \tag{4.27}$$

where z_i is a $p+1$-dimensional vector ensuring that the gradient is actually a vector of the correct dimension. The roots of this multidimensional function can be approximated numerically using the `multiroot` function from the `rootSolve` package (Soetaert et al., 2022).

4.3.2 Maximization of the latency

Recall that the task is to maximize an objective function of the following form:

$$g(\lambda) = \sum_{i=1}^{n} \omega_i \log\{S(t_i|x_i)\} + \upsilon_i \delta_i \log\{\lambda(t_i|x_i)\}, \tag{4.28}$$

where λ is a vector representing the hazard. The hazard is modelled using the Cox proportional hazards model: $\lambda(t_i|x_i) = \lambda_0(t_i)\exp\{\beta^\top x_i\}$, where β is a vector of regression coefficients of length p. The baseline hazard λ_0 is modelled non-parametrically. It has nonzero entries on all time points on which an event was observed. Denote L the number of events and $t'_1, t'_2, ..., t'_L$ the time-points at which we observed an event. Then, $\lambda_0(t'_j) > 0$ for $j = 1, 2, ..., L$ and zero for all other time-points t. Furthermore, let $\mathcal{R}(t) \subseteq \{1, 2, ..., n\}$ be the risk set at time t and $\mathcal{D}(t) \subseteq \{1, 2, ..., n\}$ the set of tied observations at time t of the respective type. It can be seen from (4.21) that the log-likelihoods for the hazards of the competing events do not depend on each other. So we can consider the event of a fixed type for the maximization. Those who experience an event of the other type are considered censored observations. The survival function can then be written as follows:

$$S(t_i|x_i) = [S_0(t_i)]^{\exp\{\beta^\top x_i\}} = \left[\exp\left\{-\sum_{t'_j \le t_i} \lambda_0(t'_j)\right\}\right]^{\exp\{\beta^\top x_i\}}. \tag{4.29}$$

A Breslow-type method to incorporate tied observations is used. The previous results yield the following expression for the objective function:

$$\begin{aligned} g(\beta, \lambda_0) &= \sum_{i=1}^{n} \left\{ \delta_i \upsilon_i (\beta^\top x_i + \log\{\lambda_0(t_i)\}) - \sum_{t'_j \le t_i} \omega_i \exp\{\beta^\top x_i\} \lambda_0(t'_j) \right\} \\ &= \sum_{j=1}^{L} \left\{ \beta^\top \cdot \sum_{l \in \mathcal{D}(t'_j)} \upsilon_l x_l + \log\{\lambda_0(t'_j)\} \cdot \sum_{l \in \mathcal{D}(t'_j)} \upsilon_l - \lambda_0(t'_j) \cdot \sum_{l \in \mathcal{R}(t'_j)} \omega_l \exp\{\beta^\top x_l\} \right\}. \end{aligned} \tag{4.30}$$

The first sum is a sum over all individuals in the data, while the latter is a sum over all distinct event time points. This objective function depends on both β and λ_0. A partial likelihood approach will be used to estimate β independently of the baseline hazard, and then optimize with respect to the baseline hazard. First, we will derive a Nelson-Aalen-type estimator for the baseline hazard which depends on β. Then, this expression is plugged back into the objective function g and it is used to derive a partial likelihood for β.

To find the non-parametric Nelson-Aalen-type estimators for the baseline hazard, we compute the roots of the gradient of g with respect to λ_0. The entries of the gradient are given by:

$$\frac{\partial g(\beta,\lambda_0)}{\partial \lambda_0(t'_j)} = \frac{\sum_{l\in\mathcal{D}(t'_j)} \upsilon_j}{\lambda_{0,j}} - \sum_{l\in\mathcal{R}(t'_j)} \omega_l \exp\{\beta^\top x_l\}. \tag{4.31}$$

Setting these expressions equal to zero yields a Nelson-Aalen-type estimator for the baseline hazard:

$$\widehat{\lambda}_{0,j} = \frac{\sum_{l\in\mathcal{D}(t'_j)} \upsilon_j}{\sum_{l\in\mathcal{R}(t'_j)} \omega_l \exp\{\beta^\top x_l\}}. \tag{4.32}$$

The partial likelihood can be found by plugging this estimator in the objective function g. This yields the following partial likelihood for β:

$$\prod_{j=1}^{L} \frac{\exp\Big\{\beta^\top \sum_{l\in\mathcal{D}(t'_j)} \upsilon_l x_l\Big\}}{\Big(\sum_{l\in\mathcal{R}(t'_j)} \omega_l \exp\{\beta^\top x_l\}\Big)^{\sum_{l\in\mathcal{D}(t'_j)} \upsilon_l}}. \tag{4.33}$$

Maximizing the partial likelihood (4.33) with respect to β is equivalent to maximizing:

$$\sum_{j=1}^{L} \left\{ \beta^\top \sum_{l\in\mathcal{D}(t'_j)} \upsilon_l x_l - \log\left\{ \sum_{l\in\mathcal{R}(t'_j)} \omega_l \exp\{\beta^\top x_l\} \right\} \sum_{l\in\mathcal{D}(t'_j)} \upsilon_l \right\}. \tag{4.34}$$

This function differs from the standard Breslow-type weighted partial likelihood. Therefore, standard functions available in `R` cannot be used. The standard Breslow-type estimator for the baseline hazard would be given if $\omega_i = \upsilon_i$ for all $i = 1, 2, ..., n$. The expression (4.34) will be maximized numerically using the `nlm` function in `R`.

4.4 Extending the estimation procedure

The estimation procedure explained above is – among others – designed to estimate the parameters related to the cure structure: π_1, π_2^0 and π_2^1. We have also discussed models where the cure structure parameters were a priori fixed: complete cure and cure for the event of interest. In order to estimate these models the estimation procedure needs to be adjusted. This section discusses how the algorithm can be adjusted or modified to accommodate different models.

4.4.1 estimation procedure for complete cure

The simpler model constructed in Section 3.4 is now considered for two competing events, i.e. $K = 2$. This model assumes a complete cure structure but does not assume the independence of potential survival times. Therefore, the marginal survival functions are not used since they do not have a survival function interpretation in this setting. We, therefore, reintroduce the quantity $S_k^*(t) = \exp\Big\{-\int_0^t \lambda_k(u)du\Big\}$, where $\lambda_k(t)$ is the cause-specific hazard. Furthermore, recall that since cure happens simultaneously, we considered a single cure status B. Define:

$$\phi := \mathbb{E}_{W|\hat{\theta},\mathcal{O}}[B] = \mathbb{P}(B = 1 \mid x_i, t_i, \delta_i, \hat{\theta}). \tag{4.35}$$

Under this model, ϕ can be expressed as:

$$\begin{aligned}\phi &= \mathbf{1}(\delta_i = 0)\frac{\mathbb{P}(B = 1 \mid x_i, \hat{\theta})}{\mathbb{P}(B = 1 \mid x_i, \hat{\theta}) + \mathbb{P}(B = 0 \mid x_i, \hat{\theta})\mathbb{P}(\min\{T_1, T_2\} > t_i \mid x_i, \hat{\theta})} \\ &= \mathbf{1}(\delta_i = 0)\frac{\pi}{\pi + (1-\pi)S(t_i)}.\end{aligned} \tag{4.36}$$

Moreover, note that the likelihood differs under this simplified model. The observed likelihood is specified in (3.29). The contributions to the complete log-likelihood are given by:

$$\begin{aligned}\log L_0^c(\theta) &= \log\big\{\pi B + (1-\pi)S(t)(1-B)\big\}, \\ \log L_1^c(\theta) &= \log\big\{(1-\pi)\lambda_1(t)S(t)(1-B)\big\}, \\ \log L_2^c(\theta) &= \log\big\{(1-\pi)\lambda_2(t)S(t)(1-B)\big\}.\end{aligned} \tag{4.37}$$

Since these quantities differ from the ones considered above, the conditional expectation of the complete log-likelihood changes. This is the equivalent of (4.21) and it is given by:

$$\begin{aligned}\mathbb{E}_{W|\hat{\theta},\mathcal{O}}[\ell^c(\theta; \mathcal{O}, W)] &= \delta^0\phi\log\{\pi\} + \delta^0(1-\phi)\log\{1-\pi\} \\ &\quad + \delta^1(1-\phi)\log\{(1-\pi)\lambda_1(t)S(t)\} \\ &\quad + \delta^2(1-\phi)\log\{(1-\pi)\lambda_2(t)S(t)\} \\ &= \delta^0\phi\log\{\pi\} + (1-\phi)\log\{1-\pi\} \\ &\quad + (1-\phi)\big[\log\{S_1(t) + \delta^1\log\{\lambda_1(t)\}\big] \\ &\quad + (1-\phi)\big[\log\{S_2(t) + \delta^2\log\{\lambda_2(t)\}\big].\end{aligned} \tag{4.38}$$

Note that it is of a simpler form than (4.21). The maximization procedure, therefore, does not need to change, but we can use a standard weighted Cox maximization procedure to find the optimum of the sums related to the latency. The maximization procedure for the parameters related to the incidence does not change. This can be done using the `coxph` function from the `survival` package (Therneau et al., 1990). Additionally, the weights in the EM algorithm need to be adjusted. These are specified by the formula for the conditional expectation of the log-likelihood in (4.38).

Chapter 5

A simulation experiment

In Chapter 3 the theoretical identifiability properties of the cure model in a setting with competing risks were studied. In this chapter, we will investigate whether these parameters are also identifiable in practice. The estimation procedure from the previous chapter will be used to estimate the parameters of the model in several simulated settings. It was partly shown and partly conjectured that – in the case of independent potential survival times – the cure structure and the time-to-event distributions were identifiable. It will be investigated whether this is also the case in practice.

This chapter illustrates two simulation studies. First, a general overview of the data generation procedure for each simulation study is given. Then the results are presented

5.1 Generating competing risks cure data

In this section, we elaborate on the general structure of the data generation process for the competing risks cure model. For the different simulation studies, a different data generation process is required. The specifications will therefore be given per study in the upcoming sections. In broad terms, the data generation process is as follows. First, the cure probability π_1 for event 1 is computed. This probability depends on the covariates through the logistic function. Given this probability, the cure status for event 1 is generated $B_1 \sim \text{Ber}(\pi_1)$. Then, given the cure status of event 1, the cure status for event 2 is determined. This can be done in several ways dependent on the goal of the study. For example, the cure statuses can be chosen a priori or can be generated using either $B_2 \sim \text{Ber}(\pi_2^0)$ or $B_2 \sim \text{Ber}(\pi_2^1)$. Then the potential survival times are generated using a Cox-Weibull model. At last, the follow-up time is computed by taking the minimum of the potential survival times and a censoring time. The censoring time is generated uniformly $C \sim U[0, t_{\max}]$ and independently of the other random variables.

The potential survival times are generated using the inverse probability transformation method. The survival times T_1 and T_2 are both distributed according to the Cox model with Weibull baseline survival distributions with parameters, respectively, denoted by (α_1, κ_1) and (α_2, κ_2). The baseline hazard functions of these distributions are given by:

$$\lambda_1^0(t) = \alpha_1 \kappa_1 t^{\kappa_1 - 1} \quad \text{and} \quad \lambda_2^0(t) = \alpha_2 \kappa_2 t^{\kappa_2 - 1}. \tag{5.1}$$

Lemma 11. *Let $U \sim \mathrm{U}(0,1)$, $x \in \mathcal{X}$ a vector of covariates and T bee given by:*

$$T = \left(-\frac{\log U}{\alpha \exp\{\beta^\top x\}}\right)^{1/\kappa}, \tag{5.2}$$

with $\alpha, \kappa > 0$ and $\beta \in \mathbb{R}^p$. Then T is distributed according to the Cox model with regression coefficients β and a Weibull baseline survival distribution with parameters (α, κ).

Proof. The complement of the cumulative distribution function of T can be derived from the distribution of U. It is given by:

$$\begin{aligned}
\mathbb{P}(T > t \mid x) &= \mathbb{P}\left(\left(-\frac{\log U}{\alpha \exp\{\beta^\top x\}}\right)^{1/\kappa} > t\right) \\
&= \mathbb{P}\Big(\log U < -\alpha t^\kappa \exp\big\{\beta^\top x\big\}\Big) \\
&= \mathbb{P}\left(U < (\exp\{-\alpha t^\kappa\})^{\exp\left\{\beta^\top x\right\}}\right) \\
&= (\exp\{-\alpha t^\kappa\})^{\exp\left\{\beta^\top x\right\}}.
\end{aligned} \tag{5.3}$$

Note that $\alpha t^\kappa = \int_0^t \alpha\kappa u^{\kappa-1} du$, i.e. it is the cumulative hazard of the Weibull distribution according to the parametrization given in (5.1). This implies that $\mathbb{P}(T > t \mid x) = S_0(t)^{\exp\left\{\beta^\top x\right\}}$ with S_0 the baseline survival function of the Weibull distribution with the correct parameters. Thus T has the specified distribution according to (2.11). □

Note that the Cox-Weibull distribution of the potential survival times does not meet the cure threshold condition. For this assumption to be satisfied, the survival times are truncated at the 99% quantile of the Weibull distributions. The respective cure thresholds τ_1 and τ_2 are given by:

$$\tau_k = \left(\frac{\log 100}{\alpha_k}\right)^{1/\kappa_k} \quad \text{for} \quad k = 1, 2. \tag{5.4}$$

This truncation leads to a positive probability of seeing equal time-to-events. In order to reduce this probability, the included covariate is chosen to be exponentially distributed. A positive covariate with positive Cox regression coefficients leads to accelerated time-to-event and thus a lower probability of seeing equal event times. In the simulation study, only one covariate will be included. Details for this choice are provided in Section 5.2.2.

Independent and dependent potential survival times will be simulated. Simulating independent survival times is straightforward, while some more work is required for the simulation of dependent survival times. The approach from Beyersmann et al. (2009) is adopted and can be summarized as follows:

1. Choose the cause-specific hazards $\lambda_1(t|x)$ and $\lambda_2(t|x)$ dependent on the covariates.
2. Simulate actual survival time T from the distribution specified by the all-cause hazard $\lambda_1(t|x) + \lambda_2(t|x)$.
3. Decide whether event 1 occurred at time T based on a binomial experiment with success probability $\lambda_1(t|x)/\big(\lambda_1(t|x) + \lambda_2(t|x)\big)$, otherwise event 2 occurred.

4. Simulate censoring time C.

This completes the overview of the data generation process for competing risks cure data. In the next section details about the data generation procedure for this simulation study are given.

5.2 Simulation study I

In this simulation study, we study the performance of the estimation procedure for a general cure structure. We investigate whether the estimation procedure proposed in Chapter 4 is able to properly estimate the parameters of the competing risks cure model described in Section 3.2. It was claimed that the parameters related to the cure structure were identifiable. We will investigate with a simulation study whether this is indeed plausible.

The section is structured as follows. First, we describe the data generation process and motivate the choice of the parameters in the simulation study. After that, we investigate the characteristics of the simulated data for the chosen parameters. Then we present the results of the simulation. To evaluate the results of the simulations study bias, variance and mean square error (MSE) are presented.

5.2.1 Set-up of the study and data generation

The procedure for generating the data has been discussed before. The parameters that ought to be chosen are:

- $\boldsymbol{(\alpha_1, \kappa_1)}$: shape and rate parameter of the Weibull distribution for the baseline survival of event 1;
- $\boldsymbol{(\alpha_2, \kappa_2)}$: shape and rate parameter of the Weibull distribution for the baseline survival of event 2;
- $\boldsymbol{\beta_1}$: regression coefficients of the Cox model for event 1;
- $\boldsymbol{\beta_2}$: regression coefficients of the Cox model for event 2;
- $\boldsymbol{\gamma_1}$: regression coefficients of the logistic model for cure probability of event 1;
- $\boldsymbol{\gamma_2^0}$: regression coefficients of the logistic model for cure probability of event 2 conditional on uncure for event 1;
- $\boldsymbol{\gamma_2^1}$: regression coefficients of the logistic model for cure probability of event 2 conditional on cure for event 1;
- $\boldsymbol{t}_{\max}$: end-of-study time.

A note of caution in the choice of the parameters is required since it is not immediately evident what data characteristics they lead to due to the complexity of the model and data-generation process. There are some aspects that we need to take into account when choosing the parameters. First of all, the four subgroups – as explained in Section 3.1 – must contain approximately the same amount of subjects. Enough events of both types must be present in each subgroup. With respect to the cure thresholds, there needs to be time in between them and they need to be of the following order: $\tau_1 < \tau_2$. In addition, some events of type 2 must occur after τ_1 and enough censored observations should be present after τ_2.

For an optimal setting of the simulation experiment, three different sets of parameters (Table 5.1) which meet the requirements above discussed are chosen through trial and error. Each scenario is simulated with three different sample sizes: $n = 500, 1000, 2500$ and with a fixed end-of-study time: $t_{\max} = 2$.

Parameters	Scenario 1	Scenario 2	Scenario 3
$\boldsymbol{(\alpha_1, \kappa_1)}$	(20, 15)	(8,1)	(6, 7.5)
$\boldsymbol{(\alpha_2, \kappa_2)}$	(2, 2.5)	(3,1)	(3, 2.5)
$\boldsymbol{\beta_1}$	0.5	0.5	2
$\boldsymbol{\beta_2}$	1	0.5	0.25
$\boldsymbol{\gamma_1}$	(0.25, -0.5)	(0.25, -1)	(-0.5, 1)
$\boldsymbol{\gamma_2^0}$	(1, -0.5)	(1, -0.5)	(0, 1)
$\boldsymbol{\gamma_2^1}$	(-0.5, 0.5)	(-0.5, 0.5)	(0, -1)

Table 5.1: Three parameter scenarios used in the simulation study.

Since the gamma parameters consist of two components, we will write $\gamma_1 = (\gamma_{1_0}, \gamma_{1_1})$, to indicate the entry related to the intercept and the regression coefficient, respectively. The simulated data is complex and it is not immediately evident how the data looks by looking at the parameters. In the next section we, therefore, provide some characteristics about the data corresponding to three choices of parameters.

An overview of the data-generating procedure for the above-described model can be found in Algorithm 1.

Algorithm 1 Simulate time-to-event data according to the competing risks cure model

```
set end-of-study time: $t_{\max} = 2$
for $i = 1, 2, ..., n$ do
    generate covariates: $x_i \sim \exp(1)$
    compute cure probability: $\pi_1(x_i)$
    generate cure status: $B_1 \sim \text{Ber}\big(\pi_1(x_i)\big)$
    if $B_1 = 1$ then
        compute cure probability: $\pi_2^1(x_i)$
        generate cure status: $B_2 \sim \text{Ber}\big(\pi_2^1(x_i)\big)$
    else
        compute cure probability: $\pi_1(x_i)$
        generate cure status: $B_2 \sim \text{Ber}\big(\pi_2^0(x_i)\big)$
    end if
    generate survival probability: $U_{1,i}, U_{2,i} \sim \text{U}(0,1)$ and censoring time: $C_i \sim \text{U}[0, t_{\max}]$
    compute potential survival times:
        $$T_{1,i} \leftarrow \left(-\frac{\log U_{1,i}}{\alpha_1 \exp\{\beta_1^\top x_i\}}\right)^{1/\kappa_1} \quad \text{and} \quad T_{2,i} \leftarrow \left(-\frac{\log U_{2,i}}{\alpha_2 \exp\{\beta_2^\top x_i\}}\right)^{1/\kappa_2}$$
    if $T_{1,i} > \tau_1$ then
        truncate survival time: $T_{1,i} \leftarrow \tau_1$
    end if
    if $T_{2,i} > \tau_2$ then
        truncate survival time: $T_{2,i} \leftarrow \tau_2$
    end if
    if $B_1 = B_2 = 0$ then
        set follow-up time: $T_i^* \leftarrow \min\{C_i, T_{1,i}, T_{2,i}\}$
    else if $B_1 = 0$ and $B_2 = 1$ then
        set follow-up time: $T_i^* \leftarrow \min\{C_i, T_{1,i}\}$
    else if $B_1 = 1$ and $B_2 = 0$ then
        set follow-up time: $T_i^* \leftarrow \min\{C_i, T_{2,i}\}$
    else if $B_1 = 1$ and $B_2 = 1$ then
        set follow-up time: $T_i^* \leftarrow C_i$
    end if
    if $T_i^* = T_{1,i}$ then
        set competing risks status indicator: $d_i \leftarrow 1$
    else if $T_i^* = T_{2,i}$ then
        set competing risks status indicator: $d_i \leftarrow 2$
    else
        set competing risks status indicator: $d_i \leftarrow 0$
    end if
end for
return competing risks survival data: $(T_i^*, d_i, x_i)$ for $i = 1, 2, ..., n$
```

5.2.2 Data characteristics

The following statistics for each simulation experiment are included:

- **Subgroup sizes**: the four subgroups are presented in Section 3.1 (as percentages of the whole population);
- **Events in subgroups**: number and type of events in the uncured subgroups;
- **Cure thresholds**: observed cure thresholds for both events;
- **Events after τ_1**: number of events of type 2 after τ_1;
- **Censoring rate**: number of censored observations.

The number of observations (censored/uncensored) is presented as a percentage with respect to the whole population. The characteristics are simulated with a population size of $N = 1000$ and Monte-Carlo replicates of $M = 10.000$. The relative group sizes are given in Figure 5.1.

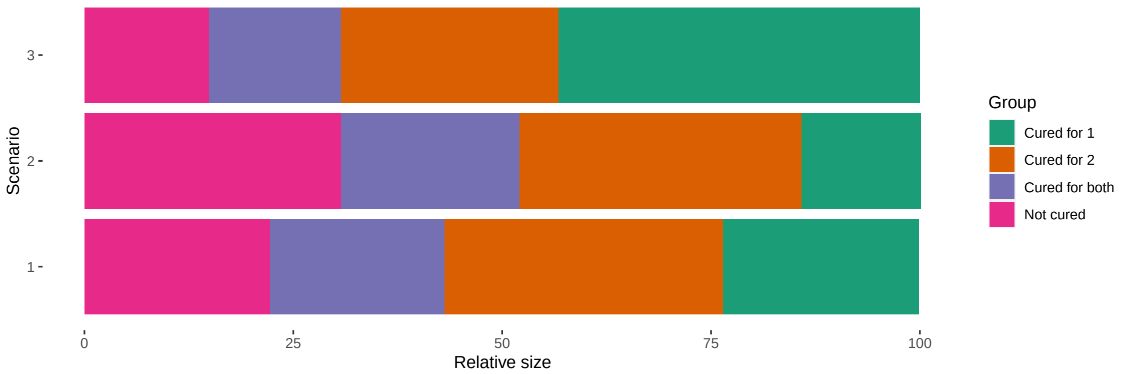

Figure 5.1: Three stacked bar charts of the relative group sizes (in percentages) for each scenario.

In Table 5.2 the relative group sizes are shown along with the 95% confidence intervals. The percentage of observations for each competing event and each group are shown in Table 5.3. Note that all groups contain a substantial amount of observations. This is crucial for obtaining proper results in the simulated examples, as mentioned above.

	Not cured	**Cured for event 1**	**Cured for event 2**	**Cured for both events**
1	22.2% (22.1% - 22.3%)	23.5% (23.4% - 23.6%)	33.3% (33,2% - 33.4%)	20.9% (20.8% - 21.0%)
2	30.7% (30.6% - 30.8%)	14.3% (14.2% - 14.3%)	33.7% (33.6% - 33.8%)	21.4% (21.3% - 21.4%)
3	14.9% (14.8% - 15.0%)	43.3% (43.2% - 43.4%)	26.0% (25.9% - 26.0%)	15.8% (15.7% - 15.9%)

Table 5.2: Relative sizes for the four different cure status groups with 95%-confidence intervals for each of the three scenarios in Table 5.1.

	Not cured		**Cured for event 1**	**Cured for event 2**
	Event 1	Event 2	Event 2	Event 1
1	1.3% (1.2% - 1.4%)	16.7% (16.6% - 16.8%)	17.2% (17.1% - 17.3%)	20.5% (20.4% - 20.6%)
2	21.7% (21.6% - 21.8%)	8.1% (8.1% - 8.3%)	12.6% (12.5% - 12.7%)	34.4% (34.3% - 34.5%)
3	3.3% (3.2% - 3.4%)	7.9% (7.8% - 8.0%)	32.6% (32.5% - 32.7%)	18.0% (17.8% - 18.1%)

Table 5.3: Percentage of observations per event in the respective groups with respect to the total population size with 95%-confidence intervals for each of the three scenarios in Table 5.1.

The relative group sizes can also be seen in the (simulated) survival plots (Figure 5.2) for the three scenarios. The plots are from one simulation ($M = 1$) with sample size $n = 50.000$ and are indicative of the courses of the survival curves over time. Due to the large sample size, the approximation with the theoretical survival curves is good.

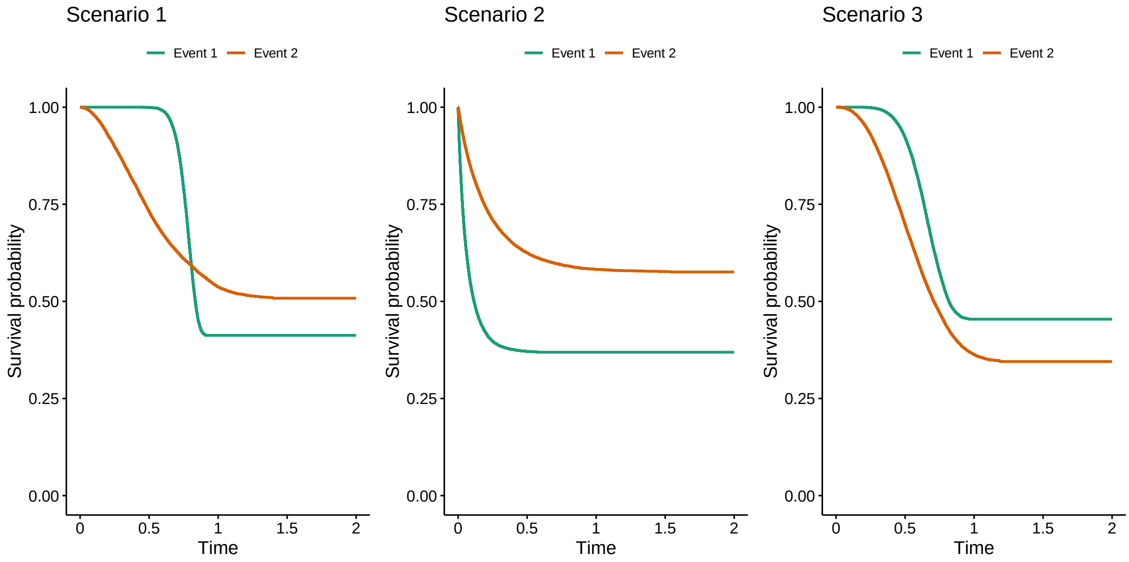

Figure 5.2: Simulated survival curves for the three scenarios.

Since the survival times were truncated at the 99%-quantile of the baseline distribution, there is a positive probability for two survival times to be equal at the truncation point. In order to reduce this probability, the regression coefficients for the latency were chosen positive. The covariate follows an exponential distribution and – thus – has an accelerating effect on the time-to-event. This can be seen in the plot where there are extremely small jumps around the theoretical quantiles. Moreover, this leads to observed cure thresholds which are lower than their (baseline) theoretical counterparts. This can be seen in Table 5.4 where the average simulated observed cure thresholds are shown.

	τ_1	τ_2
1	0.901 (CI: 0.900 - 0.901, true: 0.907)	1.245 (CI: 1.239 - 1.251, true: 1.396)
2	0.518 (CI: 0.514 - 0.522, true: 0.576)	1.118 (CI: 1.104 - 1.132, true: 1.535)
3	0.925 (CI: 0.922 - 0.926, true: 0.965)	1.142 (CI: 1.139 - 1.145, true: 1.187)

Table 5.4: Estimated cure thresholds along with theoretical value and 95%-confidence intervals for each of the three scenarios in Table 5.1.

The two cure thresholds are sufficiently apart in time from each other and this is enough to satisfy the condition stated in Theorem 1. To ensure this condition, it must also be checked that there are uncensored observations of event 2 in between these two cure thresholds. The average percentage of observations of event 2 between the two cure thresholds are shown in Table 5.5 (a). In Table 5.5 (b), the censoring rate for each scenario is shown. In addition the percentage of censored observations after the last cure threshold τ_2 in the plateau is illustrated in Table 5.5 (c).

	Events of type 2
1	2.0% (1.9% - 2.2%)
2	2.2% (2.1% - 2.3%)
3	1.7% (1.6% - 1.7%)

(a)

	Censoring rate
1	44.2% (44.1% - 44.3%)
2	25.1% (25.0% - 25.2%)
3	38.3% (38.1% - 38.4%)

(b)

	Observations after τ_2
1	6.7% (6.7% - 6.8%)
2	9.4% (9.2% - 9.6%)
3	7.9% (7.8% - 8.0%)

(c)

Table 5.5: (a) Number of events of type two observed in between the two cure thresholds, (b) censoring rate and (c) percentage of (censored) observation after the last cure threshold with 95% confidence interval for each of the three scenarios in Table 5.1

5.2.3 Results

In this section, the simulation results for each scenario are presented (see Table 5.6 – 5.8). Bias, variance and mean square error (MSE) are reported for each parameter, together with the true and average estimated value.

n		$\boldsymbol{\gamma_{1_0}}$	$\boldsymbol{\gamma_{1_1}}$	$\boldsymbol{\gamma_{2_0}^0}$	$\boldsymbol{\gamma_{2_1}^0}$	$\boldsymbol{\gamma_{2_0}^1}$	$\boldsymbol{\gamma_{2_1}^1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{S_1(1/4)}$	$\boldsymbol{S_1(1/2)}$	$\boldsymbol{S_1(3/4)}$	$\boldsymbol{S_2(1/4)}$	$\boldsymbol{S_2(1/2)}$	$\boldsymbol{S_2(3/4)}$
500	True	0.25	-0.5	1	-0.5	-0.5	0.5	0.5	1	1	0.999	0.765	0.939	0.702	0.377
	Average	0.480	-0.282	31.684	-7.968	-0.605	0.207	0.494	0.969	1.000	0.999	0.758	0.934	0.673	0.310
	Bias	0.230	0.218	30.684	-7.468	-0.105	-0.293	-0.006	-0.031	0.000	0.000	-0.007	-0.006	-0.029	-0.068
	Variance	0.033	0.093	12454.86	813.013	0.055	0.218	0.009	0.007	0.000	0.000	0.001	0.000	0.001	0.002
	MSE	0.086	0.140	13396.382	868.781	0.066	0.304	0.009	0.008	0.000	0.000	0.001	0.000	0.002	0.006

n		$\boldsymbol{\gamma_{1_0}}$	$\boldsymbol{\gamma_{1_1}}$	$\boldsymbol{\gamma_{2_0}^0}$	$\boldsymbol{\gamma_{2_1}^0}$	$\boldsymbol{\gamma_{2_0}^1}$	$\boldsymbol{\gamma_{2_1}^1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{S_1(1/4)}$	$\boldsymbol{S_1(1/2)}$	$\boldsymbol{S_1(3/4)}$	$\boldsymbol{S_2(1/4)}$	$\boldsymbol{S_2(1/2)}$	$\boldsymbol{S_2(3/4)}$
1000	True	0.25	-0.5	1	-0.5	-0.5	0.5	0.5	1	1	0.999	0.765	0.939	0.702	0.377
	Average	0.495	-0.268	11.492	-2.735	-0.595	0.142	0.485	0.969	1.000	0.999	0.756	0.934	0.674	0.309
	Bias	0.245	0.232	10.492	-2.235	-0.095	-0.358	-0.015	-0.031	0.000	0.000	-0.009	-0.006	-0.028	-0.068
	Variance	0.015	0.066	3595.327	192.759	0.024	0.115	0.004	0.003	0.000	0.000	0.001	0.000	0.001	0.001
	MSE	0.075	0.120	3705.415	197.754	0.033	0.243	0.004	0.004	0.000	0.000	0.001	0.000	0.001	0.006

n		$\boldsymbol{\gamma_{1_0}}$	$\boldsymbol{\gamma_{1_1}}$	$\boldsymbol{\gamma_{2_0}^0}$	$\boldsymbol{\gamma_{2_1}^0}$	$\boldsymbol{\gamma_{2_0}^1}$	$\boldsymbol{\gamma_{2_1}^1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{S_1(1/4)}$	$\boldsymbol{S_1(1/2)}$	$\boldsymbol{S_1(3/4)}$	$\boldsymbol{S_2(1/4)}$	$\boldsymbol{S_2(1/2)}$	$\boldsymbol{S_2(3/4)}$
2500	True	0.25	-0.5	1	-0.5	-0.5	0.5	0.5	1	1	0.999	0.765	0.939	0.702	0.377
	Average	0.491	-0.247	5.485	-1.320	-0.565	0.084	0.488	0.968	1.000	0.999	0.757	0.934	0.675	0.309
	Bias	0.241	0.253	4.485	-0.820	-0.065	-0.416	-0.012	-0.032	0.000	0.000	-0.008	-0.006	-0.027	-0.068
	Variance	0.006	0.045	1590.839	62.180	0.009	0.061	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	MSE	0.065	0.109	1610.957	62.852	0.013	0.233	0.002	0.002	0.000	0.000	0.000	0.000	0.001	0.005

Table 5.6: Simulation results (study I) for scenario 1.

n		$\boldsymbol{\gamma_{1_0}}$	$\boldsymbol{\gamma_{1_1}}$	$\boldsymbol{\gamma_{2_0}^0}$	$\boldsymbol{\gamma_{2_1}^0}$	$\boldsymbol{\gamma_{2_0}^1}$	$\boldsymbol{\gamma_{2_1}^1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{S_1(1/4)}$	$\boldsymbol{S_1(1/2)}$	$\boldsymbol{S_1(3/4)}$	$\boldsymbol{S_2(1/4)}$	$\boldsymbol{S_2(1/2)}$	$\boldsymbol{S_2(3/4)}$
500	True	0.25	-1	-0.5	0.5	1	-1	0.5	0.5	0.135	0.018	0.002	0.472	0.223	0.105
	Average	0.527	-0.880	14.176	-0.804	0.660	-1.122	0.452	0.481	0.099	0.017	0.009	0.342	0.148	0.063
	Bias	0.277	0.120	14.676	-1.304	-0.340	-0.122	-0.048	-0.019	-0.036	-0.001	0.007	-0.131	-0.076	-0.043
	Variance	0.020	0.023	4458.489	225.851	0.052	0.090	0.004	0.021	0.001	0.000	0.000	0.004	0.002	0.001
	MSE	0.097	0.037	4673.885	227.551	0.167	0.105	0.006	0.021	0.002	0.000	0.000	0.021	0.008	0.003

n		$\boldsymbol{\gamma_{1_0}}$	$\boldsymbol{\gamma_{1_1}}$	$\boldsymbol{\gamma_{2_0}^0}$	$\boldsymbol{\gamma_{2_1}^0}$	$\boldsymbol{\gamma_{2_0}^1}$	$\boldsymbol{\gamma_{2_1}^1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{S_1(1/4)}$	$\boldsymbol{S_1(1/2)}$	$\boldsymbol{S_1(3/4)}$	$\boldsymbol{S_2(1/4)}$	$\boldsymbol{S_2(1/2)}$	$\boldsymbol{S_2(3/4)}$
1000	True	0.25	-1	-0.5	0.5	1	-1	0.5	0.5	0.135	0.018	0.002	0.472	0.223	0.105
	Average	0.526	-0.880	9.695	-0.126	0.649	-1.103	0.465	0.478	0.100	0.015	0.005	0.341	0.146	0.060
	Bias	0.276	0.120	10.195	-0.626	-0.351	-0.103	-0.035	-0.022	-0.036	-0.004	0.003	-0.132	-0.077	-0.045
	Variance	0.012	0.012	2274.545	96.059	0.026	0.041	0.002	0.010	0.000	0.000	0.000	0.002	0.001	0.001
	MSE	0.088	0.027	2378.48	96.451	0.149	0.052	0.003	0.011	0.002	0.000	0.000	0.019	0.007	0.003

n		$\boldsymbol{\gamma_{1_0}}$	$\boldsymbol{\gamma_{1_1}}$	$\boldsymbol{\gamma_{2_0}^0}$	$\boldsymbol{\gamma_{2_1}^0}$	$\boldsymbol{\gamma_{2_0}^1}$	$\boldsymbol{\gamma_{2_1}^1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{S_1(1/4)}$	$\boldsymbol{S_1(1/2)}$	$\boldsymbol{S_1(3/4)}$	$\boldsymbol{S_2(1/4)}$	$\boldsymbol{S_2(1/2)}$	$\boldsymbol{S_2(3/4)}$
2500	True	0.25	-1	-0.5	0.5	1	-1	0.5	0.5	0.135	0.018	0.002	0.472	0.223	0.105
	Average	0.525	-0.872	5.829	-0.170	0.636	-1.088	0.464	0.476	0.100	0.013	0.002	0.340	0.145	0.058
	Bias	0.275	0.128	6.329	-0.670	-0.364	-0.088	-0.036	-0.024	-0.035	-0.006	-0.001	-0.132	-0.078	-0.048
	Variance	0.005	0.004	409.098	11.509	0.010	0.017	0.001	0.004	0.000	0.000	0.000	0.001	0.000	0.000
	MSE	0.080	0.021	449.150	11.958	0.142	0.025	0.002	0.005	0.001	0.000	0.000	0.018	0.006	0.002

Table 5.7: Simulation results (study I) for scenario 2.

n		$\boldsymbol{\gamma_{1_0}}$	$\boldsymbol{\gamma_{1_1}}$	$\boldsymbol{\gamma_{2_0}^0}$	$\boldsymbol{\gamma_{2_1}^0}$	$\boldsymbol{\gamma_{2_0}^1}$	$\boldsymbol{\gamma_{2_1}^1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{S_1(1/4)}$	$\boldsymbol{S_1(1/2)}$	$\boldsymbol{S_1(3/4)}$	$\boldsymbol{S_2(1/4)}$	$\boldsymbol{S_2(1/2)}$	$\boldsymbol{S_2(3/4)}$
500	True	-0.5	1	0	1	0	-1	2	0.25	1	0.967	0.5	0.911	0.588	0.232
	Average	-0.091	0.840	2.953	3.656	-0.325	-0.851	1.926	0.180	1.000	0.960	0.469	0.893	0.522	0.170
	Bias	0.409	-0.160	2.953	2.656	-0.325	0.149	-0.074	-0.070	0.000	-0.007	-0.030	-0.018	-0.066	-0.062
	Variance	0.048	0.032	1103.961	256.148	0.059	0.043	0.037	0.003	0.000	0.000	0.003	0.000	0.001	0.001
	MSE	0.215	0.058	1112.683	263.201	0.165	0.065	0.043	0.008	0.000	0.000	0.004	0.001	0.006	0.005

n		$\boldsymbol{\gamma_{1_0}}$	$\boldsymbol{\gamma_{1_1}}$	$\boldsymbol{\gamma_{2_0}^0}$	$\boldsymbol{\gamma_{2_1}^0}$	$\boldsymbol{\gamma_{2_0}^1}$	$\boldsymbol{\gamma_{2_1}^1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{S_1(1/4)}$	$\boldsymbol{S_1(1/2)}$	$\boldsymbol{S_1(3/4)}$	$\boldsymbol{S_2(1/4)}$	$\boldsymbol{S_2(1/2)}$	$\boldsymbol{S_2(3/4)}$
1000	True	-0.5	1	0	1	0	-1	2	0.25	1	0.967	0.5	0.911	0.588	0.232
	Average	-0.080	0.844	0.689	4.234	-0.338	-0.841	1.909	0.160	1.000	0.96	0.465	0.890	0.514	0.161
	Bias	0.420	-0.156	0.689	4.234	-0.338	0.159	-0.091	-0.090	0.000	-0.007	-0.034	-0.020	-0.074	-0.071
	Variance	0.026	0.016	2.215	41.901	0.028	0.021	0.017	0.001	0.000	0.000	0.001	0.000	0.001	0.000
	MSE	0.202	0.04	2.690	52.359	0.142	0.046	0.025	0.009	0.000	0.000	0.003	0.001	0.006	0.005

n		$\boldsymbol{\gamma_{1_0}}$	$\boldsymbol{\gamma_{1_1}}$	$\boldsymbol{\gamma_{2_0}^0}$	$\boldsymbol{\gamma_{2_1}^0}$	$\boldsymbol{\gamma_{2_0}^1}$	$\boldsymbol{\gamma_{2_1}^1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{S_1(1/4)}$	$\boldsymbol{S_1(1/2)}$	$\boldsymbol{S_1(3/4)}$	$\boldsymbol{S_2(1/4)}$	$\boldsymbol{S_2(1/2)}$	$\boldsymbol{S_2(3/4)}$
2500	True	-0.5	1	0	1	0	-1	2	0.25	1	0.967	0.5	0.911	0.588	0.232
	Average	-0.066	0.831	0.667	3.368	-0.353	-0.828	1.911	0.182	1.000	0.961	0.466	0.893	0.523	0.167
	Bias	0.434	-0.169	0.667	2.368	-0.353	0.172	-0.089	-0.068	0.000	-0.007	-0.033	-0.017	-0.066	-0.065
	Variance	0.010	0.006	0.120	7.154	0.012	0.008	0.006	0.001	0.000	0.000	0.001	0.000	0.000	0.000
	MSE	0.199	0.035	0.566	12.763	0.136	0.038	0.014	0.005	0.000	0.000	0.002	0.000	0.005	0.004

Table 5.8: Simulation results (study I) for scenario 3.

5.3 Simulation study II

The second simulation study investigates the estimation performance when the 'complete cure' assumption is made. Complete cure means that when a patient is cured for an arbitrary event, he is simultaneously cured for all events. In Section 3.4.1 it was shown that the cure probabilities and cause-specific hazards are identifiable even without the assumption of independent survival times. The goal of this simulation study is to show that the parameters are also in practice identifiable via the estimation procedure described in Section 4.4.1.

Recall that there is one cure status indicator for all risks simultaneously. Therefore only the incidence parameter γ needs to be estimated. This is the vector with regression coefficients (including an intercept) for the probability of being cured for all events conditional on the covariates. It is introduced in Section 3.4. The parameters that ought to be chosen are the following:

- $\boldsymbol{\alpha_1}$: shape parameter of the Weibull distribution for the baseline survival of event 1;
- $\boldsymbol{\alpha_2}$: shape parameter of the Weibull distribution for the baseline survival of event 2;
- $\boldsymbol{\kappa}$: rate parameter of the Weibull distribution for the baseline survival of both events;
- $\boldsymbol{\beta_1}$: regression coefficients of the Cox model for event 1;
- $\boldsymbol{\beta_2}$: regression coefficients of the Cox model for event 2;
- $\boldsymbol{\gamma}$: regression coefficients of the logistic model for cure probability of both events;
- $\boldsymbol{t}_{\max}$: end-of-study time.

We will consider two scenarios and three sample sizes $n = 500, 100, 2500$. Similar to the previous simulation study, the end-of-study time will be fixed: $t_{\max} = 2$. The two parameter scenarios are chosen as:

Parameters	Scenario 1	Scenario 2
$\boldsymbol{\alpha_1}$	4	5
$\boldsymbol{\alpha_2}$	5	7.5
$\boldsymbol{\kappa}$	1.2	2.5
$\boldsymbol{\beta_1}$	0.5	-0.5
$\boldsymbol{\beta_2}$	-0.25	0.25
$\boldsymbol{\gamma}$	(0,-0.5)	(-0.25 ,0)

Table 5.9: Parameter scenarios used in the simulation study.

The data generation process is different from Simulation Study I. This is partly due to the fact that the survival times are generated dependently. The approach from Beyersmann et al. (2009) is used (see Section 5.1). The shape parameters are chosen to be equal in both scenarios. Therefore, the all-cause hazard simplifies to:

$$\lambda_1(t|x) + \lambda_2(t|x) = \big(\alpha_1 \exp\big\{\beta_1^\top x\big\} + \alpha_2 \exp\big\{\beta_2^\top x\big\}\big)\kappa t^{\kappa-1}. \tag{5.5}$$

This implies that the actual survival time can be simulated using the inverse transform method proved in Lemma 11. Since the all-cause hazard also follows a Weibull distribution. The probability for experiencing event 1 at the simulated actual survival time t is then given by:

$$\frac{\lambda_1(t|x)}{\lambda_1(t|x)+\lambda_2(t|x)} = \frac{\alpha_1 \exp\{\beta_1^\top x\}}{\alpha_1 \exp\{\beta_1^\top x\} + \alpha_2 \exp\{\beta_2^\top x\}} = \frac{\alpha_1}{\alpha_1 + \alpha_2 \exp\{(\beta_2-\beta_1)^\top x\}}. \tag{5.6}$$

The data generation procedure for the complete cure model with dependent potential survival times is summarized in Algorithm 2.

To evaluate the performance of the estimation procedure, the bias, variance and mean square error are computed. Instead of the marginal probabilities on fixed time points, we will estimate the cumulative incidence function for both competing events and compare this to the theoretical value of the cumulative incidence function. The cumulative incidence function for event 1 can be computed as follows:

$$\begin{aligned}
I_1(t) &= \int_0^t \lambda_1^0(du) S_0(u) du \\
&= \int_0^t \alpha_1 \kappa u^{\kappa-1} \exp\left\{ - \int_0^u (\alpha_1+\alpha_2)\, \kappa v^{\kappa-1} dv \right\} du \\
&= \int_0^t \alpha_1 \kappa u^{\kappa-1} \exp\left\{ - (\alpha_1+\alpha_2)\, u^{\kappa} \right\} du \\
&= \frac{\alpha_1}{\alpha_1+\alpha_2}\left(1 - e^{-(\alpha_1+\alpha_2)t^\kappa}\right).
\end{aligned} \tag{5.7}$$

The cumulative incidence function for event 2 can be computed analogously.

Algorithm 2 Simulate time-to-event data for the complete cure

```
set end-of-study time: $t_{\max} = 2$
for $i = 1, 2, ..., n$ do
    generate covariates: $x_i \sim \exp(1)$
    compute cure probability: $\pi(x_i)$
    generate cure status: $B \sim \mathrm{Ber}\big(\pi(x_i)\big)$
    generate censoring time: $C_i \sim \mathrm{U}[0, t_{\max}]$
    if $B = 0$ then
        generate survival probability: $U_i \sim \mathrm{U}(0, 1)$
        compute actual survival time:
            $$T_i \leftarrow \left( - \frac{\log U_i}{\alpha_1 \exp\{\beta_1^\top x_i\} + \alpha_2 \exp\{\beta_2^\top x_i\}} \right)^{1/\kappa}$$
        generate status for event 1:
            $$\delta_i^1 \sim \mathrm{Ber}\left( \frac{\alpha_1}{\alpha_1 + \alpha_2 \exp\{(\beta_2 - \beta_1)^\top x_i\}} \right)$$
        if $\delta_i^1 = 1$ then
            if $T_i > \tau_1$ then
                truncate survival time: $T_i \leftarrow \tau_1$
            end if
            set follow-up time: $T_i^* \leftarrow \min\{C_i, T_i\}$
            if $T_i^* = T_i$ then
                set competing risks indicator: $d_i \leftarrow 1$
            else
                set competing risks indicator: $d_i \leftarrow 0$
            end if
        else
            if $T_i > \tau_2$ then
                truncate survival time: $T_i \leftarrow \tau_2$
            end if
            set follow-up time: $T_i^* \leftarrow \min\{C_i, T_i\}$
            if $T_i^* = T_i$ then
                set competing risks indicator: $d_i \leftarrow 2$
            else
                set competing risks indicator: $d_i \leftarrow 0$
            end if
        end if
    else
        set follow-up time: $T_i^* \leftarrow C_i$
        set competing risks indicator: $d_i \leftarrow 0$
    end if
end for
return competing risks survival data: $(T_i^*, d_i, x_i)$ for $i = 1, 2, ..., n$
```

5.3.1 Data characteristics

The data is relatively simple compared to the simulated data from the previous simulation study. Nevertheless, some descriptive statistics are given to better comprehend the simulated data. The following simulated data characteristics are presented: censoring rate, cure threshold and percentage of censoring in the tail. The procedure for obtaining these characteristics is identical to the previous simulation study.

First, the (theoretical) cumulative incidences (5.3) for each event per scenario can be computed using (5.7).

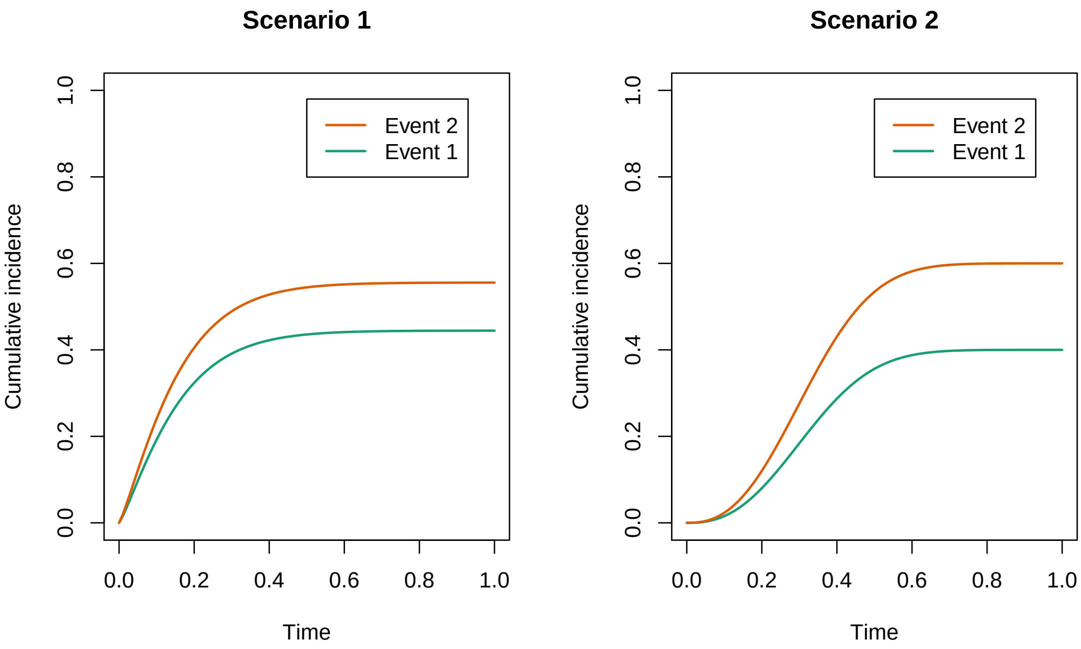

Figure 5.3: Cumulative incidence functions per event for each scenario.

It can be seen from the behaviour of the cumulative incidence functions that the probability of experiencing an event levels off after around 0.6 for the first scenario and around 0.8 for the second. This can also be seen from the simulated cure thresholds. These are given in Table 5.10 (a). The censoring rate and percentage of observations in the tail are given in Table 5.10 (b) – (c).

	τ
1	0.687 (CI: 0.680 - 0.693)
2	0.750 (CI: 0.747 - 0.753)

(a)

	Censoring rate
1	42.5% (42.4% - 42.6%)
2	52.7% (52.6% - 52.8%)

(b)

	Observations after τ
1	25.3% (25.1% - 25.5%)
2	27.5% (27.4% - 27.6%)

(c)

Table 5.10: (a) The cure threshold, (b) censoring rate and (c) percentage of (censored) observation after the cure threshold along with 95% confidence interval per scenario.

5.3.2 Results

Bias, variance and mean squared error are presented for each parameter and for the two estimated cumulative incidence functions at fixed time points. Results show that the estimation procedure is rather robust. The only exception is the γ_1 coefficient for the first combination of parameters.

n		$\boldsymbol{\gamma_0}$	$\boldsymbol{\gamma_1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{I_1(1/4)}$	$\boldsymbol{I_1(1/2)}$	$\boldsymbol{I_2(1/4)}$	$\boldsymbol{I_2(1/2)}$
500	True	0	-0.5	0.5	-0.25	0.364	0.444	0.455	0.555
	Average	-0.049	0.006	0.5	-0.259	0.366	0.449	0.461	0.565
	Bias	-0.049	0.506	0.000	-0.009	0.003	0.005	0.006	0.01
	Variance	0.000	0.000	0.001	0.012	0.002	0.003	0.004	0.003
	MSE	0.003	0.256	0.001	0.012	0.002	0.003	0.004	0.003

n		$\boldsymbol{\gamma_0}$	$\boldsymbol{\gamma_1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{I_1(1/4)}$	$\boldsymbol{I_1(1/2)}$	$\boldsymbol{I_2(1/4)}$	$\boldsymbol{I_2(1/2)}$
1000	True	0	-0.5	0.5	-0.25	0.364	0.444	0.455	0.555
	Average	-0.049	0.006	0.5	-0.254	0.366	0.446	0.46	0.562
	Bias	-0.049	0.506	0.000	-0.004	0.003	0.003	0.005	0.007
	Variance	0.000	0.000	0.000	0.006	0.001	0.001	0.002	0.001
	MSE	0.003	0.256	0.000	0.006	0.001	0.001	0.002	0.002

n		$\boldsymbol{\gamma_0}$	$\boldsymbol{\gamma_1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{I_1(1/4)}$	$\boldsymbol{I_1(1/2)}$	$\boldsymbol{I_2(1/4)}$	$\boldsymbol{I_2(1/2)}$
2500	True	0	-0.5	0.5	-0.25	0.364	0.444	0.455	0.555
	Average	-0.049	0.006	0.499	-0.246	0.368	0.449	0.455	0.555
	Bias	-0.049	0.506	-0.001	0.004	0.005	0.005	0.000	0.000
	Variance	0.000	0.000	0.000	0.003	0.000	0.001	0.001	0.001
	MSE	0.002	0.256	0.000	0.003	0.000	0.001	0.001	0.001

Table 5.11: Simulation results (study II) for scenario 1.

n		$\boldsymbol{\gamma_0}$	$\boldsymbol{\gamma_1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{I_1(1/4)}$	$\boldsymbol{I_1(1/2)}$	$\boldsymbol{I_2(1/4)}$	$\boldsymbol{I_2(1/2)}$
500	True	-0.25	0	-0.5	0.25	0.129	0.399	0.194	0.599
	Average	-0.174	0.01	-0.528	0.252	0.134	0.412	0.195	0.603
	Bias	0.076	0.01	-0.028	0.002	0.004	0.013	0.001	0.004
	Variance	0.001	0.000	0.036	0.001	0.002	0.005	0.001	0.004
	MSE	0.007	0.000	0.037	0.001	0.002	0.005	0.001	0.004

n		$\boldsymbol{\gamma_0}$	$\boldsymbol{\gamma_1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{I_1(1/4)}$	$\boldsymbol{I_1(1/2)}$	$\boldsymbol{I_2(1/4)}$	$\boldsymbol{I_2(1/2)}$
1000	True	-0.25	0	-0.5	0.25	0.129	0.399	0.194	0.599
	Average	-0.175	0.01	-0.515	0.25	0.133	0.407	0.194	0.601
	Bias	0.075	0.01	-0.015	0.000	0.004	0.008	0.000	0.002
	Variance	0.000	0.000	0.016	0.000	0.001	0.002	0.000	0.002
	MSE	0.006	0.000	0.017	0.000	0.001	0.002	0.000	0.002

n		$\boldsymbol{\gamma_0}$	$\boldsymbol{\gamma_1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{I_1(1/4)}$	$\boldsymbol{I_1(1/2)}$	$\boldsymbol{I_2(1/4)}$	$\boldsymbol{I_2(1/2)}$
2500	True	-0.25	0	-0.5	0.25	0.129	0.399	0.194	0.599
	Average	-0.175	0.01	-0.504	0.251	0.129	0.401	0.194	0.601
	Bias	0.075	0.01	-0.004	0.001	0.000	0.002	0.000	0.002
	Variance	0.000	0.000	0.006	0.000	0.000	0.001	0.000	0.001
	MSE	0.006	0.000	0.006	0.000	0.000	0.001	0.000	0.001

Table 5.12: Simulation results (study II) for scenario 2.

Chapter 6

Discussion

In this thesis, identifiability problems which arise when modelling both cure and competing risks have been studied. First theoretical aspects have been investigated where the concept of a cure structure was defined. An estimation procedure is developed to estimate the parameters of the different models. Then, the practical identifiability properties of a specific selection of the models were studied in two simulation studies.

The notion of cure was not properly defined in the context of competing risks. To capture the concept of cure when an individual can fail from multiple events, we defined the cure structure. The cure structure allows for different individuals to be cured from different subsets of competing events and can be represented through the (conditional) cure probabilities π_1, π_2^0 and π_2^1, where π_1, π_2^0 and π_2^1 are, respectively, the probability of being cured for event 1, event 2 conditional on not being cured for event 1 and event 2 conditional on being cured for event 1. Specific choices of cure structure in the literature become special cases of this generalized notion.

Identifiability of the cure competing risks model with independent potential survival times was partly proven (Theorem 1) and partly claimed (Conjecture 3). From a theoretical perspective, the claim seems plausible. However, the simulations may contradict our theoretical intuitions. Simulation Study I showed that under this model there are practical identifiability problems with the parameters related to the incidence. Particularly, the identification of the parameters π_2^0 appeared to be quite hard. Is this π_2^0 parameter not identifiable or do we have practical identification problems possibly related to the estimation procedure? These contradictory statements require further research from both theoretical and practical perspective.

Special attention was paid to the complete cure structure. The complete cure structure is of great practical importance since it is most often used in (clinical) practice. It was proven that the parameters related to the incidence and latency are identifiable in several cases, e.g. when the potential survival times are independent (Theorem 5), and also when the survival times are dependent. In the latter situation, both the subdistribution hazard (Theorem 9) and the cause-specific hazard (Theorem 8) can be uniquely determined from the data. Some of the identification problems still open in the literature (Choi et al., 2015; Zhang et al., 2019) have been addressed and solved by this thesis. In a simulation study, the estimation procedure developed in this work based on the EM algorithm was able to estimate the model parameters in two different settings.

This simulation study is limited to two suitably chosen sets of parameters, one covariate and two competing events. It should be investigated how the estimation procedure performs when a more general competing events model is studied, with more covariates included in the model and other less suitably chosen sets of parameters.

Furthermore, some models were identifiable under strict assumptions. It was shown (Theorem 7) that in general – without the suitable conditions – the parameter related to the cure structure and latency of the uncured were not identifiable. These assumptions are thus necessary, but can sometimes be very restrictive. For example, assuming independence of potential survival times is in many applications, not a realistic assumption. Although the practical evaluation of these assumptions is feasible, it is not always unambiguous. It can be difficult to make a distinction between a plateau in the survival or a very flat tail of the distribution. Misspecification is therefore lurking in the background. It should be investigated how robust the estimation procedures are. This is relevant for further research.

This thesis does not contain a practical application. It would be of great value to apply the theory and estimation procedure developed in this thesis to real-life data. Since identifiability for a general cure structure only holds for independent potential survival times, it might be hard to find a suitable data set. Clinical data often contains competing events that are not independent. The data used by Zhang et al. (2019) might be a good option. It would also allow for a comparison between the estimation procedure developed in this thesis and the one from Zhang et al. (2019)'s paper.

In conclusion, this thesis shed light on the identifiability of the parameters in the context of both theoretical and simulation-based analyses. This research contributes to a better understanding of statistical methods for handling competing risks and cure models. Results from this work are beneficial in clinical and non-clinical settings.

Bibliography

Amico, M. and Van Keilegom, I. (2018). Cure models in survival analysis. *Annual Review of Statistics and Its Application*, 5:311–342.

Austin, P. C. and Fine, J. P. (2017). Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Statistics in Medicine*, 36(27):4391–4400.

Austin, P. C., Steyerberg, E. W., and Putter, H. (2021). Fine-Gray subdistribution hazard models to simultaneously estimate the absolute risk of different event types: cumulative total failure probability may exceed 1. *Statistics in Medicine*, 40(19):4200–4212.

Barlow, W. E. and Prentice, R. L. (1988). Residuals for relative risk regression. *Biometrika*, 75(1):65–74.

Basu, S. and Tiwari, R. C. (2010). Breast cancer survival, competing risks and mixture cure model: a bayesian analysis. *Journal of the Royal Statistical Society. Series A, Statistics in Society*, 173(2):307 – 329.

Berkson, J. and Gage, R. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47:501–515.

Beyersmann, J., Latouche, A., Buchholz, A., and Schumacher, M. (2009). Simulating competing risks data in survival analysis. *Statistics in Medicine*, 28(6):956–971.

Bleyer, W. A. (1990). Acute lymphoblastic leukemia in children. Advances and prospectus. *Cancer*, 65(3 Suppl):689–695.

Boag, J. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society - Series B*, 11:15–53.

Chen, C., Shen, P., Lin, C., and Wu, C. (2020). Semiparametric mixture cure model analysis with competing risks data: Application to vascular access thrombosis data. *Statistics in Medicine*, 40(17):4086–4099.

Choi, S., Huang, X., and Cormier, J. N. (2015). Efficient semiparametric mixture inferences on cure rate models for competing risks. *Canadian Journal of Statistics*, 43(3):420–435.

Choi, S., Zhu, L., and Huang, X. (2017). Semiparametric accelerated failure time cure rate mixture models with competing risks. *Statistics in Medicine*, 37(1):48–59.

Cox, D. R. (1959). The analysis of exponentially distributed lifetimes with 2 types of failure. *Journal of the Royal Statistical Society, Series B*, 21:411–421.

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

Crowder, M. J. (2001). *Classical competing risks.* Chapman Hall/CRC, Boca Raton.

Crowder, M. J. (2012). *Multivariate survival analysis and competing risks.* Chapman Hall/CRC, Boca Raton.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38.

Farewel, V. T. (1977). A model for binary variable with time-censored observations. *Biometrika*, 64(1):43–46.

Farewel, V. T. (1982). The use of a mixture model for the analysis of survival data with long-term survivors. *Biometrics*, 38:1041–1046.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal American Statistical Association*, 94:496 – 509.

Friedman, J. H., Tibshirani, R., and Hastie, T. (2017). *The elements of statistical learning.* Springer-Verlag, New York.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292):698–707.

Hanin, L. and Huang, L. (2014). Identifiability of cure models revisited. *Journal of Multivariate Analysis*, 130:261–274.

Hsu, W. W. and Todem, D., K. K. M. (2016). A sup-score test for the cure fraction in mixture models for long-term survivors. *Biometrics*, 72(4):1348–1357.

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

Koller, T. K., Raatz, H., Steyerberg, E. W., and Wolbers, M. (2012). Competing risks and the clinical community: irrelevance or ignorance? *Statistics in Medicine*, 31(11-12):1089–1097.

Legrand, C. and Betrand, A. (2019). Cure models in cancer clinical trials. *Textbook of Clinical Trials in Oncology: A Statistical Perspective*, page 465–493.

Li, C., Taylor, J. M. G., and Sy, J. P. (2001). Identifiability of cure models. *Science Probability Letters*, 54:389–395.

Maller, R. A. and Zhou, S. (1996). *Survival analysis with long term survivors.* Wiley, New York.

Nicolaie, M. A., Taylor, J. M. G., and Legrand, C. (2019). Vertical modeling: analysis of competing risks data with a cure fraction. *Lifetime Data Analysis*, 25:1–25.

Paoletti, X. and Asselain, B. (2010). Survival analysis in clinical trials: old tools or new techniques. *Surgical Oncology*, 19(2):55–61.

Parsa, M. and Van Keilegom, I. (2023). Accelerated failure time vs Cox proportional hazards mixture cure models: David vs Goliath? *Statistical Papers*, 64(3):835–855.

Peng, Y. and Taylor, J. M. G. (2017). Residual-based model diagnosis methods for mixture cure models. *Biometrics*, 73(2):495–505.

Peng, Y. and Yu, B. (2021). *Cure models: methods, applications, and implementation.* CRC Press, Boca Raton.

Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, 26(11):2389–2430.

Putter, H., Schumacher, M., and Van Houwelingen, H. C. (2020). On the relation between the cause-specific hazard and the subdistribution rate for competing risks data: the Fine–Gray model revisited. *Biometrical Journal*, 62(3):790–807.

R Core Team (2022). *R: a language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rutqvist, L. E., Wallgren, A., and Nilsson, B. (1984). Is breast cancer a curable disease? A study of 14,731 women with breast cancer from the Cancer Registry of Norway. *Cancer*, 53(8):1793–1800.

Sargent, D., Sobrero, A., Grothey, A., O'Connell, M. J., Buyse, M., Andre, T., Zheng, Y., Green, E., Labianca, R., O'Callaghan, C., Seitz, J. F., Francini, G., Haller, D., Yothers, G., Goldberg, R., and De Gramont, A. (2009). Evidence for cure by adjuvant therapy in colon cancer: observations based on individual patient data from 20,898 patients on 18 randomized trials. *Journal of clinical oncology*, 27(6):872–877.

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

Soetaert, K., Hindmarsh, A. C., Eisenstat, S. C., Moler, C., Dongarra, J., and Saad, Y. (2022). *rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations.* R package version 1.8.2.3.

Therneau, T. M., Grambsch, and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160.

Therneau, T. M., Lumley, T., Atkinson, E. G., and Crowson, C. S. (2021). *survival: Survival Analysis.* R package version 3.2-11.

Tong, E. N., Mues, C., and Thomas, L. C. (2014). Mixture cure models in credit scoring: if and when borrowers default. *European Journal of Operation Research*, 218(1):846–857.

Tsiatis, A. (1975). A nonidentifiability asspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the United States of America*, 72(1):20–22.

Van Walraven, C. and McAlister, F. A. (2016). Competing risk bias was common in Kaplan-Meier risk estimates published in prominent medical journals. *Journal of Clinical Epidemiology*, 69:170–173.

Wileyto, E. P., Li, Y., Chen, J., and Heitjan, D. F. (2013). Assessing the fit of parametric cure models. *Biostatics*, 14(2):340–350.

Yilmaz, Y. E., Lawless, J. F., Andrulis, I. L., and Bull, S. B. (2013). Insights from mixture cure modeling of molecular markers for prognosis in breast cancer. *Journal of Clinical Oncology*, 31(16):2047–2054.

Zhang, N., Yang, Q., Kelleher, A., and Wujun, S. (2019). A new mixture cure model under competing risks to score online consumer loans. *Quantitative Finance*, 19(7):1243–1253.

Zhao, Y., Lee, A. H., Yau, K. K. W., Burke, V., and McLachlan, G. J. (2009). A score test for assessing the cured proportion in long-term survivor mixture model. *Statistics in Medicine*, 28(27):3454–3466.

Appendix

This appendix contains all the code used during the project. Most of it was used in the simulation experiments. It is presented in the following fashion. First, the code with all general functions is given. These remain the same across the different simulation experiments. Then the code for the two different simulation studies is given. As these comprise estimation procedures for three different models and thus differ substantially, although their structure may seem identical. The code will also be published on GitHub in the nearby future.

General functions

```
### General functions ###

# This file contains the general functions required for the simulation studies.

# Load the required packages
library(rootSolve)


# Function which prints a message using shell echo.
# Useful for printing messages from inside mclapply when running in Rstudio.
message_parallel <- function(...){
  system(sprintf(’echo "\n%s\n"’, paste0(..., collapse="")))
}


# Risk and tie set functions
risk.set <- function(times, t) which(times >= t)
ties.set <- function(times, t) which(times == t)


# Function for baseline survival for all observed time points.
baseline_survival <- function(bh, times){

  # Compute the baseline survival
  unique_times <- unique(times)
  surv <- exp(-sum(bh[which(times <= times)]))

  return(surv)
}


# Function computes baseline survival for specific choice.
```

```r
baseline_survival_timepoint <- function(bh, times, t){

  # Compute baseline survival for specific time point t.
  unique_times <- unique(times)
  surv <- exp(-sum(bh[which(times <= t)]))

  return(surv)
}


# Compute survival probabilities per individual.
survival_pred <- function(bh, times, status, cov, beta){

  times <- as.vector(times)
  unique_event_times <- unique(times[status == 1])
  baseline_survival <- rep(NA, length(times))

  # Compute survival probabilities per time point.
  bh_unique_times=bh[match(unique_event_times,times)]
  for(i in 1:length(times)){
    baseline_survival[i] <- exp(-sum(bh_unique_times[which(unique_event_times <= times[i])]))
  }

  # Compute the individual survival probabilities based on observed covariates.
  survival <- baseline_survival^(exp( cov %*% beta))
  max_obs <- max(times[which(status==1)])
  survival[which(times>max_obs)] <- 0
  return(survival)
}


# Estimated cumulative incidence function from cause-specific hazard for 2 events.
CIF <- function(t, times, status, bh1, bh2){

  # Define the all-cause baseline hazard.
  times <- as.vector(times)
  unique_event_times <- unique(times[status != 0])
  all_cause_surv <- rep(NA, length(times))
  bh <- bh1 + bh2

  # Only on the unique time points
  bh_unique_times <- bh[match(unique_event_times,times)]

  for(i in 1:length(times)){
    all_cause_surv[i] <- exp(-sum(bh_unique_times[which(unique_event_times < times[i])]))
  }

  # Compute both the cumulative incidence functions.
  p_k1 <- (bh1*all_cause_surv)
  p_k2 <- (bh2*all_cause_surv)
  cif1 <- sum(p_k1[times <= t])
  cif2 <- sum(p_k2[times <= t])

  return(list(cif1 = cif1,
              cif2 = cif2))
}


# Partial log-likelihood with different weights for the Cox model
logpart <- function(beta, times, status, cov, weights1, weights2){

  # Define the uncensored event times
  unique_event_times <- unique(times[which(status != 0)])
  n.event <- length(unique_event_times)

  # Compute the risk and ties set for each uncensored event timepoint
  rs <- lapply(as.matrix(unique_event_times), risk.set, times = times)
  ts <- lapply(as.matrix(unique_event_times), ties.set, times = times)

  # create variable to store values of the loglikelihood
  a <- b <- c <- NA
  temp <- vector()
  cov <- as.matrix(cov)
  weights1 <- as.vector(weights1)
  for(i in 1:n.event){
    a <- sum(((weights1 * cov) %*% beta)[ts[[i]]])
    b <- sum((weights2 * exp(cov %*% beta))[rs[[i]]])
    c <- sum(weights1[ts[[i]]])

    temp[i] <- a -c*log(b)
  }

  return(-sum(temp))
}



# Wrapper function for the computation of cox coefficients
weighted_partial <- function(beta_est, times, status, cov, weights1, weights2){

  # Maximize the log partial likelihood
  suppressWarnings({
    max <- nlm(p = beta_est, f = logpart, cov = cov, times = times,
               status = status, weights1 = weights1, weights2 = weights2)
  })

  return(list(beta = max$estimate))
}



# Define the derivative of the log-likelihood function of the cure statuses
dloglik <- function(gamma, cov, weights1, weights2) {
  cov <- cbind(rep(1, nrow(cov)), cov)
  eta <- cov %*% gamma
  dloglik <- t(cov) %*% (weights1 - (weights1 + weights2) * exp(eta)/(1 + exp(eta))
      )
  return(-dloglik)
}

```

```

# Wrapper function for the optimization of the cure statuses
weighted_IRLS <- function(times, cov, weights1, weights2, gamma_init){

  # Find the roots of the score function
  gamma <- multiroot(f=dloglik, start=gamma_init,
                     cov=cov, weights1 = weights1, weights2 = weights2)$root

  # Return the estimated gamma coefficients
  return(list(gamma = gamma))
}



# Computes the (NA-type) estimator of the baseline hazard with different weights
baseline_hazard <- function(times, status, cov, beta, weights1, weights2) {

  # Initialize an empty vector to store the baseline hazard
  hazard <- rep(0, length(times))

  # Compute the baseline hazard for each unique event time
  for (i in 1:length(times)) {

    if(status[i] != 0){

      # Compute the number of events and weights at the current time
      numerator <- sum(( status*weights1 )[times == times[i]])
      denominator <- sum(( weights2*exp(cov %*% beta) )[times >= times[i]])

      # Compute the baseline hazard at the current time
      hazard[i] <- numerator / denominator

    }
  }

  # Return the baseline hazard at eacht time point
  return(hazard)
}
```

Code for Simulation Study I

```
### Simulation study I ###


# Load the required packages.
library(parallel)
library(survival)
library(rootSolve)


# Function computes the three cure probabilities based on gamma estimates.
cure_pred <- function(cov, gamma1, gamma20, gamma21){

  # Design matrix
  x <- cbind(rep(1, nrow(cov)), cov)

  # Compute the (conditional) cure probabilities per individual.
  pi1 <- plogis(x %*% gamma1)
  pi20 <- plogis(x %*% gamma20)
  pi21 <- plogis(x %*% gamma21)

  return(list(pi1 = pi1, pi20 = pi20, pi21 = pi21))
}


# Function computes the conditional expectations based on current estimates
# in the EM algorithm.
EM_weights <- function(status, surv1, surv2, cure){

  # Denote the estimated cure proportions (vector of size n).
  pi1 <- cure$pi1
  pi20 <- cure$pi20
  pi21 <- cure$pi21

  # Compute the updated weights.
  phi <- as.numeric(status == 0) / (1 + (1-pi1)/pi1 * (pi20*surv1 + (1-pi20)*surv1*
      surv2) / (pi21 + (1-pi21)*surv2) )   +
    as.numeric(status == 2) / (1 + (1-pi1)*(1-pi20)*surv1/(pi1*(1-pi21)))
  psi0 <- as.numeric(status != 2)* pi20 / (pi20 + (1-pi20)*surv2)
  psi1 <- as.numeric(status != 2)* pi21 / (pi21 + (1-pi21)*surv2)

  # Might create NaN; due to 0/0.
  phi[is.na(phi)] <-  0
  psi0[is.na(psi0)] <- 0
  psi1[is.na(psi1)] <- 0

  return(list(phi = phi, psi0 = psi0, psi1 = psi1))
}


# Estimate the model parameters using the EM algorithm.
em <- function(times, CR_status, cov, eps, emmax){

  # Define the different status indicators. (The argument status is assumed to
  # to contain 0, 1 and 2.)
  s0 <- as.numeric(CR_status != 0)
  s1 <- as.numeric(CR_status == 1)
  s2 <- as.numeric(CR_status == 2)

  # Initialize the parameters.
  gamma1 <- gamma20 <- gamma21 <- c(0,0)
  beta1 <- beta2 <- 0
  bh1 <- baseline_hazard(times, s1, cov, beta1,
                         rep(1, length(times)), rep(1,length(times)))
  bh2 <- baseline_hazard(times, s2, cov, beta2,
                         rep(1,length(times)), rep(1,length(times)))

  # Keep track of the convergence of the algorithm.
  i <- 1
  convergence <- 100

  while (convergence > eps & i < emmax){

    # Compute the estimated conditional expectations given the current estimates.
    cure <- cure_pred(cov, gamma1, gamma20, gamma21)
    surv1 <- survival_pred(bh1, times, s1, cov, beta1)
    surv2 <- survival_pred(bh2, times, s2, cov, beta2)
    cond_exp <- EM_weights(CR_status, surv1, surv2, cure)
    phi <- cond_exp$phi
    psi0 <- cond_exp$psi0
    psi1 <- cond_exp$psi1


    # Define the different weights for each part of the model.
    wp1 <- phi*( 1-s0 + s2*(1-psi1) )
    wp1C <- (1-phi)*( 1-s0 + s1 + s2*(1-psi0) )
    wp21 <- (1-s0)*phi*psi1
    wp21C <- phi*( (1-s0)*(1-psi1) + s2*(1-psi1) )
    wp20 <- (1-phi)*psi0*( 1-s0 + s1 )
    wp20C <- (1-phi)*(1-psi0)*( 1-s0 + s1 + s2 )
    wS1 <- (1-s0)*(1-phi) + s1*(1-phi) + s2*(1-phi)*(1-psi0)
    wL1 <- s1*(1-phi)
    wS2 <- (1-s0)*( phi*(1-psi1) + (1-phi)*(1-psi0) ) + s1*(1-phi)*(1-psi0) +
      s2*( (1-phi)*(1-psi0) + phi*(1-psi1) )
    wL2 <- s2*(phi*(1-psi1) + (1-phi)*(1-psi0))


    # Store the old parameters before updating them.
    par_old <- c(gamma1,gamma20,gamma21,beta1,beta2,bh1,bh2)

    # Estimate the new parameters from the updated weigths.
    gamma1 <- weighted_IRLS(times, cov, wp1, wp1C, gamma1)$gamma
    gamma20 <- weighted_IRLS(times, cov, wp20, wp20C, gamma20)$gamma
    gamma21 <- weighted_IRLS(times, cov, wp21, wp21C, gamma21)$gamma
    beta1 <- weighted_partial(beta1, times, s1, cov, wS1, wL1)$beta
    beta2 <- weighted_partial(beta2, times, s2, cov, wS2, wL2)$beta
    bh1 <- baseline_hazard(times, s1, cov, beta1, wS1, wL1)
    bh2 <- baseline_hazard(times, s2, cov, beta2, wS2, wL2)

    # Compute the distance to the old parameters.
    convergence <- sum((par_old-c(gamma1,gamma20,gamma21,beta1,beta2,bh1,bh2))^2)
    i <- i + 1
  }

  # Estimate baseline survival for reporting the results of the simulation.
  S1.0.25 <-  baseline_survival_timepoint(bh1, times, 0.25)
  S1.0.50 <-  baseline_survival_timepoint(bh1, times, 0.50)
  S1.0.75 <-  baseline_survival_timepoint(bh1, times, 0.75)
  S2.0.25 <-  baseline_survival_timepoint(bh2, times, 0.25)
  S2.0.50 <-  baseline_survival_timepoint(bh2, times, 0.50)
  S2.0.75 <-  baseline_survival_timepoint(bh2, times, 0.75)
  surv <- c(S1.0.25, S1.0.50, S1.0.75, S2.0.25, S2.0.50, S2.0.75)

  return(list(times1 = sort(unique(times[which(CR_status == 1)])),
              times2 = sort(unique(times[which(CR_status == 2)])),
              times = times,
              bh1 = bh1,
              bh2 = bh2,
              gamma1 = gamma1,
              gamma20 = gamma20,
              gamma21 = gamma21,
              beta1 = beta1,
              beta2 = beta2,
              conv=convergence,
              it=i,
              surv = surv))
}


# Simulate the cure competing risks data and fit the model.
sim_cure <- function(n, gamma1, gamma20, gamma21, beta1, beta2, kappa1, kappa2, alpha1, alpha2){

  # Simulate the covariates and define a design matrix.
  x0 <- rep(1, n)
  x1 <- rexp(n, 1)
  X <- matrix(c(x0, x1), ncol = 2)
  Z <- matrix(c(x1), ncol = 1)

  # Generate the cure status for risk 1.
  B1 <- rbinom(n, 1, plogis(X %*% gamma1))

  # Generate cure status for risk 2 conditional on risk 1.
  B2 <- vector(length = n)
  B2[which(B1 == 0)] <- rbinom(sum(B1 == 0), 1, plogis(X %*% gamma20)[which(B1 == 0)])
  B2[which(B1 == 1)] <- rbinom(sum(B1 == 1), 1, plogis(X %*% gamma21)[which(B1 == 1)])

  # Generate survival times using Cox model with Weibull baseline survival.
  u <- runif(n)
  v <- runif(n)
  T1 <- (-log(u)/(alpha1*exp( Z %*% beta1 )))^(1/kappa1)
  T2 <- (-log(v)/(alpha2*exp( Z %*% beta2 )))^(1/kappa2)

  # Set the cure thresholds at the 99% quantiles (truncate the survival times).
  tau1 <- qweibull(0.99, kappa1, scale=alpha1^(-1/kappa1))
  tau2 <- qweibull(0.99, kappa2, scale=alpha2^(-1/kappa2))
  T1 <- pmin(T1, tau1)
  T2 <- pmin(T2, tau2)
  T1[which(B1 == 1)] <- 100  #cured patients have infinite survival times
  T2[which(B2 == 1)] <- 100

  # Event time and CR status indicator.
  T <- pmin(T1, T2)
  D <- 1 + as.numeric(T2 <= T1)

  # Censoring and follow-up time (with truncation)
  tau <-  2  #end of study
  C <- runif(n,min(tau1,tau2),tau) #rexp(n, 0.1)
  C <- pmin(C, rep(tau, n))  #truncate censoring times
  Y <- pmin(T, C)
  status <- as.numeric(C>=T)
  D <- status * D

  # Estimate model parameters for the simulated data using the EM algorithm.
  result <- em(times = Y, CR_status = D, cov = Z, eps = 10^(-5), emmax = 500)

  return(list(times = Y,
              CR_status = D,
              cov = Z,
              cure_status1 = B1,
              cure_status2 = B2,
              cens_rate = cens_rate,
              cure_stats = cure_stats,
              plateau = plateau,
              tau1 = tau1,
              tau2 = tau2,
              gamma1 = result$gamma1,
              gamma20 = result$gamma20,
              gamma21 = result$gamma21,
              beta1 = result$beta1,
              beta2 = result$beta2,
              conv = result$convergence,
              it = result$i,
              surv = result$surv))
}


# Perform the simulation study and keep track of the results.
MC_cure2 <- function(X, n, gamma1, gamma20, gamma21, beta1, beta2, kappa1, kappa2, alpha1, alpha2){

  estimates <- matrix(ncol = 14, nrow = 1)
  colnames(estimates) <- c("gamma1 (intercept)", "gamma1 (coef)",
                           "gamma20 (intercept)", "gamma20 (coef)",
                           "gamma21 (intercept)", "gamma21 (coef)",
                           "beta1", "beta2",
                           "S1(0.25)", "S1(0.50)", "S1(0.75)",
                           "S2(0.25)", "S2(0.50)", "S2(0.75)")

  # Perform the simulation until it converges. This usually takes 1 iteration.
  while (error) {
    result <- tryCatch(sim_cure(n, gamma1, gamma20, gamma21, beta1, beta2, kappa1, kappa2, alpha1, alpha2),
                       error=function(e) {
                         print("ERROR")
                         return(NA)
                       },
                       warning=function(w) {
                         print("WARNING")
                       }
    )
    if(length(result) == 18){
      error <- FALSE

    }
  }

  #message_parallel(c("n ", n, " Iteration ", X, " Conv: ", result$it))

  estimates[1,] <- c(result$gamma1, result$gamma20, result$gamma21, result$beta1, result$beta2, result$surv)

  return(estimates)
  print(c("Number of errors:", N_e))
}


##### Code for actually performing the simulations #####


# Parameter combination 1
gamma1 <- c(1/4, -1/2)
gamma20 <- c(1, -0.5)
gamma21 <- c(-0.5, 0.5)
beta1 <- c(1/2)
beta2 <- c(1)
kappa1 <- 15
kappa2 <- 2.5
alpha1 <- 20
alpha2 <- 2


# Parameter combination 2
gamma1 <- c(1/4, -1)
gamma20 <- c(-0.5, 0.5)
gamma21 <- c(1, -1)
beta1 <- c(0.5)
beta2 <- c(0.5)
kappa1 <- 1
kappa2 <- 1
alpha1 <- 8
alpha2 <- 3


#Parameter combination 3
gamma1 <- c(-1/2, 1)
gamma20 <- c(0, 1)
```

```r
gamma21 <- c(0, -1)
beta1 <- c(2)
beta2 <- c(0.25)
kappa1 <- 7.5
kappa2 <- 2.5
alpha1 <- 6
alpha2 <- 3


# Compute true survival probabilities.
surv_true <- c(exp(-alpha1*0.25^kappa1), exp(-alpha1*0.5^kappa1), exp(-alpha1*0.75^kappa1),
               exp(-alpha2*0.25^kappa2), exp(-alpha2*0.5^kappa2), exp(-alpha2*0.75^kappa2))

# Set simulation size, sample size and number of cores:
n <- 1000
M <- 1000
cores <- system("nproc", intern=TRUE)
print(paste("Using ",cores, " cores"))

# THE SIMULATION
estimates <- mcmapply(X = 1:M, FUN = MC_cure2,
                      MoreArgs = list(gamma1 = gamma1,
                                      gamma20 =gamma20,
                                      gamma21 = gamma21,
                                      n = n,
                                      beta1=beta1, beta2=beta2,
                                      kappa1=kappa1, kappa2=kappa2,
                                      alpha1=alpha1, alpha2=alpha2),
                      mc.cores=cores)


# Present the results in the correct format.
estimates <- t(estimates)
summary <- matrix(ncol = 14, nrow = 5)
colnames(summary) <- c("gamma1 (intercept)", "gamma1 (coef)",
                       "gamma20 (intercept)", "gamma20 (coef)",
                       "gamma21 (intercept)", "gamma21 (coef)",
                       "beta1", "beta2",
                       "S1(0.25)", "S1(0.50)", "S1(0.75)",
                       "S2(0.25)", "S2(0.50)", "S2(0.75)")
rownames(summary) <- c("True", "MC estimate", "Bias", "Variance", "MSE")
summary[1,] <- c(gamma1, gamma20, gamma21, beta1, beta2,surv_true)
summary[2,] <- colMeans(estimates, na.rm = TRUE)
summary[3,] <- colMeans(estimates, na.rm = TRUE) - c(gamma1, gamma20, gamma21, beta1, beta2, surv_true)
summary[4,] <- colVars(estimates, na.rm = TRUE)
summary[5,] <- summary[4,] + (summary[3,])^2

print(paste("PARAM-COMBI 1 /// n = ", n, "  /// M = ", M, "/// note: "))
print(summary)
```

Code for Simulation Study II

```r
### Simulation study II ###


# Load the required packages.
library(dplyr)
library(matrixStats)
library(parallel)
library(survival)


# Function computes the three cure probabilities based on gamma estimates.
cure_pred <- function(cov, gamma){

  # Design matrix
  x <- cbind(rep(1, nrow(cov)), cov)

  # Compute the (conditional) cure probabilities per individual.
  pi <- plogis(x %*% gamma)

  return(list(pi = pi))
}


# Function computes the conditional expectations based on current estimates
# in the EM algorithm.
EM_weights <- function(status, surv1, surv2, cure){

  # Denote the estimated cure proportions (vector of size n).
  pi <- cure$pi

  # Compute the updated weights.
  phi <- as.numeric(status == 0)* pi / (pi + (1-pi)*surv1*surv2)

  # Might create NaN; due to 0/0.
  phi[is.na(phi)] <-  0

  return(list(phi = phi))
}


# Estimate the model parameters using the EM algorithm.
em <- function(times, CR_status, cov, eps, emmax){

  # Define the different status indicators. (The argument status is assumed to
  # to contain 0, 1 and 2.)
  s0 <- as.numeric(CR_status != 0)
  s1 <- as.numeric(CR_status == 1)
  s2 <- as.numeric(CR_status == 2)

  # Initialize the parameters.
  gamma <- c(0,0)
  beta1 <- unname(coxph(Surv(times, s1) ~ cov)$coef)
  beta2 <- unname(coxph(Surv(times, s2) ~ cov)$coef)
  bh1 <- baseline_hazard(times, s1, cov, beta1,
                         rep(1, length(times)), rep(1,length(times)))
  bh2 <- baseline_hazard(times, s2, cov, beta2,
                         rep(1,length(times)), rep(1,length(times)))

  # Keep track of the convergence of the algorithm.
  i <- 1
  convergence <- 100

  while (convergence > eps & i < emmax){

    # Compute the estimated conditional expectations given the current estimates
    cure <- cure_pred(cov, gamma)
    surv1 <- survival_pred(bh1, times, s1, cov, beta1)
    surv2 <- survival_pred(bh2, times, s2, cov, beta2)
    cond_exp <- EM_weights(CR_status, surv1, surv2, cure)
    phi <- cond_exp$phi

    # Store the old parameters before updating them.
    par_old <- c(gamma,beta1,beta2,bh1,bh2)

    # Estimate the new parameters from the updated weigths.
    gamma <- weighted_IRLS(times, cov, s0*(1-phi), (1-phi), gamma)$gamma
    beta1 <- weighted_partial(beta1, times, s1, cov, (1-phi), (1-phi))$beta
    beta2 <- weighted_partial(beta2, times, s2, cov, (1-phi), (1-phi))$beta
    bh1 <- baseline_hazard(times, s1, cov, beta1, (1-phi), (1-phi))
    bh2 <- baseline_hazard(times, s2, cov, beta2, (1-phi), (1-phi))

    # Compute the distance to the old parameters.
    convergence <- sum((par_old-c(gamma,beta1,beta2,bh1,bh2))^2)
    i <- i + 1
  }

  # Estimate cumulative incidences for reporting the results of the simulation.
  CIF1.0.25 <- CIF(0.25, times, CR_status, bh1, bh2)$cif1
  CIF1.0.50 <- CIF(0.75, times, CR_status, bh1, bh2)$cif1
  CIF2.0.25 <- CIF(0.25, times, CR_status, bh1, bh2)$cif2
  CIF2.0.50 <- CIF(0.75, times, CR_status, bh1, bh2)$cif2
  CIF_estimate <- c(CIF1.0.25, CIF1.0.50, CIF2.0.25, CIF2.0.50)

  return(list(times1 = sort(unique(times[which(CR_status == 1)])),
              times2 = sort(unique(times[which(CR_status == 2)])),
              times = times,
              bh1 = bh1,
              bh2 = bh2,
              gamma = gamma,
              beta1 = beta1,
              beta2 = beta2,
              CIF = CIF_estimate,
              conv=convergence,
              it=i))
}


# Simulate the cure competing risks data and fit the model.
sim_cure <- function(n, gamma, beta1, beta2, kappa, alpha1, alpha2){

  # Simulate the covariates and define a design matrix.
```

```r
  x0 <- rep(1, n)
  x1 <- rexp(n)
  X <- matrix(c(x0, x1), ncol = 2)
  Z <- matrix(c(x1), ncol = 1)

  # Generate the cure status for risk 1.
  B <- rbinom(n, 1, plogis(X %*% gamma))

  # Generate survival times using Cox model with Weibull baseline survival.
  u <- runif(n)
  T <- (-log(u)/(alpha1*exp( Z %*% beta1 ) + alpha2*exp( Z %*% beta2)))^(1/kappa)

  #generate event status
  D <- rep(2, n)
  event1 <- rbinom(n, 1, alpha1/(alpha1 + alpha2*exp( Z %*% (beta2 - beta1) )))
  D <- D - event1

  # Set the cure thresholds at the 99% quantiles (truncate the survival times).
  tau1 <- qweibull(0.99, kappa, scale=alpha1^(-1/kappa))
  tau2 <- qweibull(0.99, kappa, scale=alpha2^(-1/kappa))
  T[D == 1] <- pmin(T[D == 1], tau1)
  T[D == 2] <- pmin(T[D == 2], tau2)
  T[which(B == 1)] <- 100  #cured patients have infinite survival times

  # Censoring and follow-up time (with truncation)
  tau <-  2  #end of study
  C <- runif(n, 0, tau)
  C <- pmin(C, rep(tau, n))  #truncate censoring times
  Y <- pmin(T, C)
  D[Y == C] <- 0

  # Estimate model parameters for the simulated data using the EM algorithm.
  result <- em(times = Y, CR_status = D, cov = Z, eps = 10^(-8), emmax = 500)

  return(list(times = Y,
              CR_status = D,
              cov = Z,
              tau1 = tau1,
              tau2 = tau2,
              gamma = result$gamma,
              beta1 = result$beta1,
              beta2 = result$beta2,
              CIF = result$CIF,
              conv = result$convergence,
              it = result$i))
}


# Perform the simulation study and keep track of the results.
MC_cure2 <- function(X, n, gamma, beta1, beta2, kappa, alpha1, alpha2){

  estimates <- matrix(ncol = 8, nrow = 1)
  colnames(estimates) <- c("gamma (intercept)", "gamma (coef)",
                           "beta1", "beta2",
                           "CIF1(0.25)", "CIF1(0.50)", "CIF2(0.25)", "CIF2(0.50)")

  # Store the results.
  result <- sim_cure(n, gamma, beta1, beta2, kappa, alpha1, alpha2)
  estimates[1,] <- c(result$gamma, result$beta1, result$beta2, result$CIF)

  return(estimates)
}


##### Code for actually performing the simulations #####


# Parameter combination 1
gamma <- c(0, -0.5)
beta1 <- c(0.5)
beta2 <- c(-0.25)
kappa <- 6/5
alpha1 <- 4
alpha2 <- 5


# Parameter combination 2
gamma <- c(-1/4, 0)
beta1 <- c(-0.5)
beta2 <- c(1/4)
kappa <- 2.5
alpha1 <- 5
alpha2 <- 7.5


# Compute the theoretical values of the cumulative incidence fucntions.
true_CIF1 <- function(t, alpha1, alpha2, kappa){
  CIF <- (alpha1/(alpha1+alpha2))*(1-exp(-(alpha1+alpha2)*t^kappa))
  return(CIF)
}

true_CIF2 <- function(t, alpha1, alpha2, kappa){
  CIF <- (alpha2/(alpha1+alpha2))*(1-exp(-(alpha1+alpha2)*t^kappa))
  return(CIF)
}

Theor_CIF <- c(true_CIF1(0.25, alpha1, alpha2, kappa),
               true_CIF1(0.75, alpha1, alpha2, kappa),
               true_CIF2(0.25, alpha1, alpha2, kappa),
               true_CIF2(0.75, alpha1, alpha2, kappa))


# Set simulation parameters
n <- 1000
M <- 1000
cores <- system("nproc", intern=TRUE)
print(paste("Using ",cores, " cores"))


# THE SIMULATION STUDY
estimates <- mcmapply(X = 1:M, FUN = MC_cure2,
                      MoreArgs = list(n = n,
                                      gamma = gamma,
                                      beta1 = beta1,
                                      beta2 = beta2,
                                      kappa = kappa,
                                      alpha1 = alpha1,
                                      alpha2 = alpha2),
                      mc.cores=cores)


# Present the results in the correct format.
estimates <- t(estimates)
summary <- matrix(ncol = 8, nrow = 5)
colnames(summary) <- c("gamma (intercept)", "gamma (coef)",
                       "beta1", "beta2",
                       "CIF1(0.25)", "CIF1(0.50)", "CIF2(0.25)", "CIF2(0.50)")
rownames(summary) <- c("True", "MC estimate", "Bias", "Variance", "MSE")
summary[1,] <- c(gamma, beta1, beta2, Theor_CIF)
summary[2,] <- colMeans(estimates, na.rm = TRUE)
summary[3,] <- colMeans(estimates, na.rm = TRUE) - c(gamma,beta1,beta2,Theor_CIF)
summary[4,] <- colVars(estimates, na.rm = TRUE)
summary[5,] <- summary[4,] + (summary[3,])^2

print(summary)
print(paste("PARAM-COMBI 1 /// n =", n, " /// M =", M))
```