Dynamic Aspects of Competing Risks with Application to Medical Data

Mioara Alina Nicolaie

Cover design: Danielle de Laminne de Bex, Genval, Belgium
Printed by: Off Page

ISBN: 978-94-6182-384-7

Research leading to this thesis was supported by the Netherlands Organization for Scientific Research Grant ZONMW-912-07-018'Prognostic modeling and dynamic prediction for competing risks and multi-state models'.

Dynamic Aspects of Competing Risks with Application to Medical Data

Proefschrift

ter verkrijging van de graad van Doctor aan de Universiteit Leiden, op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker, volgens besluit van het College voor Promoties te verdedigen op woensdag 8 januari 2014 klokke 13:45 uur

door

Mioara Alina Nicolaie
geboren te Braşov, Roemenië
in 1977

PROMOTIECOMMISSIE

Promotores:
Prof.dr. H. Putter
Em. prof.dr. H. C. van Houwelingen

Overige leden:
Prof.dr. F. Willekens, Max Planck Institute for Demographic Research, Rostock, Germany
Prof.dr. C. Legrand, Catholic University of Louvain, Louvain-la-Neuve, Belgium
Dr. R. B. Geskus, Academic Medical Center, Amsterdam, The Netherlands

Table of Contents

1 Introduction

Dynamic Aspects of Competing Risks with Application to Medical Data

Mioara Alina Nicolaie

1.1 Introduction

Analysis of lifetime duration, often termed survival analysis, is a topic of broad interest; its applications overlay a variety of fields such as medical sciences, economics, engineering, demography and social sciences. Our particular preference concerns medical applications; for this reason, the terminology used in the following will involve aspects of life.

The focus of this thesis is on inference in survival models, a subject of topical interest nowadays. This analytical methodology aims at describing the outcome *survival*, which refers to having experienced death or any type of clinical event which hampers the life course of an individual; examples of situations which are addressed by it will be given in the remainder of the thesis. Survival data consists of 1. time to occurrence of an event of interest (e.g., patient lifetime in case the event is death of the patient), which we call *survival time* or *time-to-event* variable, and 2. some additional clinical information believed to be relevant to the outcome of interest, which is incorporated in what we call *covariates* (also called predictors). The time-to-event variable can be partially or fully observed during the follow-up period, depending on whether the patient is lost to follow-up or not; the latter situation occurs due to a certain *censoring* mechanism. The use of statistical models for the analysis of time-to-event data has been extensively addressed in the biostatistical literature; nevertheless, challenges provided by clinical questions lead researchers to account for broader aspects of the data. Some may suspect the presence of several, competing endpoints of interest in the data, such that the occurrence of one precludes the occurrence of others; we shall refer to this as survival data as *competing risks*. In competing risks, patient lifetime (uniquely defined) and time to a specific event (one event out of several) are, in principle, two distinct concepts but they overlap in reality, because what we observe as patient lifetime is nothing but the time to the very first event occurring; time to any of the competing events becomes a latent survival time.

Clinical information collected at the start of the follow-up may be taken into account in modeling (e.g. type of treatment administered, type of surgery, specific disease markers). However, patients might respond differently to certain clinical conditions set at the beginning of the experiment/clinical trial, and this updated information might be relevant to the outcome of interest. To put things into perspective, we anticipate the use of patient history in the analysis and we shall quantify it through so-called *time-dependent* covariates.

The outline of the Introduction is as follows: in Section 1.1.1 we give an overview of the main concepts used in competing risks and introduce some inference methods for competing risks, relied on in the remainder of the thesis. In Section 1.1.2 we motivate the need for an optimal quality of the information on causes of death. In Section 1.1.3 we provide some background on the topic of

dynamic prediction.

1.1.1 Approaches to survival data with competing risks

Competing risks data consist of subjects who are susceptible to several types of events j, where $j = 1, \ldots, J$; we assume there are no other causes aside this specified group. Typically, what we observe for a particular patient is the realization of the time-to-event variable T, where T represents either the time to the first event occurring or time to censoring, the realization of the indicator δ of the cause of failure j ($\delta = j$) or censoring ($\delta = 0$), and the realizations of some baseline or time-dependent covariates $\mathbf{Z}$. A competing risks model can be used to analyse this type of data (Putter et al., 2007).

A key assumption in these models is that the independent censoring mechanism applies, that is the survival time and the (potential) censoring time are independent, given covariates (Lawless, 2003, p.54). Note that this condition may hardly be checked given the typical observable data. We also assume accurate measurements of the covariates and reliable diagnosis of the causes of failure. Carrol et al. (2006) give a comprehensive treatment of the topic of measurement errors in covariates and their consequences in survival analysis. In van Rompaye et al. (2010) the case is discussed when one can not rely on the cause of death ascertainment.

Basically, a competing risks model can be viewed as adding structure to the standard survival model. The standard survival model specifies the (overall) *survival function* $S(t) = P(T > t)$, that is one minus the distribution function of the time-to-event variable T. The clinical interpretation of survival function has a strong appeal to physicians; typical clinical questions addressed by it could be:

Doctor: *What is the duration of an illness*?

or

Patient: *Doctor, (when) will I be cured*?

The survival function is interpreted at a population level, and it is obtained by averaging the individual survival functions. In competing risks, the distribution of the time-to-event can specialize on the type of events; however, it is possible to collect all failure types into a single category (an "all-causes" category) such that we preserve the interpretation of the time-to-event distribution as defined in ordinary survival (time to a specific event in competing risks becomes time to "all-causes"). One way to estimation is to specify a non-/semi-parametric model for the survival function. A famous non-parametric estimator is the Kaplan-Meier estimator (Kaplan and Meier, 1958), that is

$$\widehat{S}(t) = \prod_{t_i \leq t} (1 - \frac{d_i}{n_i})$$

where the product is taken over all event time points $t_i \leq t$, d_i is the number of failures at time t_i irrespective of their cause and n_i is the number of patients in the risk set $\mathcal{R}(t_i)$ at time t_i. Yet another perspective on the modeling and estimation of S is obtained if we notice that the survival function is completely specified by the hazard function of "all-causes", that is

$$S(t) = \exp\Big[-\int_0^t \lambda(u)du\Big] \tag{1.1}$$

where the hazard function of "all-causes" is given by

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} \tag{1.2}$$

being interpreted as the instantaneous failure rate at time t irrespective of its type, given no event before time t. As a consequence, a regression model specified for the hazard of "all-causes" implicitly leads to a model for the survival function.

If we account for the presence of J competing risks, (1.2) becomes the so-called *cause-specific hazard* of cause j, that is

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t, D = j | T \geq t)}{\Delta t}, \qquad j = 1, \ldots, J. \tag{1.3}$$

Note that $\lambda(t) = \sum_{j=1}^{J} \lambda_j(t)$. The $\lambda_j(\cdot)$'s allow us to specify competing risks as a process-based model. Cortese and Andersen (2010) have studied these models in a variety of contexts. Basically, we can view competing risks as a Markov process, whose state space comprises a transient state (the state at time 0) and several absorbing states (the competing events) and whose transition intensities are the cause-specific hazards.

Because the $\lambda_j(\cdot)$'s are the natural observable quantities in competing risks and can be estimated directly from the data, the cause-specific hazard approach has been the most widely used procedure to the analysis of competing risks data in medical research (Prentice et al., 1978). There have been various parametric, semi-parametric or non-parametric models for the cause-specific hazards, among which the most popular is the Cox proportional hazards semi-parametric regression model (Cox, 1972) given by

$$\lambda_j(t|\mathbf{Z}) = \lambda_{j0}(t) \exp(\beta_j^\top \mathbf{Z}), \tag{1.4}$$

where $\lambda_{j0}(\cdot)$ is a baseline hazard function, β_j stands for a vector of unknown regression parameters and $\mathbf{Z}$ is a vector of covariates; the baseline hazard is interpreted as the hazard corresponding to $\mathbf{Z} = 0$ and β_j's stand for the effects of covariates on the hazard scale, $j = 1, \ldots, J$. In this model, no shape on $\lambda_{j0}(\cdot)$ is assumed and the effects of covariates, called *hazard ratios*, are the same at all

time points t. A typical relaxation of the proportionality assumption found in the literature consist of allowing time-varying effects of covariates ($\beta_j = \beta_j(t)$), i.e., a deviation from the proportionality assumption. Typical clinical questions addressed by the model (1.4) could be:

Doctor: *Does this clinical study show that the new drug brings benefits in terms of preventing disease-specific deaths*?
or

Patient: *Doctor, what are the chances of dying from the disease within* 5 *years with this new drug compared with the standard drug*?

Denote $\beta = (\beta_j)_{j=1,\dots,J}$ and $\gamma = (\lambda_{j0})_{j=1,\dots,J}$; we gather them in $\theta = (\beta, \gamma)$, the multidimensional vector of parameters in the competing risks model. Estimation of θ is usually done by means of maximum likelihood procedures. The full likelihood, under the assumption of independent censoring mechanism, can be expressed in terms of cause-specific hazards:

$$\begin{aligned}\mathcal{L}(\theta) &= \prod_i \prod_j \big[\lambda_j(t_i)\big]^{1\{\delta_i=j\}} S(t_i) \\ &= \prod_i \prod_j \big[\lambda_j(t_i)\big]^{1\{\delta_i=j\}} \exp\big(-\sum_{j=1}^{J} \int_0^{t_i} \lambda_j(u)du\big).\end{aligned}$$

This likelihood is proportional to the joint probability distribution of $(t_i, \delta_i, \mathbf{Z}_i)$, for all patients i. In principle, to obtain an estimator for θ, we can calculate the maximizer of the $\log(\mathcal{L}(\theta))$, but in practice this can be quite difficult to obtain. Instead, a different method can be used, which may lead to an easier estimation procedure.

Cox (1975) has introduced a modified likelihood, called *partial likelihood*, in order to avoid estimation of nuisance parameters as required by the full likelihood, such as the baseline cause-specific hazards; this likelihood especially addresses the situation in which one is interested primarily in the effects of covariates in (1.4), but not in the shape of the hazard or in parameters corresponding to other additional variables of less interest. Another example of such nuisance parameters could be those corresponding to the censoring distribution model. The new perspective brought by the partial likelihood on the data suggests that sampling is done from the conditional probability of a failure at time t, given the number of patients at risk at time t instead from the joint density of data as done when using full likelihood above; thus, it does not require specification of the baseline hazard. We denote this partial likelihood by $\mathcal{L}_{PL}(\beta)$, where

$$\mathcal{L}_{PL}(\beta) = \prod_{j=1}^{J} \prod_i \frac{\exp(\beta_j^\top \mathbf{Z}_i)}{\sum_{k\in\mathcal{R}(t_i)} \exp(\beta_j^\top \mathbf{Z}_k)}.$$

Inference based on $\mathcal{L}_{PL}(\beta)$ leads to maximum likelihood estimators of the regression parameters gathered in β, denoted by $\widehat{\beta}_{PL}$, which share the same asymptotic behaviour as those given by the full likelihood. However, the partial likelihood requires no ties in the data. In case of ties, Breslow (1974) or Efron (1977) approximations to the partial likelihood can be used.

Johansen (1983) argued that the partial likelihood can be obtained as a *profile* likelihood from the full likelihood; in this perspective, $\mathcal{L}_{PL}(\beta)$ is equated to $\mathcal{L}(\beta, \widehat{\gamma}(\beta))$, where $\widehat{\gamma}(\beta)$ represents the maximizer of $\mathcal{L}(\theta)$ with respect to γ for fixed β. Having obtained $\widehat{\beta}_{PL}$, estimation of γ can be obtained straightforward if we plug-in $\widehat{\beta}_{PL}$ in the expression of $\widehat{\gamma}(\beta)$ which yields

$$\widehat{\lambda}_j(t_i^{(j)}) = \frac{1}{\sum_{k \in \mathcal{R}(t_i^{(j)})} \exp\left(\widehat{\beta}_j^\top \mathbf{Z}_k\right)}$$

where $t_i^{(j)}$ is an event time point where a failure of cause j occurs, $j = 1, \ldots, J$.

Another key-quantity in competing risks is the so-called *cumulative cause-specific hazard of cause* j given by

$$\Lambda_j(t) = \int_0^t \lambda_j(t)dt, \qquad j = 1, \ldots, J.$$

This summary quantity does not have a simple probabilistic interpretation (Andersen and Keiding, 2012). A non-parametric estimator of it is the famous Nelson-Aalen estimator (Nelson, 1969; Aalen, 1975) and a semi-parametric estimator implied by (1.4) is given by the Breslow estimator (Breslow (1974)). Obviously, $S(t) = \exp\left(-\sum_{j=1}^J \Lambda_j(t)\right)$.

There is another way to measure the risk on a cumulative scale: the *cause-specific cumulative incidence function* of cause j, that is

$$F_j(t) = P(T \leq t, D = j), \qquad j = 1, \ldots, J. \tag{1.5}$$

These can be seen as natural extensions to competing risks of the time-to-event distribution $P(T \leq t)$ encountered in ordinary survival. These probabilities may be estimated non-parametrically using the Aalen-Johansen estimator (Aalen and Johansen, 1978) or semi-parametrically, exploiting their relationship with the cause-specific hazards, that is

$$F_j(t) = \int_0^t \lambda_j(u) \exp(-\sum_{l=1}^J \Lambda_l(u))du, \qquad j = 1, \ldots, J.$$

An interesting aspect appears here, revealing the complexity of modeling competing risks. It concerns the interpretability of covariate effects: a covariate may

not affect the cause-specific hazard of the cause of interest, but still affects the cumulative risk of it through the effects on the cause-specific hazards of the competing events. We will come back to this feature in this thesis. Examples of clinical questions which can be addressed with the cumulative incidence function are:

Doctor: *Does the group of young leukemia patients (aged 30 or younger) have a high risk of developing a second malignancy within two years post-transplant?* or

Patient: *Doctor, which is my chance to die due to lung cancer five years after surgery if I stop smoking soon?*

To avoid the drawback of interpretability of covariate effects on the cumulative incidence scale based on models on cause-specific hazards, several approaches have been proposed in literature for direct regression on this scale. A famous one is that of Fine and Gray (Fine and Gray, 1999); they introduced the *subdistribution hazard* of cause j which preserves a one-to-one relationship with $F_j(t)$, and is given by

$$\alpha_j(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t, D = j | (T \geq t) \cup (T \leq t, D \neq j))}{\Delta t},$$

for $j = 1, \ldots, J$.

Alternative ways to the inference and estimation in competing risks models will be discussed in this thesis. Andersen and Keiding (2012) postulated three principles which could be helpful in formulating theory and for a good practice of competing risks:

(1) Do not condition on the future;

(2) Do not regard individuals at risk after they have died; and

(3) Stick to this world.

It is worth saying that across the existing approaches to competing risks, several cannot meet these demands. Ultimately, the goal of these approaches is to understand how covariates influence the underlying process of competing risks and to predict an outcome for a new patient, given certain baseline information. An extensive discussion of how covariate effects are interpreted for different functionals in competing risks can be found in Andersen and Keiding (2012).

However, one can easily guess that research questions inspired by real-data might not readily extract answers using the existing methods. What if we are interested in patterns over time of causes of failure in competing risks? What if, in case of failure, causes of failure are just partially known? What if we are interested to predict the occurrence of a specific failure of a patient, taking into account a certain part of their history?

1.1.2 Missing causes of failure

Data might be missing for a variety of practical reasons. We are concerned with the situation when missing causes of failure in competing risks arise due to reasons related to observed data (*missing at random* assumption). It is not difficult to imagine that this might occur in the process of collecting the data (for instance, the burden of other urgent matters might prevent the treating physician from updating the patient report form).

The challenge here is to estimate different functionals (cause-specific hazards, cumulative incidences, etc.) that would have been observed in a population without missing causes of failure.

Naive methods would delete the individuals with missing causes of failure or would recode the missing cause and assign it to the cause of interest (especially in the study of a lethal disease); due to the fact that the information on the survival time of these patients is omitted, these procedures will result in biased and/or inefficient estimators.

More elaborate methods were proposed as extensions of some existing regression models in competing risks to incorporate a subpopulation with incomplete information on mortality. Discussing these different methods is beyond the scope of this introduction and we will only describe some features and indicate when these methods are appropriate.

Goetghebeur and Ryan (1995) proposed a semi-parametric regression model on cause-specific hazards of two competing causes yielding estimates of the regression coefficients in the Cox proportional hazards model. Their method could be useful even if the covariates are omitted in the modeling. A particularity of their approach is that they assume the ratio $\frac{\lambda_{20}(t)}{\lambda_{10}(t)}$ to be constant, therefore facilitating the use of the partial likelihood for inference. However, their inference method is more elaborate than the traditional partial likelihood, consisting in a two-stage maximization of (appropriate) partial likelihoods. Extension of their method to include a time-varying model for $\frac{\lambda_{20}(t)}{\lambda_{10}(t)}$ is also discussed.

Lu and Tsiatis (Lu and Tsiatis, 2001) proposed a multiple imputation procedure based on modeling the cause-specific hazards by means of Cox proportional hazards model. An important feature of their approach is that they could recover complete data by imputing the missing information from a regression model on the conditional distribution of the cause of failure given observable data.

However, the question arises whether we could capture and model somehow a natural way in which such situation occurs, that is detecting when a failure occurs, irrespective of its cause, and then observing which type of failure arose.

1.1.3 Dynamic prediction

Treating physicians are often concerned with optimal treatment decisions and, given that, they are interested to predict a clinical event for a patient taking into account the individual event history (dynamic prediction).

In this thesis, we are interested to develop dynamic prediction models for competing risks which are able to address questions such as:

Doctor: *Could we standardize the use of a certain clinical management strategy that would depend on the prognosis within first three months post surgery?*
or

Patient: *Doctor, which is my chance to die due to leukemia five years after experiencing a recurrence during the first two months post-transplantation?*

To this goal, the accuracy of the parameter estimates and the fit of the model are not the main concern, but we are interested in a model that predicts accurately.

Dynamic prediction in survival analysis has received a lot of interest recently both in terms of theoretical developments and applications. A comprehensive discussion from a technical and practical point of view can be found in van Houwelingen and Putter (2012). In competing risks, the dynamic prediction probabilities are obtained from the joint distribution of $(T, D, Z(\cdot))$, where $Z(t)$ is the covariate process; more exactly, we are interested in the conditional probabilities $P(T \leq t, D = j |\ T > s, \{Z(u) : 0 \leq u \leq s\})$, $j = 1, \ldots, J$, where the time-dependent covariates play a key role.

There are a number of ways to model and estimate the dynamic prediction probabilities. If the aim is to comprehensively model the data, a multi-state model approach (Putter et al., 2007) or a joint modeling approach (Proust-Lima and Taylor, 2009; Rizopoulos, 2011) can be used to derive a model of the prediction probabilities; these will require not only a model for the time-to-event variables, but also a model for how $Z(t)$ will develop beyond the prediction time, that is for $t > s$. If one is interested only in the prediction probabilities, a direct modeling approach can be used, overcoming the burden of assumptions on some variables not of interest for prediction. Such a direct approach which has received a lot of attention is *landmarking*. The key feature of the landmarking is the updating of the time-varying covariates at each prediction time point. Then, the challenge becomes how to incorporate the updated information either in a model for the cause-specific hazards or in a direct model for the conditional cause-specific cumulative incidence functions.

1.2 Outline of the thesis

Chapter 2 introduces a new approach to competing risks data, called *vertical modeling*. It is built on natural observable quantities in competing risks, that is it quantifies 1. the chance that a failure occurs, irrespective of its cause and 2. conditionally that a failure occurred, it quantifies the risk that the event of failure is ascertained to a certain type of failure. Vertical modeling is recognized to fulfill the three principles above (see Andersen and Keiding (2012)), due to the fact it puts forward interpretable functionals and due to its practical applicability. The vertical modeling approach directly identifies and models the patterns of causes of failure over time, and it retrieves the cumulative incidence function, making it available both for prediction and for dynamic prediction. Explicit expressions are given for the variance of the cause-specific cumulative incidence functions. By using semi-parametric regression models, our approach can be easily implemented in any of the existing statistical softwares. Our approach is compared to the non-parametric approach in a large simulation study. The utility of our method is demonstrated in the analysis of a real data set. This chapter is based on the work of Nicolaie et al. (2010).

Since collection of competing risks data might be a cumbersome process in practice, it is important to be able to deal with less optimal situations which often reality faces. Chapter 3 reveals another appealing feature of vertical modeling, that is it deals with competing risks when missing causes of failure occur. Under some reasonable assumptions, this situation is handled in a natural way by vertical modeling, because it uses all the information on the time of failure (also from those individuals with missing causes of failure), while the partial information on the causes of failure is used in an optimal way. Vertical modeling leads to correct inference; maximum likelihood estimators of regression parameters based on the observed likelihood coincide with maximum likelihood parameters obtained from our approach. Other advantages of our method to some existing methods are discussed and exemplified in the analysis of a real data set. This chapter is based on the work of Nicolaie et al. (2011).

Chapter 4 proposes a new approach to the topic of dynamic prediction in competing risks, which comes as an extension of the landmark approach in ordinary survival. It is based on combining in supermodels Cox proportional hazards models of cause-specific hazards for the cohort of survivors at each landmark time point. Supermodels, obtained by smoothing the cause-specific baseline hazards over a range of landmark time points, can handle time-varying effects of covariates or time-dependent covariates, while accounting for multiple causes of failure. Estimation of regression parameters is done by means of pseudo partial-likelihoods. The advantage of this method is reduced modeling effort, because it is targeted directly at the dynamic prediction probabilities, incorporating only

the necessary information to prediction from the complex underlying process. We validate empirically our method on a real data set. This chapter is based on the work of Nicolaie et al. (2013a).

Chapter 5 proposes an alternative to the methods of Chapter 4. Instead of using the complete prediction interval framework of Chapter 4, the new approach is targeted directly at the time point where dynamic prediction is of interest. The key ingredients are pseudo-observations computed for the cohort of survivors at each landmark time point, called dynamic pseudo-observations. They are combined in supermodels for a range of landmark time points using a generalized linear model (GLM) approach, where smoothed time-varying effects of covariates or time-varying covariates could affect their mean values. Estimation is done by means of a generalized estimating equations (GEE) method. Mathematical properties of our method concerning the asymptotic behaviour of our estimators are considered. The advantage of our method of modeling in a single time point is its robustness against model mispecification which are likelily to occur in more comprehensive models. Our method is illustrated in the analysis of a real data set. This chapter is based on the work of Nicolaie et al. (2013b).

Chapter 6 studies properties of several approaches to dynamic prediction in competing risks in simulation studies, including the methods introduced in Chapters 4 and 5. The main interest here is not in the fit of the model, but in the accuracy of the methods to predict a future event taking into account all available information at that time point. Two main scenarios are considered: first, the true, underlying stochastic process is chosen to fulfill the Markov property and, secondly, when it fails to fulfill this property. This chapter is based on the work of Nicolaie et al. (2013c).

2

Vertical modeling: a pattern mixture approach to competing risks

Abstract

We study an alternative approach for estimation in the competing risks framework, called vertical modeling. It is motivated by a decomposition of the joint distribution of time and cause of failure. The two elements of this decomposition are 1. the time of failure and 2. the cause of failure conditional on time of failure. Both elements of the model are based on observable quantities, namely the total hazard and the relative cause-specific hazards. The model can be implemented using standard software. The relative cause-specific hazards are flexibly estimated using multinomial logistic regression and smoothing splines. We show estimates of cumulative incidences from vertical modeling to be more efficient statistically than those obtained from the standard nonparametric model. We illustrate our methods using data of 8966 leukemia patients from the European Group for Blood and Marrow Transplantation.

2.1 Introduction

Competing events in medical research concern situations where individual subjects may experience multiple types of events, such that the occurrence of one event precludes the occurrence of others (e.g., death in the disease process may occur due to several mutually exclusive causes). Competing risks models provide a natural framework to describe multiple causes of failure; they can be seen as a multi-state model with one initial state 0 (alive, event-free) and a number of absorbing states $j = 1, \ldots, J$, corresponding to the different types of events Andersen et al. (2002).

Interest focuses on the joint distribution of time to failure T and cause of failure D, abstractly denoted here as $P(T, D)$. Typically this joint distribution is summarized through the cumulative incidence functions $F_j(t) = P(T \leq t, D = j)$. The standard approach of modeling them is through the cause-specific hazard functions $\lambda_j(t)$ (Prentice et al., 1978), the hazard of failing from a given cause in the presence of competing events. Unfortunately, the cause-specific hazard function does not have a direct interpretation in terms of survival probabilities for the particular failure type, since these probabilities also depend on the cause-specific hazards of the other causes.

The joint distribution of (T, D) may also be decomposed as a mixture model in two ways:

$$P(T, D) = P(T|D)P(D) \ , \tag{2.1}$$

or as

$$P(T, D) = P(D|T)P(T) \ . \tag{2.2}$$

The decomposition (2.1) has been proposed and studied by Larson and Dinse (1985) and has received modest attention (Ng and McLachlan, 2003; Lu and Peng, 2008; Kuk, 1992). Decomposition (2.1) may perhaps seem more obvious on first glance, on reflection it is a somewhat awkward construction, resulting in two major disadvantages. The first is that, from a practical point of view, an EM-algorithm typically has to be used to infer the missing cause of failure for censored observations. This makes estimation in model (2.1) time-consuming and difficult. The second disadvantage is arguably even more important. From an interpretational point of view, it implies that the cause of death is determined from the outset. Its estimated distribution will depend on the length of follow-up, which is hard to reconcile with the implied existence of such a distribution from the outset. In that respect decomposition (2.2) is arguably more natural; its constituents are the time of failure T and its cause D once failure has occurred. Also the ingredients needed to estimate $P(D|T)$ and $P(T)$ are closely connected to the natural observable quantities in the competing risks setting, the cause-specific hazards. For $P(T)$ the driving force is $\lambda_\bullet(t)$, the total or all-cause hazard,

and for $P(D|T)$ the relevant quantities are the relative cause-specific hazards $\pi_j(t) = \frac{\lambda_j(t)}{\lambda_\bullet(t)}$. Although the idea of the decomposition (2.2) appears to have been used implicitly (Hachen, 1988; Smits et al., 2000), the decomposition (2.1) has been used more widely, despite its obvious disadvantages. We feel therefore that model (2.2) deserves more attention and deeper study.

We introduce in Section 2.2.1 the notation we need to describe the model. In Sections 2.2.2 and 2.2.3 we describe the model, both with and without covariates. We contrast the vertical modeling and existing approaches in Section 2.2.4. A leukemia patients data set analysis demonstrates the utility of our methods in Section 2.3. In Section 2.4 we discuss our method and results. Technical details dealing with the computation of standard errors can be found in the Appendix.

2.2 Vertical modeling

2.2.1 Notation

Let T be the time of failure, C the censoring time, and D the cause of failure with possible values $1, \ldots, J$. Let $\mathbf{Z}$ denote a vector of covariates. We observe $(\tilde{T}_i, \Delta_i, \mathbf{Z}_i)$, for $i = 1, \ldots, n$, where $\tilde{T}_i = \min(T_i, C_i)$ is the earliest of failure and censoring time, and $\Delta_i = \mathbf{1}\{T_i < C_i\} \cdot D_i$ is the cause of failure in case of failure and 0 in case of censoring. The usual requirement of conditional independence of (T, D) and C, given $\mathbf{Z}$, is assumed to be true here as well. The aim is to estimate the joint distribution of (T, D), given by the cumulative incidence functions

$$F_j(t) = P(T \leq t, D = j) \ . \tag{2.3}$$

Viewed as a function of t, $F_j(t)$ is a possibly (or rather, probably) defective distribution function.

The standard way of estimating the cumulative incidence functions is through the cause-specific hazards

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, D = j | T \geq t)}{\Delta t} \ , \ \Lambda_j(t) = \int_0^t \lambda_j(s) ds \ . \tag{2.4}$$

Define the total, overall, or all-cause hazard

$$\lambda_\bullet(t) = \sum_{j=1}^{J} \lambda_j(t) \ , \ \Lambda_\bullet(t) = \int_0^t \lambda_\bullet(s) ds = \sum_{j=1}^{J} \Lambda_j(t) \tag{2.5}$$

and its corresponding survival function

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

In what follows, we shall refer to $\lambda_\bullet(t)$ ($\Lambda_\bullet(t)$) as the *total (cumulative) hazard*. Then the cumulative incidence function of cause j may be expressed in terms of the cause-specific hazards as

$$F_j(t) = \int_0^t \lambda_j(s) S(s-) ds \ . \tag{2.7}$$

Covariates may be incorporated in this context through proportional hazards model imposed on the cause-specific hazards, i.e. by assuming that $\lambda_j(t|\mathbf{Z}) = \lambda_{j,0}(t)\exp(\beta_j^\top \mathbf{Z})$, with $\lambda_{j,0}$ an unspecified baseline hazard and β_j an unknown vector of regression coefficients, both to be estimated. The β_j are usually supposed to be different for different causes, but they may also be taken to be identical, see for instance Putter et al. (2007). Alternatively, such proportional hazards model may be specified for the subdistribution hazards (Fine and Gray, 1999).

Define also the *relative cause-specific hazards*

$$\pi_j(t) = \frac{\lambda_j(t)}{\lambda_\bullet(t)}, \ j = 1, \ldots, J \ . \tag{2.8}$$

An important issue about the $\pi_j(t)$ is that it describes a local time behaviour, namely

$$\pi_j(t) = P(D = j|T = t) \ . \tag{2.9}$$

Also, $\sum_{j=1}^k \pi_j(t) = 1$. Although the concept of the relative hazard has already been used in the literature as the ratio of two hazards (see, e.g., Armitage and Colton (2007)), for ease of reference we shall simply refer to the concept in (2.8) as the *relative hazard*. Note that the relative hazard can be essentially any function taking values in $[0, 1]$.

The intuition for the vertical modeling is that, when estimating the probability of dying from the target cause, we first estimate the probability of failure, irrespective of the cause of death (an "overall" view) and then the conditional probability of dying from the target cause, given that the death occurred at that time. In terms of formulas, this idea leads to decomposition (2.2) of the cumulative incidence function. With regard to the modeling aspect, two models are needed: a model for the overall failure time (irrespective of its cause) and a model for the cause of failure, given the failure time, models which are described in the following.

2.2.2 No covariates

The driving force for the overall failure time distribution is the total hazard; here, all failures are considered as events, irrespective of the cause of failure. The most obvious choice for estimation is the nonparametric Kaplan-Meier estimator.

The driving force for the cause of failure given time of failure is the relative hazard, given by (2.8) and (2.9). Note that due to the independence assumption of (T, D) and C, we have

$$\begin{aligned} \pi_j(t) &= P(D = j|T = t) = P(D = j|T = t, C \geq t) \\ &= P(D = j|T = t, T \leq C) \ , \end{aligned}$$

so that we may restrict ourselves to the observed event time points and ignore the censored observations. In order to estimate it "model free", let $0 < t_1 \leq t_2 \leq \ldots \leq t_M$ be the ordered event times at which failures of any cause occur. Let d_{kj} denote the number of patients failing from cause j at time t_k, and let $d_k = \sum_{j=1}^{J} d_{kj}$ denote the total number of failures (from any cause) at time t_k. Let n_k be the number of patient at risk (i.e. that are still in follow-up (alive or not censored) and have not failed from any cause at time t_k). The relative hazard $\pi_j(t_k)$ can be estimated by means of estimates of the hazards in (2.8), namely

$$\widehat{\pi}_j(t_k) = \frac{\frac{d_{kj}}{n_k}}{\frac{d_k}{n_k}} = \frac{d_{kj}}{d_k} \ . \tag{2.10}$$

If we assume that $\pi_j(t)$ is continuous in reality and if we are interested in a smooth curve of $\pi_j(t)$, we need to smooth, because in the absence of ties, only one of the d_{kj} equals 1 for a given k, and $d_k = 1$. Therefore, no smoothing would lead to a very erratic behaviour of $\pi_j(t)$. This leads to the choice of a smoother as function of time. We will use a predefined set of functions of time $B_1(t)$, $B_2(t), \ldots, B_p(t)$, for instance spline basis functions, gathered in a vector $\mathbf{B}(t)$=$(B_1(t), B_2(t), \ldots, B_p(t))^\top$. We will assume that $\mathbf{B}(t)$ includes an intercept. Using this set of time functions, the most natural model for the relative hazard is a multinomial logistic model, which specifies that

$$\pi_j(t) = \frac{\exp(\beta_j^\top \mathbf{B}(t))}{\sum_{l=1}^{J} \exp(\beta_l^\top \mathbf{B}(t))} \ , \ j = 1, \ldots, J \ , \tag{2.11}$$

where $\beta_j = (\beta_{j1}, \ldots, \beta_{jp})$ is a row vector of p regression coefficients, $j = 1, \ldots, J$. For identifiability, we may set $\beta_1 \equiv 0$. In the case of two causes of failure, the multinomial logistic model will simplify to a binary logistic model. Both multinomial and binary logistic regression models are standard in almost all statistical

software packages.

It is worth reiterating that a vertical model, through (2.2), describes the joint distribution of time and cause of failure, just like a model based on the cause-specific hazards. Thus, the cumulative incidence functions can also be retrieved from a vertical model. From the model for the total hazard, an estimate of $\exp(-\Lambda_\bullet(t))$ can be obtained. The reversal of the definition of the relative hazard π_j gives the cause-specific hazard

$$\lambda_j(t) = \pi_j(t)\lambda_\bullet(t) \ . \tag{2.12}$$

Finally, the cumulative incidence of cause j can also be expressed in terms of relative and total hazard, through the relation

$$\begin{aligned} F_j(t) &= \int_0^t \lambda_j(u) \exp(-\Lambda_\bullet(u))du \\ &= \int_0^t \pi_j(u)\lambda_\bullet(u) \exp(-\Lambda_\bullet(u))du \\ &= \int_0^t \pi_j(u) f_\bullet(u)du \ , \end{aligned} \tag{2.13}$$

where $f_\bullet$ denotes the density corresponding to the overall failure time distribution. The last equation (2.13) is interesting; it describes the role of the relative hazard as quantifying the proportion of the overall failure density contributing to the cumulative incidence function.

Dynamic prediction (i.e., prediction from a later time point $s > 0$) is also straightforward for vertical modeling: $P_{0j}(s,t) := P(T \leq t, D = j | T > s)$ can be expressed in terms of relative and total hazards as an obvious extension of (2.13), namely $P_{0j}(s,t) = \int_s^t \pi_j(u) f_\bullet(u)du$.

In the Appendix we show how to compute the standard errors of cumulative incidences from the vertical model.

2.2.3 Covariates

When we want to incorporate covariates in the vertical modeling, we have to consider ways of incorporating these covariates in the model for the time to overall failure (the total hazard) and in the model for the cause of failure given time of failure (the relative hazards). The most obvious model for the time to failure is a proportional hazards model. Again, all failures are considered as events, irrespective of the cause of failure. In case of a single categorical variable, a nonparametric alternative is to use the Nelson-Aalen estimator for different levels of the covariate.

For the cause of failure, the most natural model would again be a multinomial regression model. The challenge here is to obtain sufficiently rich and meaningful models which are not difficult to fit. One option is to obtain sub-groups and apply smoothing methods within each subgroup. Alternatively, one could expand the multinomial logistic regression model of (2.11) to include the covariates and possibly the interactions of these covariates with the time functions. To illustrate this last possibility, we propose the following model for the relative hazard

$$\pi_j(t) = \frac{\exp(\beta_j^\top \mathbf{B}(t) * Z)}{\sum_{l=1}^{J} \exp(\beta_l^\top \mathbf{B}(t) * Z)},\ j = 1, \ldots, J\ , \tag{2.14}$$

where Z stands for the covariate and $*$ for its interaction with the time functions, and where again $\beta_1 \equiv 0$. Note that no main covariate effects are included since $\mathbf{B}(t)$ includes an intercept. In case this model is not identifiable from the data, and additive model may be used, where $\exp(\beta_j^\top \mathbf{B}(t) * Z)$ is replaced by $\exp(\beta_j^\top \mathbf{B}(t) + \gamma_j Z)$, with $\beta_1 \equiv 0$ and $\gamma_1 = 0$. A likelihood ratio test could be used to test whether the covariate by time interaction needs to be included in (2.14). Sometimes it is biologically or clinically plausible to put restrictions on the relative hazards; the relative hazard of a specific cause could be decreasing or non-decreasing, it could be higher than the relative hazard of a second cause, or it may be plausible that two relative hazards are proportional. Such restrictions are often more natural for the relative hazards than for the cause-specific hazards; they may be taken into account in the multinomial regression model, and they will result in a gain in efficiency.

2.2.4 Vertical modeling in perspective

The most commonly used methods for competing risks are based on the cause-specific hazards, which can be seen (Figure 2.1) as the transition intensities of a multi-state mode, with "Alive" as starting state and failures from the different causes as absorbing states (Andersen et al., 2002). By far the most popular model for the cause-specific hazards is the Cox model, which specifies that

$$\lambda_k(t \mid \mathbf{Z}) = \lambda_{k,0}(t) \exp(\boldsymbol{\beta}_k^\top \mathbf{Z})\ . \tag{2.15}$$

The advantage of (2.15) is that it is straightforward to fit using standard statistical software (Lunn and McNeil, 1995). The main disadvantage is that the one-to-one rate to risk relation, $F(t) = 1 - \exp(-\int_0^t \lambda(s)ds)$ (λ being the rate (hazard) and F being the risk (distribution function)) that we are used to in ordinary survival analysis no longer holds. The reason is that the effect of a covariate on the cumulative incidence of a cause j of interest not only depends on $\boldsymbol{\beta}_j$, but also on the other $\boldsymbol{\beta}_l$'s, because of the fact that the cumulative incidence

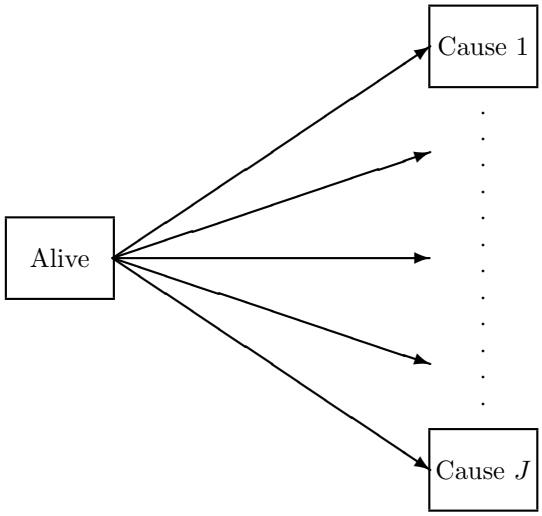


Figure 2.1: A competing risks model with J causes of failure

function $F_j(t) = \int_0^t \lambda_j(s)S(s-)ds$ depends not only on λ_j but also on the other cause-specific hazards λ_l through $S(s-) = \exp(-\sum_{l=1}^{J} \int_0^{s-} \lambda_l(u)du)$ Prentice et al. (1978); Putter et al. (2007).

This fact prompted by Fine and Gray (1999) to propose a proportional hazards model on the subdistribution hazard $\lambda_j^*(t)$ which is defined so as to satisfy the one-to-one rate to risk relation $F_j(t) = 1 - \exp(-\int_0^t \lambda_j^*(s)ds)$. The Fine-Gray model specifies

$$\lambda_j^*(t \mid \mathbf{Z}) = \lambda_{j,0}^*(t) \exp(\boldsymbol{\beta}_j^{*\top} \mathbf{Z}) \ . \tag{2.16}$$

The usefulness of the Fine-Gray model and the reason for its growing popularity, both in terms of applications and research effort (Andersen et al., 2003; Klein and Andersen, 2005) is of course that it establishes a one-to-one relation between covariates and cumulative incidence, but it has a number of disadvantages as well. The subdistribution hazard itself is an awkward construct, there are technical difficulties to be overcome when fitting the model, and the theory has not completely developed yet (for instance, time-dependent covariates and time-dependent covariate effects is still an active research area on time-dependent covariates in the Fine-Gray model (Beyersmann and Schumacher, 2008)).

Both the proportional cause-specific hazards model (2.15) and the Fine & Gray model (2.16), despite their differences, can be seen as horizontal models in

that they model what happens when we follow the arrows in Figure 2.1 from left to right (they model the rate at which particular types of events occur). Vertical modeling first describes the total intensity out of the "Alive" state in Figure 2.1, say one arrow from left to amounted right. The relative hazards then describe the transition intensities or cause-specific hazards vertically through their relative magnitudes. Our model assumes the continuity of the relative hazards, which after smoothing might facilitate their interpretation. Vertical modeling is most useful if the biological (or mechanical) system causing failures can be thought of as consisting of an overall rate of failure which may be subdivided by different causes of failure. An example might be causes of death in public health where overall mortality rate as a function of age may be subdivided by different causes of death and where interest lies in quantifying the contribution of these different causes of death to overall mortality in the course of time (age). Of course, it may also be a useful alternative in situations where the proportional hazards assumptions on the cause-specific and/or subdistribution hazards fail to hold. An example of violation of these assumptions will be given in the next section.

Just like some authors advise to analyze and report the results of proportional hazards models on both cause-specific and subdistribution hazards, vertical modeling may be used side-by-side with these two methods, since the different models emphasize different aspects of the data.

2.3 Data analysis

The data studied in this paper comes from the European Group for Blood and Marrow Transplantation (EBMT) and was also used in Fiocco et al. (2005). It consists of 8966 patients with acute myeloid leukemia (AML), acute lymphoblastic leukemia (ALL), or chronic myeloid leukemia (CML), who received an allogeneic hematopoietic stem cell transplantation (HSCT) for early leukemia. Data was reported to EBMT in three time cohorts: 1985-1989 (27%), 1990-1994 (40%), 1995-1998 (33%). The objective was to study whether patterns of causes of death changed over time, and in particular whether deaths from infections and graft-versus-host disease (GvHD) only occurred in the first year after transplantation or whether treating physicians should remain alert to the possibility of dying from infections or GvHD after one year post-transplant (Gratwohl et al., 2005). For this particular analysis we collapsed the different "death due to infection" categories. This leaves the following causes of death: death from relapse and transplant related mortality which was subdivided into death from acute or chronic graft-versus-host disease (GvHD), infections or "other" causes, as shown in Table 2.1. Median follow-up was approximately six years. Gender mismatch refer to a particular combination of donor/ recipient gender (donor female, recipient male) that is known to confer higher risk of adverse outcome. Covariate

Table 2.1: Number of events for each competing risk; 5656 patients were alive at the last follow-up

Event	Relapse	GvHD	Infections	Other
Number	1098	843	454	924
Percentage with respect to patients	12 %	9 %	5 %	10 %
events	33 %	25 %	14 %	28 %

information is described in Table 2.2.

Table 2.2: Prognostic factors for all patients

Prognostic factor		n (%)
Disease classification	AML	3514 (39%)
	ALL	1870 (21%)
	CML	3582 (40%)
Donor recipient	no gender mismatch	6758 (75%)
	gender mismatch	2208 (25%)
GvHD prevention	no TCD	4390 (49%)
	+ TCD	1720 (19%)
	unknown	2857 (32%)
Year of HSCT	1985–1989	2390 (27%)
	1990–1994	3575 (40%)
	1995–1998	3001 (33%)
Age at transplant (years)	≤ 20	1974 (22%)
	20–40	4800 (54%)
	> 40	2192 (24%)

2.3.1 No covariates

To illustrate our vertical modeling, we are first concerned with the choice of time functions. With respect to this goal, we start by using piecewise constant functions. We divide the follow-up time interval in five subintervals: $I_1 = [0, 0.25)$, $I_2 = [0.25, 0.5)$, $I_3 = [0.5, 1)$, $I_4 = [1, 2.5)$ and $I_5 = [2.5, \infty)$, corresponding to 32%, 21%, 19%, 17%, 11% of all events, respectively, and we define $B_i(t) = \mathbf{1}_{I_i}(t)$, $i = 1, \ldots, 5$. Multinomial regression can be used, or alternatively the resulting

estimated relative hazards $\widehat{\pi}_{jk}$ of cause j in interval I_k can be obtained directly from equation (2.11) as

$$\widehat{\pi}_{jk} = \frac{\exp(\widehat{\beta}_{jk})}{\sum_{l=1}^{5} \exp(\widehat{\beta}_{jl})},\ j = 1, \ldots, 4,\ k = 1, \ldots, 5\ . \tag{2.17}$$

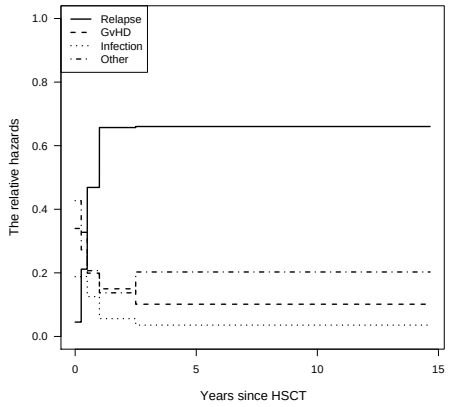
For this special case of piecewise constant time functions, the $\widehat{\pi}_{jk}$ can also be obtained directly, as in (2.10), namely, the proportion of all failures occurring in I_k that are attributed to cause j.

Figure 2.2a shows the plots of the associated relative hazards implied by our choice of time functions, equation (2.17) or, equivalently, the estimated values from Table 2.3 obtained from (2.17).

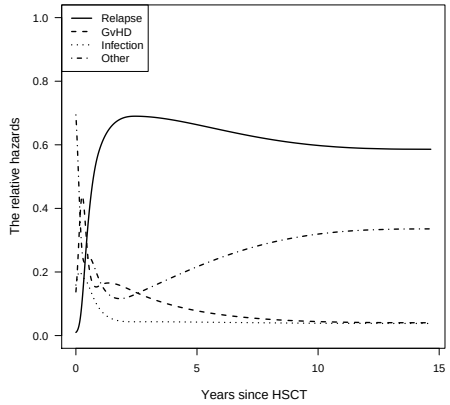
These piecewise constant functions are very useful to obtain a first idea of the relative hazards. Often, it is preferable to show relative hazards as smoothed functions obtained from (2.17). For illustration, we will use cubic splines, which consist of piecewise cubic polynomials between adjacent knots (i.e., of the form $ax^3 + bx^2 + cx + d$), are continuous and smooth at each knot, with continuous first and second derivatives (see, e.g., Schumaker (1981)). For this particular analysis, we introduce 7 knots on the time axis $(k_i)_{i=1,\ldots,7} = -1$, -0.5, 0, 0.25, 0.5, 1, 2.5 and 15 which lead to four such functions, denoted by $B_i(t), i = 1, \ldots, 4$. In Figure 2.2b we show the plots of the associated relative hazards implied by our choice of spline basis functions, equation (2.11) and the estimated regression coefficients from Table 2.4. Here we use $\beta_1 \equiv 0$ for identifiability. In the next section, we will reexamine this model, adjusting for covariates. Figures 2.2a and b are of course based on different spline functions, but qualitatively they are the same. It can be seen from Figure 2.2 that in the first year after transplantation, the majority of deaths come from the transplant related causes: GvHD, infections and other causes, while from one year after transplant relapse is the dominating cause of death. It seems that the other causes of death do not completely disappear. The behaviour of the relative hazards in the right tail is of course very uncertain due to the small number of events.

Figure 2.3 shows the plots of the cause-specific cumulative hazards implied by the cubic splines model through (2.11) and (2.12), together with their standard errors. They are compared with the nonparametric estimates. The estimates of cumulative hazards are virtually indistinguishable, but the standard errors from the vertical model are somewhat higher.

Figure 2.4 shows the plots of the cumulative incidence functions together with their standard errors, which are based on the smoothed relative hazards estima-



(a)



(b)

Figure 2.2: The piecewise constant relative hazards and the smoothed relative hazards

tors, (2.12) and (2.13). The estimates of cumulative incidences and their standard errors were obtained via the `mstate` package in `R` (see de Wreede et al. (2010)) and are based on the standard errors of cumulative hazards as derived in the Appendix. The estimates of cumulative incidences are again virtually indistinguishable, but the standard errors from the vertical model are smaller now. The fact that the standard errors of the cumulative incidences are smaller, while they were bigger for the cause-specific cumulative hazards, is caused by the fact that

Table 2.3: Estimated relative hazards and their standard errors for the multinomial model

	$[0, 0.25)$	$[0.25, 0.5)$	$[0.5, 1)$	$[1, 2.5)$	$[2.5, +\infty)$
Relapse	0.045 (0.0063)	0.212 (0.0155)	0.469 (0.0197)	0.657 (0.0201)	0.660 (0.0247)
GvHD	0.339 (0.0145)	0.328 (0.0178)	0.199 (0.0158)	0.150 (0.0151)	0.101 (0.0157)
Infections	0.188 (0.0119)	0.188 (0.0148)	0.125 (0.0131)	0.056 (0.0097)	0.036 (0.0097)
Other	0.427 (0.0151)	0.272 (0.0169)	0.207 (0.0160)	0.137 (0.0146)	0.203 (0.0210)

Table 2.4: Estimated regression coefficients and their standard errors for the multinomial model based on cubic splines

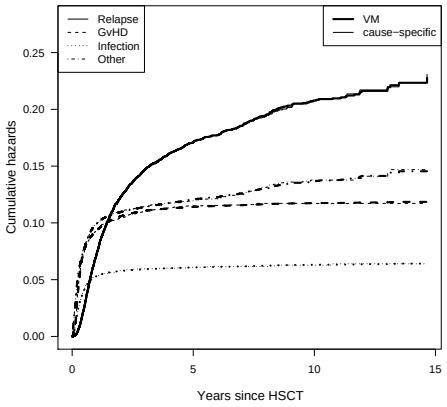
	Intercept	$B_1(t)$	$B_2(t)$	$B_3(t)$	$B_4(t)$
Relapse	–	–	–	–	–
GvHD	-2.676 (0.398)	6.514 (0.817)	4.186 (0.466)	1.385 (0.402)	1.359 (0.509)
Infections	-2.731 (0.488)	7.538 (0.886)	2.943 (0.555)	1.557 (0.494)	-0.044 (0.648)
Other	-0.557 (0.229)	7.587 (0.741)	0.758 (0.327)	0.013 (0.256)	-1.428 (0.342)

for the vertical modeling the estimated cause-specific cumulative hazards for different causes are negatively correlated, while for the nonparametric estimates they are uncorrelated.

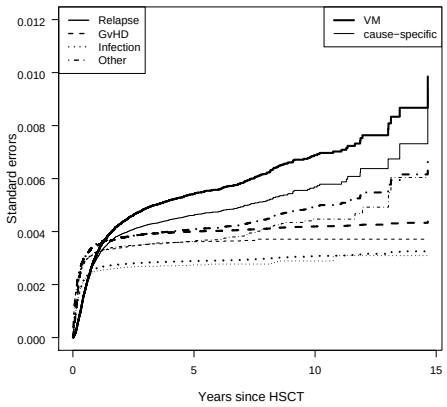
2.3.2 Covariates

The most important covariate to consider is disease subclassification of acute or early leukemia, classified as either AML (the reference category), ALL or CML. In Figure 2.5 we show the nonparametric estimates of the total cumulative hazard for each level of the disease subclassification. There is strong evidence that a Cox proportional hazards model for overall survival will not be appropriate for this particular covariate; the estimate of the cumulative hazard for CML shows a much more gradual increase than those for AML and ALL. We use these Nelson-Aalen estimators to model the total hazard for each level of disease subclassification.

With respect to the modeling of the cause of failure, we make use of the same sequence of knots and the same cubic splines $B_i(t)$, $i = 1, \ldots, 4$, as derived in Section 3.1. Table 2.5 gives the estimated regression coefficients with associated



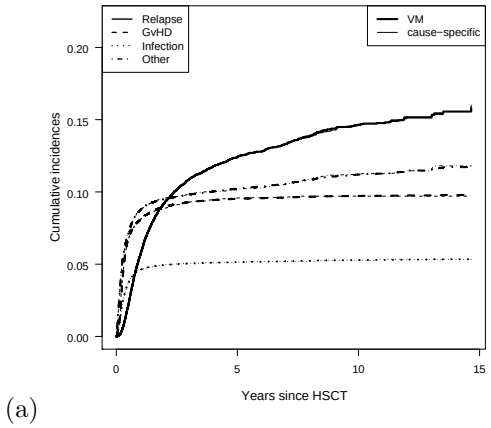
(a)



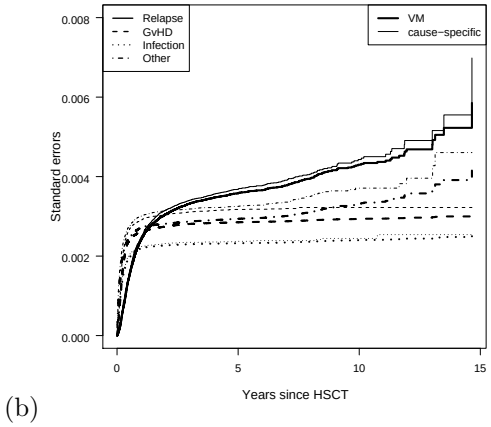
(b)

Figure 2.3: The estimated cause-specific cumulative hazards (a) and associated standard errors obtained from the vertical modeling (b)

standard errors. Again we use $\beta_1 \equiv 0$ for identifiability. A likelihood ratio test showed that the model (2.14) could not be replaced by a simpler model with only main effects for cubic splines and disease subclassification ($\chi^2 = 137.09$, $df = 9$, $p < 0.0001$). The plots of relative hazards are given in Figure 2.6. The pictures suggest that AML patients are very similar to ALL patients, with respect to their cause of death over time. The likelihood ratio test comparing model (2.14) with a model with equal coefficients for AML and ALL was not significant ($\chi^2 = 16.57$,



(a)



(b)

Figure 2.4: The cumulative incidences (a) and associated standard errors (b)

$df = 15$, $p = 0.34$).

Figure 2.7 shows the plots of the estimated cumulative cause-specific hazards implied by our model through (2.12) and based on the smoothed relative hazard estimates. While the relative hazards are similar for AML and ALL, clearly this is not the case for the cause-specific hazards; especially the cause-specific hazard for relapse is higher for ALL compared to AML.

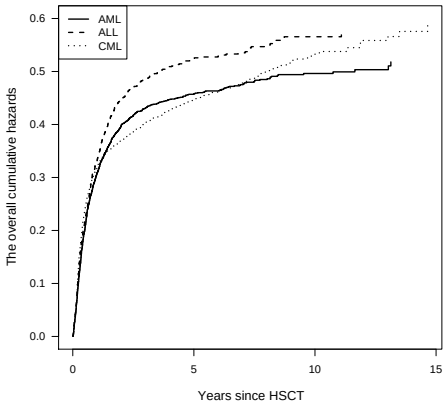


Figure 2.5: Nonparametric estimates of the overall cumulative hazard obtained for AML, ALL and CML patients separately

Figure 2.8 shows the plots of the cumulative incidences, which are based on the smoothed relative hazard estimates, (2.12) and (2.13).

Figures 2.7 and 2.8 are practically identical to the corresponding nonparametric estimates (not shown here). A lack of proportionality can be seen in Figure 2.7: especially for relapse as cause of death there is strong evidence that proportionality of cause-specific hazards among AML, ALL and CML patients does not hold. Again, for relapse as cause of death, there is strong evidence from Figure 2.8 that also proportionality of the subdistribution hazards among AML, ALL and CML patients doesn't hold. It seems clear that for disease subclassification as covariate, neither a proportional hazards model for the cause specific hazards, nor for the subdistribution hazards model is appropriate.

Finally, we illustrate our model including the remaining covariates in the analysis. In modeling the time to failure, we found that a proportional hazards model with GvHD prevention, years since HSCT and age at transplant as covariates stratified by disease subclassification and donor recipient in interaction is appropriate. The estimated hazard ratios and 95% confidence interval for GvHD prevention, years of HSCT and age at transplant are reported in Table 2.6. Figure 2.9 shows plots of the estimated cumulative baseline hazards for each of the six strata.

Table 2.5: Regression coefficients

AML	Intercept	$B_1(t)$	$B_2(t)$	$B_3(t)$	$B_4(t)$
Relapse	–	–	–	–	–
GvHD	-1.714 (0.704)	6.333 (1.358)	2.617 (0.801)	-0.552 (0.715)	-0.014 (0.892)
Infections	-1.038 (0.669)	6.648 (1.358)	0.802 (0.780)	-0.635 (0.689)	-1.941 (0.911)
Other	0.583 (0.428)	6.897 (1.218)	-0.760 (0.561)	-1.927 (0.465)	-2.976 (0.616)
ALL	Intercept	$B_1(t)$	$B_2(t)$	$B_3(t)$	$B_4(t)$
Relapse	–	–	–	–	–
GvHD	-3.376 (0.309)	6.603 (2.121)	3.915 (2.258)	1.426 (1.743)	1.259 (1.498)
Infections	-1.835 (0.318)	6.034 (1.985)	1.254 (2.362)	0.309 (1.995)	-2.659 (1.342)
Other	0.416 (0.275)	5.626 (1.167)	-1.172 (1.743)	-1.238 (1.680)	-3.357 (0.723)
CML	Intercept	$B_1(t)$	$B_2(t)$	$B_3(t)$	$B_4(t)$
Relapse	–	–	–	–	–
GvHD	-3.393 (1.698)	5.821 (0.509)	7.287 (0.462)	3.963 (2.106)	3.107 (0.847)
Infections	-4.880 (1.889)	8.105 (0.571)	7.322 (0.507)	5.182 (2.148)	3.290 (1.415)
Other	-1.609 (1.181)	7.739 (0.440)	3.981 (0.436)	2.776 (1.476)	0.590 (0.671)

Table 2.6: Hazard ratios with confidence intervals

Prognostic factor		HR (95% CI)
GvHD prevention	no TCD	1
	+ TCD	1.167 (1.065-1.279)
	unknown	1.201 (1.109-1.300)
Year of HSCT	1985–1989	1
	1990–1994	0.752 (0.693-0.815)
	1995–1998	0.600 (0.546-0.659)
Age at transplant (years)	≤ 20	1
	20–40	1.515 (1.373-1.671)
	> 40	2.045 (1.826-2.290)

With respect to the modeling of the cause of failure, we again use the same sequence of knots and the same cubic splines $B_i(t)$, $i = 1, \ldots, 4$, as derived earlier. A stepwise downward selection procedure based on minimizing the AIC selected a model with interaction between disease subclassification and splines and only main effects for donor recipient, GvHD prevention, years since HSCT and age at transplant (AIC=17217.594). Figure 2.10 shows the resulting estimates of the relative hazards separately for combinations of AML, CML (ALL is similar to

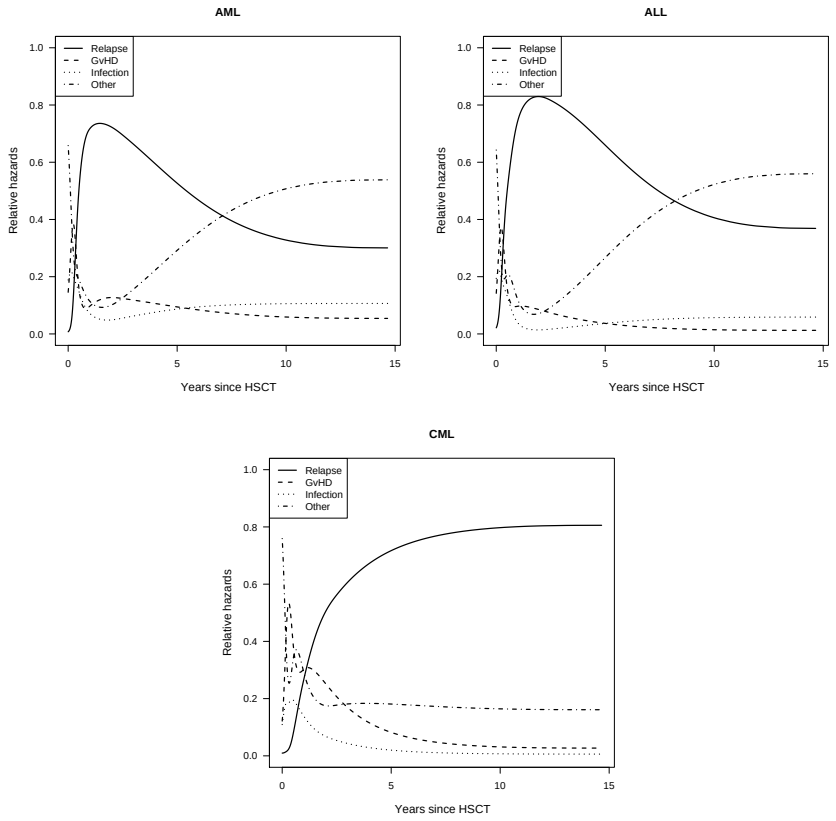


Figure 2.6: The relative hazards

AML and omitted) and age category, choosing the most common category for the remaining covariates. It can be seen that older patients die relatively more frequently of other causes and infections (only for AML).

2.3.3 Simulation experiments

In this section we present the results of three sets of simulations, based on data sets consisting of $n = 500$, 1000 and 5000 individuals, respectively. We compare the estimators of the cumulative incidence functions derived from the vertical modeling with the nonparametric estimators of cumulative incidences.

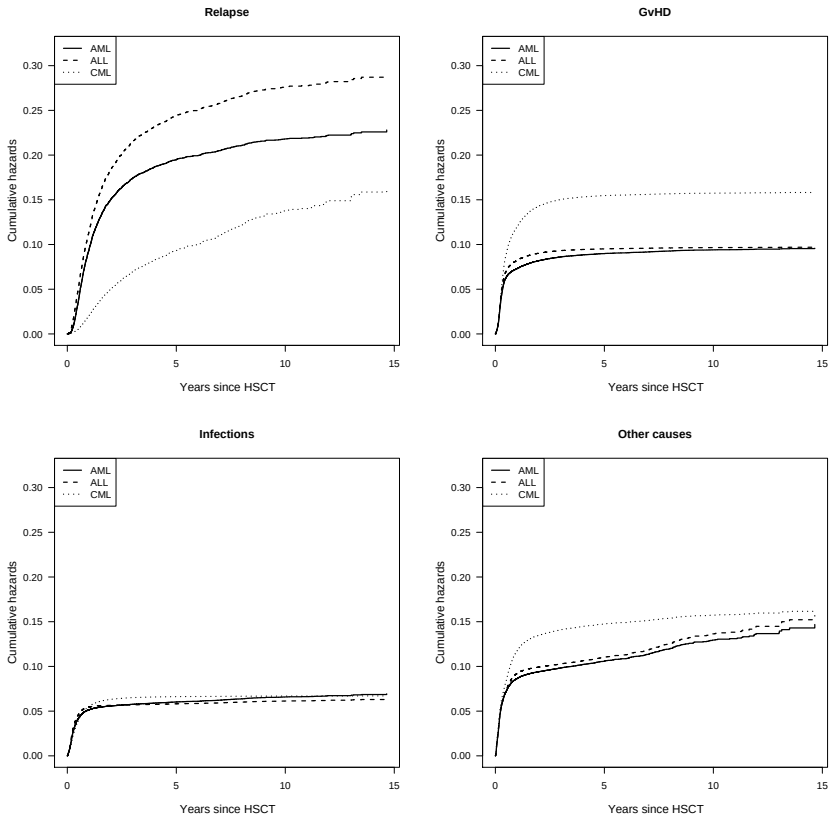


Figure 2.7: The cause-specific cumulative hazard for each level of disease subclassification

In each of these three sets of simulations, data sets were generated from a model with piecewise constant cause-specific hazards; no covariates were included and the overall failure time distribution was given by a piecewise exponential distribution on the same five time intervals $I_1 = [0, 0.25)$, $I_2 = [0.25, 0.5)$, $I_3 = [0.5, 1)$, $I_4 = [1, 2.5)$, $I_5 = [2.5, \infty)$ (see, e.g., Cox and Oakes (1984)), as in Section 2.3.1, with rates $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (0.5022, 0.3755, 0.1958, 0.0669, 0.0155)$, as estimated from the EBMT data. After generating overall failure times, depending on the interval I_j in which the failure time fell, the cause of failure was generated according to the (piecewise constant) relative hazards of Table 2.3. Right censoring times were generated uniformly on $(10, 20)$, leading to a 50-55

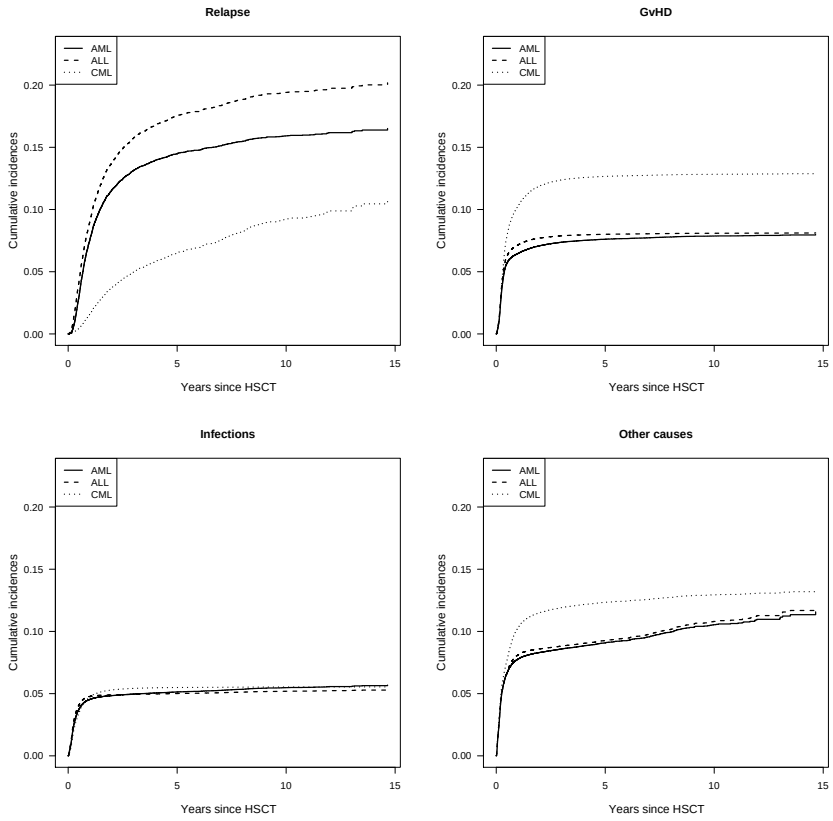


Figure 2.8: The cumulative incidences for each level of disease subclassification

per cent of censored individuals per simulated data set.

The true cumulative incidence functions at the endpoints of the five time intervals were calculated from the Chapman-Kolmogorov equation $\mathbf{P}(s,t) = \mathbf{P}(s,u)\mathbf{P}(u,t)$, $s \leq u \leq t$, and $\mathbf{P}(s_j,t_j) = \exp((t_j - s_j)\mathbf{Q_j})$, $s_j, t_j \in I_j$, where $\mathbf{Q_j}$ is the constant intensity matrix containing the exponential cause-specific failure rates on the intervals I_j and $\mathbf{P}(s,t)$ is the transition probability matrix from time s to time t (see, e.g., Iosifescu (1980)). The package `msm` package in `R` Jackson et al. (2003) was used for calculations. The cumulative incidence probabilities from the vertical model were calculated using the same cubic spline basis functions and knots as in our data analysis.

The results, in terms of bias and root mean squared error (RMSE) with respect

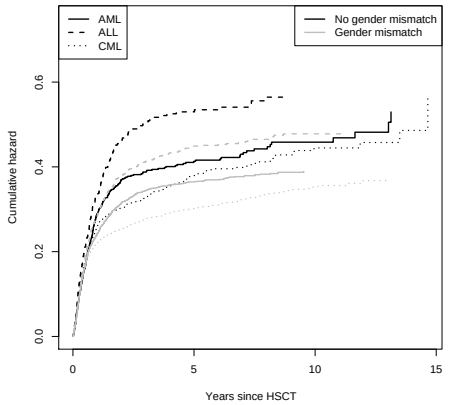


Figure 2.9: The estimated cumulative baseline hazards for each of the six strata defined by disease subclassification and gender mismatch; the line types distinguish between the disease subclassification, the grey scales between donor recipient gender mismatch yes/no

to the true cumulative incidence probabilities are reported in Table 2.8. All numbers were divided by 10^{-3}. Bias for the vertical model is larger than for the nonparametric estimates and doesn't decrease for larger sample sizes, while for smaller sample sizes, the RMSE of the vertical model is smaller than that of the nonparametric estimates. As expected, RMSE decreases for the nonparametric estimates at rate $\sqrt{n}$. The smaller RMSE for the vertical model compared to the nonparametric estimates seems to disappear for larger sample sizes. This is most probably due to the fact that we have used the same spline basis functions and knots, irrespective of the sample size and simulated data set. Probably, with regard to minimizing the RMSE of the vertical model, our choice of spline basis functions and knots was not optimal for all sample sizes and has lead to some degree of oversmoothing of the relative hazards resulting in a small bias of the estimated cumulative incidence functions. A data-driven optimal choice of (penalized) spline basis functions and knots would be subject for further research.

2.4 Discussion

We proposed a pattern mixture approach to competing risks analysis, as an alternative or supplementary analysis to the standard analyse based on cause-

Table 2.7: True cumulative incidence probabilities for simulation studies

Time	Relapse	GvHD	Infections	Other
0.25	0.005	0.040	0.022	0.050
0.50	0.022	0.065	0.037	0.071
1.00	0.057	0.080	0.046	0.087
2.50	0.102	0.091	0.050	0.096
15	0.179	0.103	0.054	0.120

Table 2.8: Vertical modeling compared with nonparametric cumulative incidence estimates. Reported are bias (root mean squared error) with respect to the true cumulative incidence probabilities; all numbers are divided by 10^{-3}.

		Vertical model				Non-parametric			
n	Time	Relapse	GvHD	Infections	Other	Relapse	GvHD	Infections	Other
500	0.25	1.25 (3.66)	0.04 (8.56)	-0.07 (6.14)	1.70 (9.56)	0.01 (3.28)	-0.31 (8.83)	-0.22 (6.37)	0.03 (9.77)
	0.50	1.67 (6.69)	-1.36 (10.97)	-0.79 (8.14)	0.27 (11.44)	0.18 (6.59)	-0.30 (11.11)	-0.33 (8.31)	0.19 (11.49)
	1.00	1.27 (10.53)	-0.37 (12.22)	-0.67 (9.21)	-0.25 (12.45)	0.19 (10.53)	-0.22 (12.30)	-0.31 (9.28)	0.26 (12.54)
	2.50	-0.05 (13.56)	-0.05 (12.72)	-0.04 (9.58)	1.02 (13.07)	0.02 (13.71)	-0.13 (12.76)	-0.36 (9.62)	0.02 (13.08)
	15.00	0.01 (17.42)	-0.01 (13.62)	-0.04 (9.98)	0.26 (14.57)	0.06 (17.65)	-0.17 (13.69)	-0.41 (10.04)	0.27 (14.78)
1000	0.25	1.25 (2.78)	0.26 (5.86)	0.12 (4.47)	-1.65 (6.92)	0.01 (2.30)	0.07 (6.04)	0.03 (4.67)	0.12 (6.97)
	0.50	1.49 (4.79)	1.10 (7.71)	-0.40 (5.87)	0.19 (8.22)	0.05 (4.64)	-0.01 (7.82)	0.04 (5.96)	0.07 (8.33)
	1.00	1.20 (7.37)	-0.19 (8.51)	-0.27 (6.64)	-0.45 (9.11)	0.07 (7.34)	-0.02 (8.57)	0.09 (6.70)	0.11 (9.15)
	2.50	-0.05 (9.50)	-0.45 (8.97)	0.00 (6.89)	0.86 (9.56)	0.14 (9.58)	-0.03 (9.04)	0.12 (6.93)	0.08 (9.62)
	15.00	0.11 (12.33)	-0.15 (9.61)	0.11 (7.21)	0.16 (10.45)	0.03 (12.51)	-0.16 (9.70)	0.09 (7.25)	0.17 (10.58)
5000	0.25	1.41 (1.77)	0.36 (2.68)	0.07 (2.02)	-1.77 (3.47)	-0.01 (1.02)	0.08 (2.75)	-0.04 (2.11)	0.04 (3.11)
	0.50	1.36 (2.43)	-0.97 (3.55)	-0.46 (2.65)	0.22 (3.63)	-0.03 (2.05)	0.08 (3.49)	-0.01 (2.67)	0.10 (3.65)
	1.00	1.10 (3.41)	-0.09 (3.81)	-0.37 (2.98)	-0.45 (3.94)	-0.02 (3.25)	0.09 (3.83)	0.00 (2.99)	0.10 (3.96)
	2.50	-0.14 (4.25)	-0.35 (4.06)	-0.15 (3.11)	0.84 (4.22)	-0.02 (4.29)	0.07 (4.08)	0.01 (3.12)	0.13 (4.17)
	15.00	0.00 (5.52)	0.12 (4.31)	0.01 (3.23)	0.11 (4.63)	0.00 (5.59)	0.12 (4.35)	0.02 (3.24)	0.09 (4.69)

specific hazards used for competing risks data. It is based on a decomposition $P(T, D) = P(D|T) \cdot P(T)$ of the joint distribution $P(T, D)$ of the time and cause of failure that is in our view more natural than the better known decomposition $P(T, D) = P(T|D) \cdot P(D)$ of Larson and Dinse (1985). The components $P(D|T)$ and $P(T)$ of the decomposition correspond to directly observable quantities, the relative hazard and the total hazard. It is straightforward to build models for the relative and overall hazards, both with and without covariates, and to estimate the underlying parameters, using standard statistical software. We illustrated this using data on different causes of death in the context of bone marrow transplantation.

Although the approach is not completely new, having appeared at last implicitly before in literature (see, e.g., Hachen (1988), Smits et al. (2000)), as far as

we know, this is the first account with a deeper study of its properties, and with the inclusion of covariates. From the EBMT data analysis it became clear that vertical modeling is a useful alternative in cases where the proportional hazards assumption on the cause-specific and/or the subdistribution hazards is not realistic (see the effect of disease subclassification on death due to relapse). The vertical modeling approach without covariates, assuming smooth underlying relative hazards and estimating these with a multinomial regression model with cubic splines, resulted in a gain in efficiency in estimating the cumulative incidence functions, relative to the nonparametric approach based on the Nelson-Aalen estimator of the cause-specific hazards.

Concerning the modeling aspect, two models are needed. The most obvious model for the time to failure is a proportional hazard model where all failures are considered as events, irrespective of the cause of failure. Here, standard statistical software can be used. For the cause of failure part, the most natural model is a multinomial logistic model. The easiest way is (multinomial) logistic regression with covariates and pre-specified functions of time and, possibly, their interactions as predictors. One advantage of this approach is that the hazard functions need not to be proportional. Another advantage of this approach is the fact that it is easier to apply dimension reduction techniques to the $P(D|T)$ than in proportional cause-specific hazards as in Fiocco et al. (2005), since reduced ranks are more widely used in generalized linear models. In dealing with multiple time functions the methods of Perperoglou et al. (2006) may be considered.

One advantage of the vertical modeling approach that is worthy of further investigation is that it is straightforward to deal with missing cause of failure. These missing causes of failure only influence the estimates of the relative hazards, not the overall hazard. Thus, the information that is present in the incomplete data is used in a natural way. Vertical modeling could be an attractive alternative to the methods of Goetghebeur and Ryan (1995), which are quite involved.

Finally, it is worth pointing out the relation between decomposition (2.2) and methods of generating competing risks data. A natural way of generating competing risks data (Dabrowska, 1995; Fiocco et al., 2008; Beyersmann et al., 2009) is to generate an overall failure time t according to $P(T)$ and then to sample from the causes of failure according to $P(D|T = t)$. The ingredients $P(T)$ and $P(D|T)$ are again the total hazard and the relative hazards. In fact we have used these ideas in our simulation study.

Appendix: The standard errors of cumulative hazards from vertical modeling

In this appendix we derive the formulas for the standard errors of the cause-specific cumulative hazards from the vertical modeling.

The Nelson-Aalen estimator of $\Lambda_\bullet(t)$, denoted by $\widehat{\Lambda}_\bullet(t)$, makes jumps of size $d\widehat{\Lambda}_\bullet(.)$ at time points $0 \leq t_1 < t_2 < \ldots < t_M < \infty$ with covariance matrix

$$\text{var}(d\widehat{\Lambda}_\bullet) = \text{var}\big((d\widehat{\Lambda}_\bullet(t_1), \ldots, d\widehat{\Lambda}_\bullet(t_M))^\top\big) = \begin{pmatrix} \tau_1^2 & 0 & & & 0 \\ 0 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 0 \\ 0 & & & 0 & \tau_M^2 \end{pmatrix}. \tag{2.18}$$

Relevant quantities for our purposes are the relative hazards $\pi_j(t)$; we model them as in formula (2.8).

Remark. Often it will convenient to retain the system (2.8) and to work with the $p \times J$ Fisher information matrix of $\boldsymbol{\beta}^\top = \big(\beta_1, \ldots, \beta_J\big)^\top$, denoted by $\mathcal{I}_{\boldsymbol{\beta}}$ which has rank $p(J-1)$ and, in particular, is not invertible. Let $\Sigma_{\boldsymbol{\beta}}$ denote a Moore-Penrose generalized inverse of $\mathcal{I}_{\boldsymbol{\beta}}$.

We are now interested to develop a formula for the covariance matrix $\text{var}(\widehat{\boldsymbol{\Lambda}}) =: \Sigma_{\boldsymbol{\Lambda}}$ of the estimator $\widehat{\Lambda}_k(t) = \sum_{s \leq t} \widehat{\pi}_k(t_s)\widehat{\lambda}_\bullet(t_s) = \sum_{s \leq t} \widehat{\lambda}_{ks}$, where $\widehat{\lambda}_{ks} = \widehat{\pi}_k(t_s)\widehat{\lambda}_\bullet(t_s)$. First, we introduce some notation, as follows:

$$\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda_\bullet(t_1), \ldots, \lambda_\bullet(t_M))^\top ,$$

$$\boldsymbol{\Lambda} = \big(\Lambda_1(t_1), \Lambda_2(t_1), \ldots, \Lambda_J(t_1), \Lambda_1(t_2), \ldots, \Lambda_J(t_2), \Lambda_1(t_M), \ldots, \Lambda_J(t_M)\big)^\top ,$$

and

$$\boldsymbol{\lambda} = \big(\lambda_{11}, \lambda_{21}, \ldots, \lambda_{J1}, \lambda_{12}, \ldots, \lambda_{J2}, \lambda_{1M}, \ldots, \lambda_{JM}\big) .$$

According to the Delta-method, we get

$$\Sigma_{\boldsymbol{\Lambda}} = \frac{\partial \boldsymbol{\Lambda}}{\partial \boldsymbol{\lambda}} \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}} \text{var}(\widehat{\boldsymbol{\theta}}) (\frac{\partial \boldsymbol{\Lambda}}{\partial \boldsymbol{\lambda}} \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}})^\top . \tag{2.19}$$

It is straightforward to see that

$$\frac{\partial \boldsymbol{\Lambda}}{\partial \boldsymbol{\lambda}} = \begin{pmatrix} I_{J\times J} & 0 & 0 & & & 0 \\ I_{J\times J} & I_{J\times J} & 0 & & & 0 \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & \ddots & 0 \\ I_{J\times J} & I_{J\times J} & & \dots & & I_{J\times J} \end{pmatrix} \tag{2.20}$$

where $I_{J\times J}$ is the identity matrix of order J. Also, we have

$$\frac{\partial \lambda_{ks}}{\partial \beta_{lu}} = -\pi_k(t_s)\pi_l(t_s)\lambda_\bullet(t_s)B_u(t_s) \ ,$$

where $k, l \in \{1, \dots, J\}$, $k \neq l$, $s \in \{1, \dots, M\}$, $u \in \{1, \dots, p\}$, and

$$\frac{\partial \lambda_{ks}}{\partial \beta_{ku}} = \pi_k(t_s)\big[1 - \pi_k(t_s)\big]\lambda_\bullet(t_s)B_u(t_s) \ ,$$

where $k \in \{1, \dots, J\}$, $s \in \{1, \dots, M\}$, $u \in \{1, \dots, p\}$.

Moreover,

$$\frac{\partial \lambda_{ks}}{\partial \lambda_\bullet(t_r)} = \pi_k(t_s)\delta_{s,r} \ ,$$

where $k \in \{1, \dots, J\}$, $s, r \in \{1, \dots, M\}$ and δ stands for the Kronecker delta.

Setting, for $t \geq 0$,

$$\begin{aligned} \boldsymbol{\Omega}(t) &= \begin{pmatrix} \pi_1(t) & 0 & \dots & 0 \\ 0 & \pi_2(t) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \pi_J(t) \end{pmatrix} \\ &- \big(\pi_1(t), \pi_2(t), \dots, \pi_J(t)\big)^\top \big(\pi_1(t), \pi_2(t), \dots, \pi_J(t)\big) \ , \end{aligned}$$

$$\boldsymbol{\alpha}(t) = \big(\lambda_\bullet(t)B_1(t), \lambda_\bullet(t)B_2(t), \dots, \lambda_\bullet(t)B_J(t)\big)$$

and

$$\boldsymbol{\Pi}(t_s) = \big(\pi_1(t_s), \pi_2(t_s), \dots, \pi_J(t_s)\big)^\top, \ s \in \{1, \dots, M\} \ ,$$

we get

$$\frac{\partial(\lambda_{1s}, \lambda_{2s}, \dots, \lambda_{Js})}{\partial(\beta_1, \beta_2, \dots, \beta_J)} = \boldsymbol{\Omega}(t_s) \otimes (\boldsymbol{\alpha}(t_s))^\top, \ s \in \{1, \dots, M\} \ , \tag{2.21}$$

where $\otimes$ stands for the Kronecker product, and finally

$$\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{\Omega}(t_1) \otimes (\boldsymbol{\alpha}(t_1))^\top & \boldsymbol{\Pi}(t_1) & 0 & 0 & & & 0 \\ \boldsymbol{\Omega}(t_2) \otimes (\boldsymbol{\alpha}(t_2))^\top & 0 & \boldsymbol{\Pi}(t_2) & 0 & & & 0 \\ & & & 0 & \ddots & \ddots & & \\ \vdots & & \vdots & & \ddots & \ddots & \ddots & \\ & & & & & \ddots & \ddots & 0 \\ \boldsymbol{\Omega}(t_M) \otimes (\boldsymbol{\alpha}(t_M))^\top & 0 & 0 & 0 & \dots & 0 & \boldsymbol{\Pi}(t_M) \end{pmatrix} \tag{2.22}$$

As a result, we obtain

$$\Sigma_{\boldsymbol{\lambda}} = \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}} \left(\begin{array}{c|ccc} \Sigma_{\boldsymbol{\beta}} & 0 & 0 & 0 \\ - & - & - & - \\ 0 & \tau_1^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \tau_M^2 \end{array} \right) (\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}})^\top \ . \tag{2.23}$$

In conclusion, using (2.19), (2.20) and (2.23), we have that

$$\Sigma_{\boldsymbol{\Lambda}} = \begin{pmatrix} \mathbf{W}_1 \Sigma_{\boldsymbol{\beta}} \mathbf{W}_1 + \widetilde{\boldsymbol{\Pi}}_1 & \mathbf{W}_1 \Sigma_{\boldsymbol{\beta}} \mathbf{W}_2 + \widetilde{\boldsymbol{\Pi}}_1 & \dots & \mathbf{W}_1 \Sigma_{\boldsymbol{\beta}} \mathbf{W}_M + \widetilde{\boldsymbol{\Pi}}_1 \\ \mathbf{W}_2 \Sigma_{\boldsymbol{\beta}} \mathbf{W}_1 + \widetilde{\boldsymbol{\Pi}}_1 & \mathbf{W}_2 \Sigma_{\boldsymbol{\beta}} \mathbf{W}_2 + \widetilde{\boldsymbol{\Pi}}_2 & \dots & \mathbf{W}_2 \Sigma_{\boldsymbol{\beta}} \mathbf{W}_M + \widetilde{\boldsymbol{\Pi}}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_M \Sigma_{\boldsymbol{\beta}} \mathbf{W}_1 + \widetilde{\boldsymbol{\Pi}}_1 & \mathbf{W}_M \Sigma_{\boldsymbol{\beta}} \mathbf{W}_2 + \widetilde{\boldsymbol{\Pi}}_2 & \dots & \mathbf{W}_M \Sigma_{\boldsymbol{\beta}} \mathbf{W}_M + \widetilde{\boldsymbol{\Pi}}_M \end{pmatrix} , \tag{2.24}$$

where

$$\mathbf{W}_k = \sum_{s=1}^{k} \boldsymbol{\Omega}(t_s) \otimes (\boldsymbol{\alpha}(t_s))^\top, \ k \in \{1, \dots, M\} \ ,$$

and

$$\widetilde{\boldsymbol{\Pi}}_k = \sum_{s=1}^{k} \tau_s^2 \boldsymbol{\Pi}(t_s) (\boldsymbol{\Pi}(t_s))^\top, \ k \in \{1, \dots, M\} \ .$$

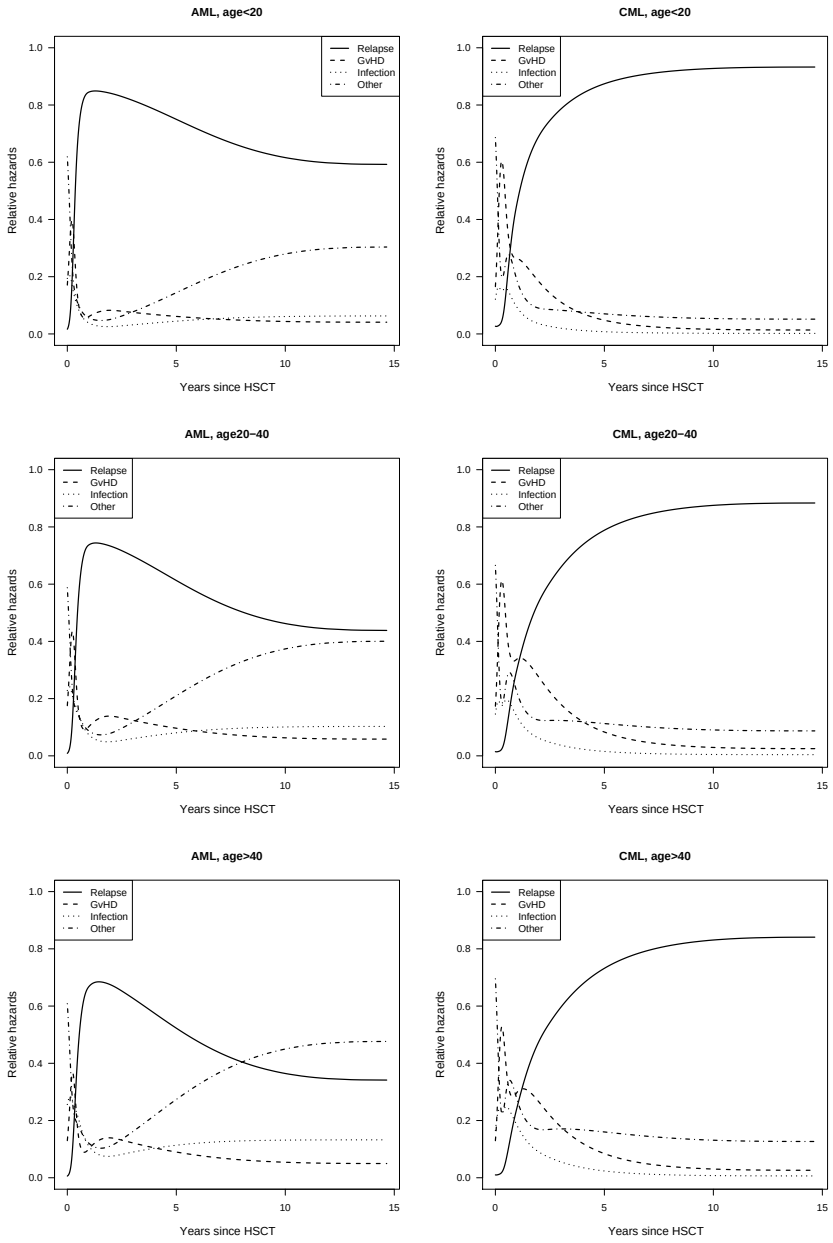


Figure 2.10: The estimated relative hazards of the multivariate model for combination of disease subclassification and age category

3

Vertical modeling: analysis of competing risks data with missing causes of failure

Abstract

We propose vertical modeling as a natural approach to the problem of analysis of competing risks data when failures types are missing for some individuals. Under a natural missing-at-random assumption for these missing failure types, we use the observed data likelihood to estimate its parameters and show that the all-cause hazard and the relative hazards appearing in vertical modeling are indeed key quantities of this likelihood. This fact has practical implications in that it suggests vertical modeling as a simple and attractive method of analysis in competing risks with missing causes of failure; all individuals are used in estimating the all-cause hazard and only those with non-missing cause of failure for relative hazards. The relative hazards also appear in a multiple imputation approach to the same problem proposed by Lu and Tsiatis and in the EM-algorithm. We compare the vertical modeling approach with the method of Goetghebeur and Ryan for a breast cancer data set, highlighting the different aspects they contribute to the data analysis.

3.1 Introduction

The problem of missing causes of failure for a subgroup of individuals in competing risks data arises frequently in practice. For instance, in the medical context information on mortality may be lost or not collected (e.g. forms are not fully completed), or the cause of failure for some individuals may be difficult to determine (e.g. patients die without autopsy). In the industrial context, the determination of the cause of failure of a system made up of multiple components connected in series may be expensive or may be very difficult to observe due to the lack of appropriate diagnostics (Park, 2005) The statistical literature addresses the inference problem in this setting. Some simple methods include analyses based on omitting cases with unknown failure type or recoding these cases as due to a certain cause (e.g., the cause of interest in case of a lethal disease) and then running a standard analysis. Here, the main drawback is substantial bias and power loss. As to assessing the covariate effects through more reasonable methods, Goetghebeur and Ryan (1995) proposed a semiparametric proportional hazards model on the cause-specific hazards, Lu and Tsiatis (2001) use multiple imputation procedures to impute the missing cause of failure, Craiu and Duchesne (2004) considered the problem via the EM algorithm on the traditional cause-specific hazards approach to competing risks, Park (2005) considered the problem via the EM algorithm on the latent failure time approach and Lu and Liang (2008) studied the semiparametric additive hazard model.

Recently, Nicolaie et al. (2010) proposed a new mixture approach to competing risks, called vertical modeling, which factorizes the joint probability distribution of time of failure T and cause of failure D according to $P(T, D) = P(T)P(D|T)$, which corresponds to natural observable quantities in these data, namely, time to failure and cause of failure given a failure occurred. In this paper, we show how this makes a natural, easy to implement approach to the above mentioned problem, because missing causes of failure affect precisely only the last component in this factorization.

The remainder of the paper is organized as follows: in Section 2 we introduce notation and general concepts in competing risks with missing causes of failure, without any particular distributional assumption. In Section 3 we discuss three methods in more detail: vertical modeling, the Goetghebeur and Ryan method (Goetghebeur and Ryan, 1995) and the Lu and Tsiatis method (Lu and Tsiatis, 2001). In Section 4 we analyze data from the Eastern Cooperative Oncology Group (ECOG) (Cummings et al., 1986) by means of two methods. In Section 4.1 we approach the data through Goetghebeur and Ryan's method. In Section 4.2 we analyze the same data by means of vertical modeling. In Section 5 we study vertical modeling in the context of the existing methods. We conclude in Section 6 with a discussion. Major technical derivations are contained in the

Appendices A and B.

3.2 Competing risks data with missing causes of failure

Notation and concepts

Suppose that data are available from n individuals each of whom can experience one of J types of failure, which we term $1, \ldots, J$, respectively, or can be subject to a noninformative censoring. Let $\tilde{T}$ denote the time of failure, C the censoring time, and D the cause of failure. Let $\mathbf{Z}$ denote a p-vector of covariates. In the absence of missing causes of failure, the observed data for individual i is $(T_i, \Delta_i, \mathbf{Z}_i)$, for $i = 1, \ldots, n$, where $T_i = \min(\tilde{T}_i, C_i)$ is the earliest of failure and censoring time, and $\Delta_i = \mathbf{1}\{\tilde{T}_i < C_i\} \cdot D_i$ is the cause of failure in case of failure and 0 in case of censoring. The usual requirement of conditional independence of $(\tilde{T}, D)$ and C, given $\mathbf{Z}$, is assumed to be true here as well. Data from different individuals are supposed to be independent.

Suppressing covariates in the notation for a moment, a key concept in competing risks modeling is the cause-specific hazard of cause j

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, D = j | T \geq t)}{\Delta t} \ , \ \Lambda_j(t) = \int_0^t \lambda_j(s) ds,$$

for $j = 1, \ldots, J$. We will assume continuity of the distribution of T and define the total, overall, or all-cause hazard

$$\lambda_\bullet(t) = \sum_{j=1}^{J} \lambda_j(t) \ , \ \Lambda_\bullet(t) = \int_0^t \lambda_\bullet(s) ds = \sum_{j=1}^{J} \Lambda_j(t).$$

In what follows, we shall refer to $\lambda_\bullet(t)$ ($\Lambda_\bullet(t)$) as the *total (cumulative) hazard*.

The survival function, defined as $S(t) = P(T > t)$, corresponds to (Putter et al., 2007)

$$S(t) = \exp(-\Lambda_\bullet(t)) \ .$$

The cumulative incidence function of cause j is defined by (Putter et al., 2007)

$$F_j(t) = \int_0^t \lambda_j(s) S(s-) ds, \ j = 1, \ldots, J. \tag{3.1}$$

As in Nicolaie et al. (2010), we define also the *relative cause-specific hazard* of

cause j

$$\pi_j(t) = \frac{\lambda_j(t)}{\lambda_\bullet(t)},\ j = 1, \ldots, J. \tag{3.2}$$

Reversal of the definition (3.2) gives the cause-specific hazard of cause j in terms of the relative hazard of cause j and total hazard as

$$\lambda_j(t) = \pi_j(t)\lambda_\bullet(t),\ j = 1, \ldots, J. \tag{3.3}$$

Missing causes of failure; assumptions

In case missing causes of failure occur, let R be an indicator variable taking values zero or one depending on whether the cause of failure is reported (including censoring) or not (in which case cause of failure is missing). In this case, the observed data for individual i is $\mathbf{Y}_i = (T_i, \Delta_i, \mathbf{Z}_i)$ if $R_i = 0$ and $\mathbf{Y}_i = (\tilde{T}_i, \mathbf{Z})$ if $R_i = 1$, independent across subjects i.

We assume that the missingness mechanism is missing at random (MAR) (Rubin, 1976) that is, the probability of missing information depends only on the observed data. In our context this means that the probability of a failure cause being missing, given failure time, covariates and given a failure occurred, does not depend on the cause; that is, for every individual i with $D_i > 0$

$$P(R_i = 1|T_i = t, D_i, \mathbf{Z}_i) = P(R_i = 1|T_i = t, \mathbf{Z}_i). \tag{3.4}$$

In fact, we will assume ignorability as well, which means that in addition to the MAR assumption, the parameters of the data model and any parameters in (3.4) are distinct. This implies that we do not need to consider the model for R for making inference about $\boldsymbol{\theta}$ based on the observed data (Little and Rubin, 1987). Assumption (3.4) implies that R_i and D_i are independent, given T_i and $\mathbf{Z}_i$ for $D_i > 0$, expressed equivalently as

$$\begin{aligned} P(D_i = j|R_i = 1, T_i = t, \mathbf{Z}_i) &= P(D_i = j|R_i = 0, T_i = t, \mathbf{Z}_i) \\ &= P(D_i = j|T_i = t, \mathbf{Z}_i), \end{aligned} \tag{3.5}$$

for $j = 1, \ldots, J$. Note that $P(D_i = j|T_i = t, \mathbf{Z}_i) = \pi_j(t|\mathbf{Z}_i),\ j = 1, \ldots, J$. Intuitively, the MAR assumption expresses the idea that patients who died at time t due to a known cause of failure are representative of all patients who died at time t, irrespective whether the cause of failure was observed or no. This implies that estimation of the parameters of a model for relative hazards uses only the patients with observed cause of failure, as expressed in (3.5).

Note that these two assumptions on the missingness mechanism are common to all the approaches described in Section 3.

Observed likelihood

Define $\mathcal{D}_j$ and $\mathcal{D}_u$ as the set of subjects with failure of cause j, $j = 1, \ldots, J$, and of unknown cause, respectively, $\mathcal{D}_{knw} = \bigcup_{j=1}^{J} \mathcal{D}_j$ as the set of subjects with known cause of failure, and let $\mathcal{D} = \mathcal{D}_{knw} \cup \mathcal{D}_u$ denote all failures.

For simplicity of notation, we suppress the dependence on covariates in the notation. The likelihood is a product of contributions of the individuals, which can be divided into three categories. The contribution to the likelihood of a patient i who is censored at time t_i is given by

$$P(\widetilde{T}_i > t_i)P(C_i = t_i).$$

A patient i who died at time t_i due to an unknown cause contributes

$$P(\widetilde{T}_i = t_i)P(C_i > t_i)P(R_i = 1|\widetilde{T}_i = t_i),$$

while a patient i who died at time t_i due to cause j contributes

$$P(\widetilde{T}_i = t_i)P(C_i > t_i)P(D_i = j|\widetilde{T}_i = t_i)P(R_i = 0|\widetilde{T}_i = t_i).$$

These equations follow from independence of $(\widetilde{T}, D)$ and C, and from the MAR assumption implying (3.5). Due to the ignorability assumption, the distribution of R can be omitted from the full observed likelihood. If we assume that the distributions of $\widetilde{T}$ and C have no common parameters, then we can omit the contribution of C to the likelihood as well. Therefore, after rearranging terms, the full likelihood is given by

$$\prod_{i=1}^{n} \left[P(\widetilde{T}_i > t_i)^{1\{D_i=0\}} P(\widetilde{T}_i = t_i)^{1\{D_i>0\}} \right] \prod_{i \in \mathcal{D}_{knw}} \prod_{j=1}^{J} P(D_i = j|\widetilde{T}_i = t_i). \quad (3.6)$$

3.3 Models for competing risks with missing causes of failure

3.3.1 Vertical modeling

We introduce the vertical modeling approach of Nicolaie et al. (2010) as a tool for dealing with missing causes of failure in competing risks.

Vertical modeling for competing risks

Conceptually, the basic idea behind vertical modeling is the decomposition $P(T, D) = P(T)P(D|T)$ of the joint distribution of time and cause of failure. The compo-

nents $P(T)$ and $P(D|T)$ of the decomposition correspond to directly observable quantities, the total hazard $\lambda_\bullet$ and the relative hazards π_j. By (3.1) and (3.3), the cumulative incidence function of cause j may be expressed according to the vertical modeling approach in terms of the previous concepts as the product of the failure time distribution multiplied by the conditional distribution of cause given a failure occurred which yields (Nicolaie et al., 2010)

$$F_j(t) = \int_0^t \pi_j(s)\lambda_\bullet(s)S(s-)ds \ . \tag{3.7}$$

Specifically, the vertical modeling approach to competing risks relies on models for the total hazard and the relative hazards, rather than models for the cause-specific hazards in order to describe the joint distribution of time and cause of failure, $F_j(t)$.

Vertical modeling for competing risks with missing causes of failure

The most appealing feature of vertical modeling in the presence of missing causes of failure is its ease of implementation. No ad-hoc or time-consuming software is needed; it can be fitted with standard statistical software.

In the presence of missing causes of failure, vertical modeling naturally separates into the two main components of competing risks, where missing causes of failure are either irrelevant (total hazard) and where missing causes of failure are relevant (relative hazards). Modeling the rate at which a failure occurs through a model for the total hazard and estimating the corresponding regression parameters involve only the use of information on the failure or censoring times and covariates, eventually, and it is therefore insensitive to missing causes of failure. In contrast, modeling the cause of failure in case of failure requires caution due to the missing information on the actual cause of failure. To clarify this point, suppose that the total hazard and the relative hazards are parameterized by parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$. Specific models for the total hazard and for the relative hazards will be considered at a later point.

Using the concepts of relative and total hazards it is straightforward to see that the observed likelihood (3.6) can be rewritten as

$$L(\boldsymbol{\theta}) = L_1(\boldsymbol{\beta}) \cdot L_2(\boldsymbol{\gamma}), \tag{3.8}$$

where

$$L_1(\boldsymbol{\beta}) = \prod_{i=1}^{n} (\lambda_\bullet(t_i))^{1\{D_i>0\}} S(t_i)$$

and

$$L_2(\boldsymbol{\gamma}) = \prod_{i \in \mathcal{D}_{knw}} \prod_{j=1}^{J} (\pi_j(t_i))^{1\{D_i=j\}}.$$

Equation (3.8) says that the likelihood factorizes into two parts, each involving one of the two ingredients of the vertical modeling approach. The first part, $L_1(\boldsymbol{\beta})$, for the survival time (total hazard) ignores the cause of failure and uses all the observations; the second part, $L_1(\boldsymbol{\gamma})$, for the cause of failure given survival time (relative hazards) uses only the failures with known cause. Since the model is parameterized in such a way that the parameters appearing in the total hazard and those in the relative hazard are distinct, we can maximize the likelihood (3.8) by separately maximizing the total hazard (fitting an overall survival model) and the relative hazard (fitting a multinomial logistic regression model on the events with known cause). This greatly simplifies analysis and makes it no more involved than in the case of data with known causes of failure. If $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are each estimated using maximum likelihood, then vertical modeling will yield maximum likelihood estimators, and hence is fully efficient.

3.3.2 The proportional cause-specific hazard model

For ease of reference we assume only two causes of failure from now on, that is, $J = 2$. Goetghebeur and Ryan (1995) proposed a method of assessing covariate effects in competing risks data when some failure types are missing, based on a standard proportional hazards structure for each of the cause-specific hazards of the two failure types, that is

$$\lambda_j(t|\mathbf{Z}) = \lambda_{j0}(t) \exp(\boldsymbol{\eta}_j^\top \mathbf{Z}),\ j = 1, 2. \tag{3.9}$$

They assume that the ratio between the baseline cause-specific hazards for the two causes of failure is constant and indicate extensions when this baseline hazard ratio is expressed through a simple parametric function of time. Denote the vector of regression parameters associated with this ratio model by $\boldsymbol{\xi}$. Then, a two-step Cox partial likelihood-like procedure is used iteratively to estimate the regression parameters $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\xi})$. In fact, this method involves calculation of the relative hazards (3.2) for failures with missing cause of death as weighted contributions of the unknown deaths to the score equations. The first step consists in maximizing

a Cox partial likelihood, defined as

$$L_{GR}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2|\boldsymbol{\xi}) = \prod_{i\in\mathcal{D}_1} \frac{e^{\boldsymbol{\eta}_1^\top \mathbf{Z}_i}}{\sum_{j\in R(t_i)} e^{\boldsymbol{\eta}_1^\top \mathbf{Z}_j}} \prod_{i\in\mathcal{D}_2} \frac{e^{\boldsymbol{\eta}_2^\top \mathbf{Z}_i+\xi}}{\sum_{j\in R(t_i)} e^{\boldsymbol{\eta}_2^\top \mathbf{Z}_j+\xi}} \cdot \prod_{i\in\mathcal{D}_u} \frac{e^{\boldsymbol{\eta}_1^\top \mathbf{Z}_i} + e^{\boldsymbol{\eta}_2^\top \mathbf{Z}_i+\xi}}{\sum_{j\in R(t_i)} \left(e^{\boldsymbol{\eta}_1^\top \mathbf{Z}_j} + e^{\boldsymbol{\eta}_2^\top \mathbf{Z}_j+\xi}\right)}, \tag{3.10}$$

based on the conditional probabilities of a specific event given that one event of that type occurs from the risk set at that time. This will result only in estimators of the regression coefficients $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$, due to the fact that ξ is assumed to be known. The second step consists in maximizing a Cox partial likelihood-like, defined as

$$L^*_{GR}(\boldsymbol{\xi}|\boldsymbol{\eta}_1, \boldsymbol{\eta}_2) = \frac{\prod_{i\in\mathcal{D}_1} e^{\boldsymbol{\eta}_1^\top \mathbf{Z}_i} \prod_{i\in\mathcal{D}_2} e^{\boldsymbol{\eta}_2^\top \mathbf{Z}_i+\xi} \prod_{i\in\mathcal{D}_u} \left(e^{\boldsymbol{\eta}_1^\top \mathbf{Z}_i} + e^{\boldsymbol{\eta}_2^\top \mathbf{Z}_i+\xi}\right)}{\prod_{i\in D} \sum_{j\in R(t_i)} \left(e^{\boldsymbol{\eta}_1^\top \mathbf{Z}_j} + e^{\boldsymbol{\eta}_2^\top \mathbf{Z}_j+\xi}\right)}, \tag{3.11}$$

based on the conditional probabilities of an event of specific type, given that one event occurs, but without conditioning on the type of event. This second partial likelihood uses the estimated values of the regression parameters from the first step $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ and results in an estimator of ξ. Although the method is specifically aimed at estimating the cause-specific hazard ratio, without covariates it reduces to estimating the (possibly, time-varying) ratio between the baselines. We shall refer to his approach as the proportional cause-specific hazard model with constant baseline hazard ratio.

An important difference between vertical modeling and the above mentioned model is that vertical modeling separates the parameters for the total hazard and relative hazards, while the proportional cause-specific hazard model does not, which can be seen from the identity (3.3) which retrieves the cause-specific hazards for cause j from the vertical modeling approach. From a vertical model we can obtain the covariate effect on the cause-specific hazards through (3.3). The hazard ratio following from a vertical model will typically *not* be time-constant. We will come back to this issue in Section 5.

3.3.3 Multiple imputation

Lu and Tsiatis (2001) proposed a multiple imputation procedure to the same problem of missing causes of failure, where the cause-specific hazard for the cause of interest is modeled through a proportional hazards relationship. To impute the failure type for cases with missing cause of failure a semiparametric model on the relative hazard for the cause of interest is proposed. Parameters of the

model for the relative hazard are estimated based on the subjects with known cause of failure, as for vertical modeling. If these imputed values were used, together with a model for the total hazard, then multiple imputation could just be considered as a random version of the vertical modeling with the same models for relative and total hazards. The appeal of multiple imputation lies in the fact that the complete data sets could be used subsequently for any model, such as proportional hazards model on the cause-specific and subdistribution hazards.

3.4 Data analysis

We analyze data on 169 elderly women (over the age of 65) with stage II breast cancer prospectively randomized to receive either tamoxifen or placebo for 24 months in a clinical trial conducted by the ECOG. In the original paper, with a median follow-up of 55 months, at 4 years 80% of the patients treated with tamoxifen were still alive and 74% following placebo. No significant treatment differences were noted in overall survival, with a log-rank p-value of 0.26. The same data have been studied in the paper of Goetghebeur and Ryan (1995).

Covariate information is described in Table 3.1. Cummings et al. (1986)

Table 3.1: Prognostic factors for all patients.

Prognostic factor		n (%)
Number of positive nodes	1-3	90 (53%)
	≥ 4	79 (47%)
Estrogen receptor	negative	6 (4%)
	positive	163 (96%)
Treatment	placebo	83 (49%)
	tamoxifen	86 (51%)

reported two covariates, the number of positive axillary nodes and the estrogen receptor status (ER) of their primary tumor, as being significantly associated with overall survival.

The Kaplan-Meier survival curves for each combination of these two covariates are shown in Figure 3.1. There was one patient with 1-3 positive nodes and negative estrogen receptor status who was censored. This patient is included in the analysis but not shown in Figure 3.1.

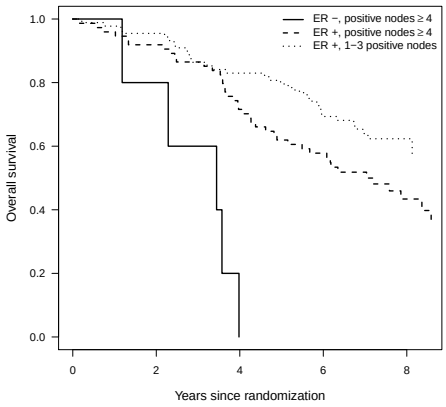


Figure 3.1: Kaplan-Meier survival curves of time to death for each combination of covariates.

In what follows, we will be interested in analyzing death due to cause 1: breast cancer and due to cause 2: other causes. Complicating factor is that for a relatively large number of patients, cause of death is unknown. A total of 79 patients died; 44 (56%) of these died of cancer, 17 (21%) of other causes. For 18 patients (21%), the cause of death is unknown. The number of events for each combination of these two covariates is reported in Table 3.2.

Table 3.2: Number of events per each combination of covariates.

Group			Events			
Number of positive nodes	Estrogen receptor	Number of patients	Cancer	Other	Unknown	Censored
1-3	negative	1	0	0	0	1
1-3	positive	89	18	6	9	56
≥ 4	negative	5	5	0	0	0
≥ 4	positive	74	21	11	9	33

Since the effect of treatment was neither significant for survival, nor for the cause of death, we will ignore treatment from now on and study the effect of the covariates number of positive nodes and estrogen receptor status, coded as indicators $Z_1 = \mathbf{I}(\geq 4$ positive nodes) and $Z_2 = \mathbf{I}$(estrogen receptor status is negative),

and let $\mathbf{Z} = (Z_1, Z_2)$.

For all methods used here, the MAR assumption is used. We shall discuss the appropriateness of this assumption in Section 6.

3.4.1 The proportional cause-specific hazard approach with constant baseline hazard ratio

We first analyze cause-specific mortality in these data using the approach of Goetghebeur and Ryan, based on modeling the cause-specific hazards as in (3.9). We follow Goetghebeur and Ryan's original assumption of the baseline hazards of cancer to be proportional to the baseline hazard of other causes, that is

$$\lambda_{20}(t) = \lambda_{10}(t) \exp(\xi). \tag{3.12}$$

The resulting estimates are shown in Table 3.3. Higher number of positive axillary lymph nodes and negative estrogen receptor status increase the cause-specific hazard of death due to cancer, and higher number of positive axillary lymph nodes increases the death rate due to other causes. No parameter estimate of estrogen receptor status is given for death due to other causes, because no patient with negative estrogen receptor status died of a cause other than cancer (see Table 3.2). In Figure 3.2 we show the estimated cumulative incidences of death due to

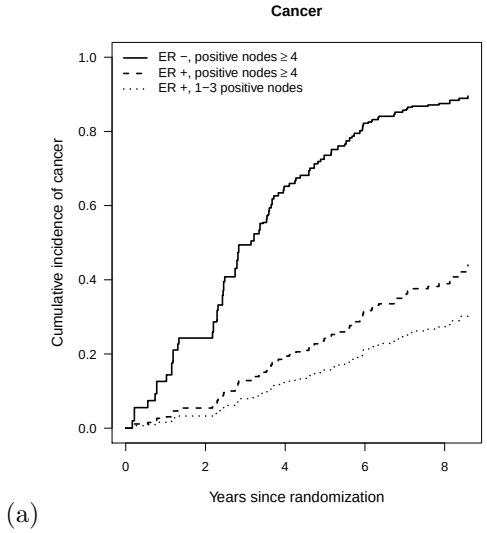
Table 3.3: A proportional hazards model.

Covariate	Cancer	Other causes	
	$\boldsymbol{\eta}_1$	η_2	ξ
Number of positive nodes ≥ 4	0.52	0.78	
Estrogen receptor ER-	1.60		
Other causes vs cancer			-1.02

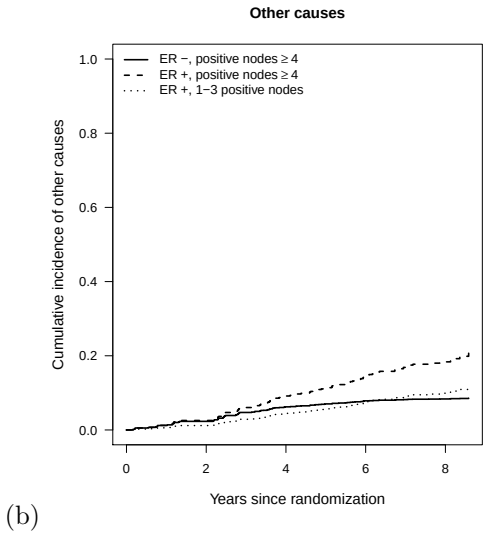
cancer and other causes for each combination of covariate values following from this model, calculated by means of the `mstate` package in `R` (de Wreede et al., 2010). For the combination 1-3 positive nodes and negative estrogen receptor status both cumulative incidence functions are identically zero.

3.4.2 Vertical modeling

Next we analyze cause-specific mortality using vertical modeling. For this purpose we need to model the total hazard and the relative cause-specific hazards. As for



(a)



(b)

Figure 3.2: Cumulative incidences of cancer (a) and other causes (b).

a model for the total hazard, we take a Cox proportional hazards model with Z_1

and Z_2 as covariates, that is

$$\lambda_\bullet(t|\mathbf{Z}) = \lambda_0(t)\exp(\beta_1 Z_1 + \beta_2 Z_2).$$

The estimated regression coefficient (SE) are 0.59 (0.23) and 1.20 (0.47), respectively. These correspond to hazard ratios of 1.80 (95% CI $1.15 - 2.84$) and 3.31 (95% CI $1.32 - 8.29$), respectively, confirming that higher number of positive axillary lymph nodes and negative estrogen are associated with lower survival. In Figure 3.3 we show the estimated survival curves of time to death for each combination of these two covariates implied by our model. They look quite similar to the nonparametric curves of Figure 3.1, which would indicate that the proportionality assumption is reasonable.

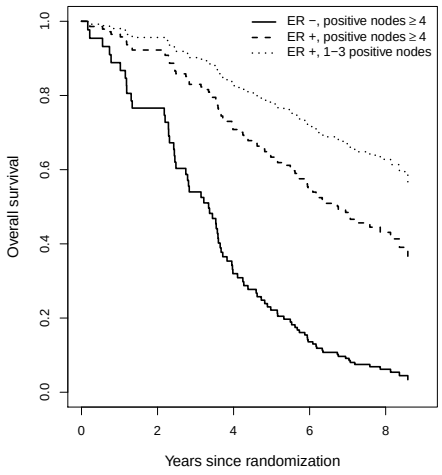


Figure 3.3: Model based survival curves of time to death for each combination of covariates.

For the relative hazards, we fitted a logistic regression model for death due to cancer on the subset of patients whose cause of death is known:

$$\text{logit}(\pi_1(t|\mathbf{Z})) = \kappa^\top \mathbf{B}(t) + \nu^\top \mathbf{Z} + \delta^\top \mathbf{B}(t) * \mathbf{Z},$$

so that $\boldsymbol{\gamma} = (\kappa^\top, \nu^\top, \delta^\top)^\top$. This model incorporates dependence of $\pi_1(t|\mathbf{Z})$ on time t, covariates and, possibly, their interactions. Here $\mathbf{B}(t)$ is a vector of time

functions, which could for instance be polynomials, piecewise constant or spline basis functions. In our application we choose spline functions of degree 2 with 2 knots on the time axis, $t_1 = 3$ and $t_2 = 6$, such that roughly the same number of events occur in each of the three intervals defined by these knots. This yields $\mathbf{B}(t) = (1, t, t^2, (t - t_1)^{+^2}, (t - t_2)^{+^2})$, where for any number a, the notation a^+ stands for $\max\{0, a\}$. Deviance and AIC for several models are presented in Table 3.4. Compared to only time, the inclusion of estrogen receptor did

Table 3.4: Deviance and AIC for different logistic regression models for the relative hazards.

Model	Deviance	AIC
$\mathbf{B}(t)$	68.370	78.370
$\mathbf{B}(t)$ and Z_2	64.804	76.804
$\mathbf{B}(t)$ and Z_1	68.339	80.339
$\mathbf{B}(t)$ and Z_1 and Z_2	64.778	78.778
$\mathbf{B}(t)$ and Z_2 and $\mathbf{B}(t)$ * Z_2	64.803	82.803

not significantly improve the model fit: the p-value of estrogen receptor status was 0.058. Nodal status had no significant effect on the relative hazards, and interactions with time were also not significant. Based on the lowest AIC, we chose the logistic regression model with main effects of time and estrogen receptor as the model for the relative hazards of cancer and other causes respectively, that is

$$\pi_1(t|\mathbf{Z}) = \frac{\exp(\kappa^\top \mathbf{B}(t) + \nu Z_2)}{1 + \exp(\kappa^\top \mathbf{B}(t) + \nu Z_2)},\ \pi_2(t|\mathbf{Z}) = 1 - \pi_1(t|\mathbf{Z}).$$

In Figure 3.4 we show a plot of the associated relative hazards of death due to cancer implied by our model for positive and negative estrogen receptor status. For ER- patients, the estimated relative hazard of death due to cancer equals one since no ER- patient died of other causes. In Figure 3.5 we show the estimated cumulative incidences of death due to cancer and other causes for each combination of the covariates, based on our model through Equation (3.7). For comparison, the cumulative incidence curves of Figure 3.2 are repeated in lighter gray lines. We see that, minor differences notwithstanding, they are qualitatively similar. The most striking difference is in the estimate of the cumulative incidence function of death due to other causes for the subgroup negative estrogen receptor and higher number of positive axillary lymph nodes: although there is no death due to other causes for these patients in the data, the estimate from the Goetghebeur and Ryan approach is increasing, while our estimate is always zero. This could be

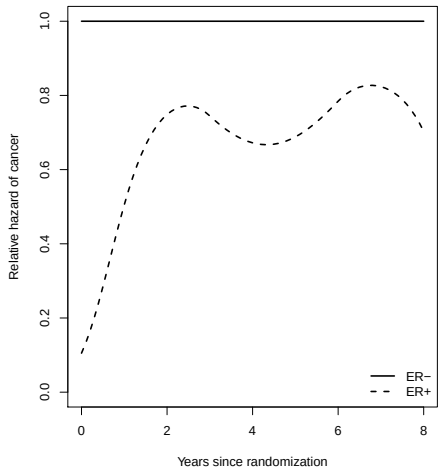


Figure 3.4: The relative hazards of death due to cancer for ER- and ER+.

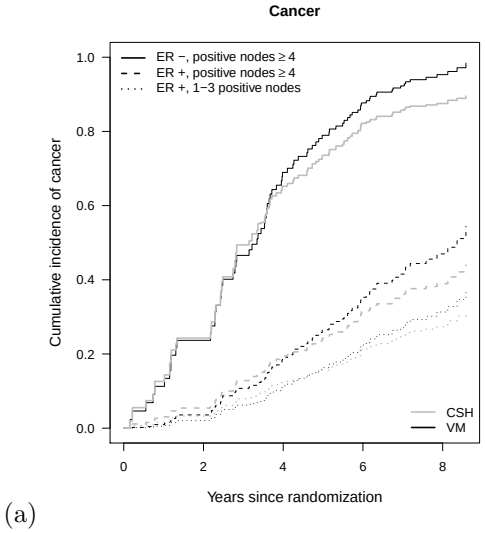
explained by the fact that in the Goetghebeur and Ryan approach occurrence of an event impacts both cause-specific hazards at the same time, via the common baseline hazard $\lambda_{10}(t)$ they share. Vertical modeling acknowledges the fact that no deaths due to other causes were observed for the subgroup negative estrogen receptor and higher number of positive axillary lymph nodes, and thus mirrors the data more closely. This feature of vertical modeling is indeed apparent from (3.7), which shows that the proportion of the density of the survival time which is attributed to the cumulative incidence function of death due to other causes is dictated by the relative hazard of other causes, and the proportion for this particular subgroup of patients is zero.

The estimates presented in this Section were obtain via the `vm` function in `R`, to be made available in the **vm** package.

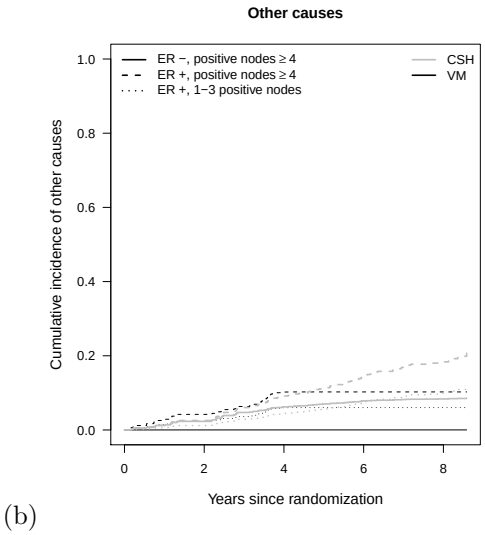
3.5 Vertical modeling in context

Vertical modeling and cause-specific hazards

It is of interest to compare the Goetghebeur and Ryan approach with vertical modeling. We take the results of Section 4 as starting points. Assume vertical



(a)



(b)

Figure 3.5: Cumulative incidences of cancer (a) and other causes (b) from Goetghebeur's approach (CSH) versus vertical modeling (VM).

modeling consists of modeling the total hazard and relative hazards as follows

$$\lambda_\bullet(t|\mathbf{Z}) = \lambda_0(t)\exp(\beta_1 Z_1 + \beta_2 Z_2)$$

and

$$\pi_1(t|\mathbf{Z}) = \frac{\exp(\kappa^\top \mathbf{B}(t) + \nu Z_2)}{1 + \exp(\kappa^\top \mathbf{B}(t) + \nu Z_2)},\ \pi_2(t|\mathbf{Z}) = 1 - \pi_1(t|\mathbf{Z}), \tag{3.13}$$

respectively.

Goetghebeur and Ryan's approach is based on the aim to estimate cause-specific hazard ratios $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ (see model (3.9)) in the presence of missing causes of failure. On the other hand, from the relation $\lambda_1(t|\mathbf{Z}) = \lambda_\bullet(t|\mathbf{Z})\pi_1(t|\mathbf{Z})$ it is straightforward to see that vertical modeling will, in general, not result in time-invariant cause-specific hazard ratios, but in a relation like

$$\lambda_j^{(0)}(t|\mathbf{Z}) = \lambda_{j0}(t)\exp(\widetilde{\boldsymbol{\eta}}_j^\top(t)\mathbf{Z}),\ j = 1, 2. \tag{3.14}$$

More specifically, from (3.3) it is straightforward to see the regression coefficient for Z_1 in model (3.14) is $\widetilde{\eta}_{11}(t) \equiv \beta_1$, while for Z_2 the corresponding regression coefficient in the same model is

$$\widetilde{\eta}_{12}(t) = \beta_2 + \log\frac{\pi_1(t|Z_2 = 1)}{\pi_1(t|Z_2 = 0)}.$$

In Figure 3.6 we show how $\widetilde{\eta}_{12}(t)$ compares with the corresponding coefficient from the Goetghebeur and Ryan approach.

Moreover, it might be interesting to investigate how a summary cause-specific hazard ratio can be obtained when fitting a vertical model, therefore when the model implies a time-varying hazard ratio (see the implied model (3.14)). By extending to competing risks a result of van Houwelingen (2007) for the case of ordinary survival, an approximation of a weighted average of $\widetilde{\boldsymbol{\eta}}_1(t)$ could be obtained given by

$$\boldsymbol{\eta}_1^* \approx \int_0^\infty \mathbf{w}(t)\widetilde{\boldsymbol{\eta}}_1(t)dt, \tag{3.15}$$

where the weights are given by

$$\mathbf{w}(t) = \frac{S(t)C(t)V_{\widetilde{\boldsymbol{\eta}}_1(t)}(\mathbf{Z}|t)\lambda_1(t)}{\int_0^\infty S(u)C(u)V_{\widetilde{\boldsymbol{\eta}}_1(u)}(\mathbf{Z}|u)\lambda_1(u)du},$$

$S(t)$ and $C(t)$ are the marginal distributions of time to failure and censoring respectively, $\lambda_1(t) = \lim_{\Delta t\to 0}\frac{1}{\Delta t}P(t \le T \le t + \Delta t, D = 1|T \ge t)$ is the marginal cause-specific hazard, and

$$\mathbf{V}_{\boldsymbol{\eta}_1(t)}(\mathbf{Z}|t) = \frac{\sum_{j\in R(t)}\mathbf{Z}_j^2\exp(\widetilde{\boldsymbol{\eta}}_1(t)^\top\mathbf{Z}_j)}{\sum_{j\in R(t)}\exp(\widetilde{\boldsymbol{\eta}}_1(t)^\top\mathbf{Z}_j)} - \left(\frac{\sum_{j\in R(t)}\mathbf{Z}_j\exp(\widetilde{\boldsymbol{\eta}}_1(t)^\top\mathbf{Z}_j)}{\sum_{j\in R(t)}\exp(\widetilde{\boldsymbol{\eta}}_1(t)^\top\mathbf{Z}_j)}\right)^2$$

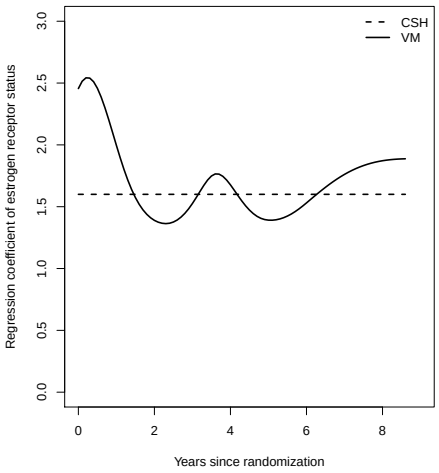


Figure 3.6: Implied time-varying regression coefficient of estrogen receptor status on the cause specific hazard of death due to cancer from Goetghebeur and Ryan's approach (CSH) versus vertical modeling (VM).

is the weighted variance of $\mathbf{Z}$ given $(T = t, D = 1)$ under the Cox model, with $\widetilde{\boldsymbol{\eta}}_1(t)$ as true effect. Technical details are given in Appendix A. Replacing the parameters in (3.15) by estimates will yield an alternative to the approach of Goetghebeur and Ryan.

The proportional hazards model with time-varying baseline hazard ratio

Now, suppose the model is parameterized by means of a Cox proportional hazards model on the cause-specific hazards, where the ratio between the baseline hazards is time-varying, rather than time constant as in Goetghebeur and Ryan's approach. For simplicity, we consider two causes of failure, with cause-specific hazards

$$\begin{aligned} \lambda_1(t \,|\, \mathbf{Z}) &= \lambda_0(t) \exp(\boldsymbol{\eta}_1^\top \mathbf{Z}) \ , \\ \lambda_2(t \,|\, \mathbf{Z}) &= \lambda_0(t) \exp(\boldsymbol{\xi}^\top \mathbf{B}(t) + \boldsymbol{\eta}_2^\top \mathbf{Z}) \ , \end{aligned}$$

with $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ denoting the effects of the covariates on the cause-specific hazards of cause 1 and 2, respectively, $\boldsymbol{\xi}$ parameterizing the (time-varying) ratio between

the baseline cause-specific hazard of cause 2 with respect to that of cause 1, and $\mathbf{B}(t)$, as before, a vector of given time functions. This approach implies that the total hazard is given by

$$\lambda_\bullet(t \,|\, \mathbf{Z}) = \lambda_0(t)\Big[\exp(\boldsymbol{\eta}_1^\top \mathbf{Z}) + \exp(\boldsymbol{\xi}^\top \mathbf{B}(t) + \boldsymbol{\eta}_2^\top \mathbf{Z})\Big] \ ,$$

and the relative hazards are given by

$$\pi_1(t \,|\, \mathbf{Z}) = \frac{\exp(\boldsymbol{\eta}_1^\top \mathbf{Z})}{\exp(\boldsymbol{\eta}_1^\top \mathbf{Z}) + \exp(\boldsymbol{\xi}^\top \mathbf{B}(t) + \boldsymbol{\eta}_2^\top \mathbf{Z})} \ , \ \pi_2(t \,|\, \mathbf{Z}) = 1 - \pi_1(t \,|\, \mathbf{Z}) \ .$$

Define

$$w_i^{(1)} = \exp(\boldsymbol{\eta}_1^\top \mathbf{Z}_i) \ , \ w_i^{(2)}(t) = \exp(\boldsymbol{\xi}^\top \mathbf{B}(t) + \boldsymbol{\eta}_2^\top \mathbf{Z}_i) \ , \ w_i^{(u)}(t) = w_i^{(1)} + w_i^{(2)}(t) \ .$$

In Appendix B it is shown that the following partial likelihood can be obtained as a profile likelihood from (3.8):

$$\frac{\prod_{i\in\mathcal{D}_1} w_i^{(1)} \prod_{i\in\mathcal{D}_2} w_i^{(2)}(t_i) \prod_{i\in\mathcal{D}_u} w_i^{(u)}(t_i)}{\prod_{i\in D} \sum_{j\in R(t_i)} w_j^{(u)}(t_i)} \ . \tag{3.16}$$

For time-fixed relative hazard (i.e., a time-constant hazard ratio between the two baseline cause-specific hazards), the partial likelihood (3.16) is identical to the second partial likelihood L_{GR}^* of Goetghebeur and Ryan (1995), see (3.11). Goetghebeur and Ryan use this partial likelihood only to estimate the $\boldsymbol{\xi}$ parameter and rely on a different partial likelihood, L_{GR}, to estimate the covariate effects $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$. Their first argument is that the partial likelihood (3.16) yields non-standard estimators of the covariate effects in case all failures are known. Second, they argue, the estimators obtained from their first partial likelihood L_{GR} are generally preferable for reasons of robustness. This last argument is no longer decisive if the assumption of a time-fixed ratio between the two cause-specific hazards is relaxed to that of a time-varying ratio. The advantage of using (3.16) in that case is a likelihood that can easily be maximized and which is more efficient than the sandwich estimators used to obtain standard errors of the methods of Goetghebeur and Ryan. Appendix B provides expressions for the variances of the maximum likelihood estimator of $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\xi})$, also in the case of time-varying relative hazards.

3.6 Discussion

Missing causes of failure in competing risks is a very common problem. Unless perhaps in the context of well-controlled clinical trials, determination of the cause of death is often postponed (and hence often not done) in favor of more urgent tasks. Outside the medical context, one can also think of situations where this may occur such as industry, where a machine may fail and the cause may be expensive to determine, and in demography, where in migration analysis it may be known when an individual left the home country, but the destination is unknown. Especially in the last example, known destination is the exception, rather than the rule.

We proposed vertical modeling for the analysis of competing risks data with missing causes of failure, as an alternative approach to some ad-hoc methods (deletion or recoding) or some more reasonable methods based on modeling cause-specific hazards, and there are two main reasons for that. The first reason is related to the ingredients of vertical modeling, the total hazard and the relative hazards, which appear in a natural way in the likelihood of the data when missing causes of failure are present. The second reason concerns its simplicity, because any model where the parameters for the total and relative hazards are separated can be analyzed by separately maximizing the likelihood contributions for the total hazard, which is not affected by missing causes of failure, and for the relative hazards, for which missing causes of failure simply may be ignored under an appropriate missing at random assumption. This approach will then yield the maximum likelihood estimators.

It is important to discuss the missing at random assumption. The same assumption underlies both the Goetghebeur and Ryan (Goetghebeur and Ryan, 1995) and Lu and Tsiatis (Lu and Tsiatis, 2001) approaches. It says that in case of failure, given the failure time and covariates, the probability of the failure cause being missing does not depend on the cause. In practice, this assumption may or may not be fulfilled. One can think of situations where some causes of death are more difficult to verify than others. If these difficult to verify causes are not investigated more closely and subsequently reported as missing causes of death, then this could lead to a violation of the missing at random assumption. If one wants unbiased estimation in the case of such informative missing causes of failure, then one would have to model the missingness mechanism as well. It is certainly of interest to pursue this, but outside the scope of this paper.

Appendix A: Derivation of (3.15)

The derivation of (3.15) follows closely van Houwelingen (2007). Consider a sample of size n and define the counting process $N_i(t) = I(T_i \geq t, D_i = 1)$. Let

$Y_i(t) = I(T_i \geq t) = I(\widetilde{T}_i \geq t, C_i \geq t)$ be the "at-risk" indicator of the counting process. We have $S(t) = P(\widetilde{T}_i \geq t)$ and $C(t) = P(C_i \geq t)$. Define

$$S^{(j)}(t) = \sum_{i \in R(t)} Z_i^j \lambda_1^{(0)}(t|Z_i),\ s^{(j)}(t) = ES^{(j)}(t), \tag{3.17}$$

$$S^{(j)}(\eta_1, t) = \sum_{i \in R(t)} Z_i^j \lambda_1(t|Z_i),\ s^{(j)}(\eta_1, t) = ES^{(j)}(\eta_1, t), \tag{3.18}$$

for $j = 0, 1, 2$, where the expectation is taken with respect to the distribution of $(T, \Delta, \mathbf{Z})$ as it is implied by the model (3.14).

By similar arguments as those of Theorem 2.1 of Struthers and Kalbfleisch (1986) we can prove that formula (2.6) of Xu and O'Quigley (2000) is still valid under competing risks, namely, the maximum partial likelihood estimator of the parameter in model (3.9), when we assume a distribution of the data as derived from the model (3.14), converges to the solution of

$$\int_0^\infty \Big[\frac{s^{(1)}(t)}{s^{(0)}(t)} - \frac{s^{(1)}(\eta_1, t)}{s^{(0)}(\eta_1, t)}\Big] s^{(0)}(t)dt = 0. \tag{3.19}$$

By similar arguments as those of Xu and O'Quigley we can show that

$$E_{\eta_1(t)}(Z|t) := \frac{s^{(1)}(t)}{s^{(0)}(t)}$$

gives a consistent estimate of the conditional expectation of Z given $(T = t, D = 1)$ under the model (3.14) and

$$V_{\widetilde{\eta}_1(t)}(Z|t) := \frac{\partial}{\partial \eta_1}\Big(\frac{s^{(1)}(\eta_1, t)}{s^{(0)}(\eta_1, t)}\Big)\Big|_{\eta_1 = \widetilde{\eta}_1(t)}$$

gives a consistent estimate of the conditional variance of Z given $(T = t, D = 1)$ under the model (3.14).

We can apply the Taylor theorem to expand $\frac{s^{(1)}(\eta_1^*, t)}{s^{(0)}(\eta_1^*, t)}$ around $\widetilde{\eta}_1(t)$ for a fixed t, where η_1^* is the solution of (3.19). This yields

$$\frac{s^{(1)}(t)}{s^{(0)}(t)} - \frac{s^{(1)}(\eta_1^*, t)}{s^{(0)}(\eta_1^*, t)} \ \approx\ [\eta_1^* - \widetilde{\eta}_1(t)] V_{\widetilde{\eta}_1(t)}(Z|t). \tag{3.20}$$

Also, under the random censorship assumption

$$\begin{aligned} s^{(0)}(t) &= E[Y(t)\lambda_{10}(t)\exp(\widetilde{\eta}_1(t)Z)] \\ &= E[Y(t)]\cdot E[\lambda_{10}(t)\exp(\widetilde{\eta}_1(t)Z)] \\ &= E[I(T\geq t)]\cdot E[\lambda_1^{(0)}(t;Z)] \\ &= E[I(\widetilde{T}\geq t, C\geq t)]\cdot E[\lambda_1^{(0)}(t|Z)] \\ &= S(t)C(t)\lambda_1^{(0)}(t). \end{aligned} \tag{3.21}$$

Indeed, if we denote by $h(z|T\geq t)=P(Z=z|T\geq t)$, the conditional density of Z given $T\geq t$, we have

$$\begin{aligned} E[\lambda_1^{(0)}(t|Z)] &= E[\lambda_1^{(0)}(t|Z)|T\geq t] \\ &= \int_{-\infty}^{+\infty}\lambda_1^{(0)}(t;u)\cdot h(u|T\geq t)du \end{aligned}$$

and, further

$$\begin{aligned} &= \int_{-\infty}^{+\infty}\lim_{\Delta t\to 0}\frac{P(t\leq T\leq t+\Delta t, D=1|T\geq t, Z=u)}{\Delta t}P(Z=u|T\geq t)du \\ &= \int_{-\infty}^{+\infty}\lim_{\Delta t\to 0}\frac{1}{\Delta t}\frac{P(t\leq T\leq t+\Delta t, D=1|T\geq t, Z=u)}{P(T\geq t, Z=u)}\frac{P(Z=u, T\geq t)}{P(T\geq t)}du \\ &= \int_{-\infty}^{+\infty}\lim_{\Delta t\to 0}\frac{1}{\Delta t}P(t\leq T\leq t+\Delta t, D=1, Z=u|T\geq t)du \\ &= \lim_{\Delta t\to 0}\frac{P(t\leq T\leq t+\Delta t, D=1|T\geq t)}{\Delta t} \\ &= \lambda_1^{(0)}(t). \end{aligned}$$

Combination of (3.19), (3.20) and (3.21) yields

$$\int_0^{\infty}\left[\eta_1^*-\widetilde{\eta}_1(t)\right]V_{\widetilde{\eta}_1(t)}(Z|t)S(t)C(t)\lambda_1^{(0)}(t)dt \;\approx\; 0$$

which provides an approximation of β_1^*, namely

$$\eta_1^*\approx\frac{\int_0^{\infty}S(t)C(t)V_{\widetilde{\eta}_1(t)}(Z|t)\lambda_1^{(0)}(t)\widetilde{\eta}_1(t)dt}{\int_0^{\infty}S(t)C(t)V_{\widetilde{\eta}_1(t)}(Z|t)\lambda_1^{(0)}(t)dt}. \tag{3.22}$$

Appendix B: Profile likelihood

Consider two causes of failure, with cause-specific hazards

$$\begin{aligned} \lambda_1(t \,|\, \mathbf{Z}) &= \lambda_0(t) \exp(\boldsymbol{\eta}_1^\top \mathbf{Z}), \\ \lambda_2(t \,|\, \mathbf{Z}) &= \lambda_0(t) \exp(\boldsymbol{\xi}^\top \mathbf{B}(t) + \boldsymbol{\eta}_2^\top \mathbf{Z}) \ , \end{aligned} \tag{3.23}$$

with $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ denoting the effects of the covariates on the cause-specific hazards of cause 1 and 2, respectively, $\boldsymbol{\xi}$ parameterizing the (time-varying) ratio between the baseline cause-specific hazards of cause 1 with respect to that of cause 2, and $\mathbf{B}(t)$ as before a vector of given time functions. Let $\boldsymbol{\theta} = (\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top, \boldsymbol{\xi}^\top)^\top$. Recall $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_u$ as the set of subjects with failure of cause 1, 2, and of unknown cause, respectively, $\mathcal{D}_{knw} = \mathcal{D}_1 \cup \mathcal{D}_2$ as the set of subjects with known cause of failure, and $\mathcal{D} = \mathcal{D}_{knw} \cup \mathcal{D}_u$ as the set of subjects with failure, irrespective of the cause. Further define

$$w_i^{(1)} = \exp(\boldsymbol{\eta}_1^\top \mathbf{Z}_i) \ , \ w_i^{(2)}(t) = \exp(\boldsymbol{\xi}^\top \mathbf{B}(t) + \boldsymbol{\eta}_2^\top \mathbf{Z}_i) \ , \ w_i^{(u)}(t) = w_i^{(1)} + w_i^{(2)}(t) \ ,$$

each of these variables depending on $\mathbf{Z}_i$ and $\boldsymbol{\theta}$. Rearranging the terms in the likelihood (3.6), we see that

$$L(\boldsymbol{\theta}) = \prod_{i \in \mathcal{D}_1} \lambda_1(t_i|\mathbf{Z}_i) \cdot \prod_{i \in \mathcal{D}_2} \lambda_2(t_i|\mathbf{Z}_i) \cdot \prod_{i \in \mathcal{D}_u} \lambda_\bullet(t_i|\mathbf{Z}_i) \cdot \prod_{i=1}^{n} \exp\Big(-\Lambda_\bullet(t_i|\mathbf{Z}_i)\Big),$$

which for the present parameterization leads to

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i \in \mathcal{D}_1} w_i^{(1)} \cdot \prod_{i \in \mathcal{D}_2} w_i^{(2)}(t_i) \cdot \prod_{i \in \mathcal{D}_u} w_i^{(u)}(t_i) \cdot \\ &\cdot \prod_{i \in D} d\Lambda_0(t_i) \cdot \prod_{i=1}^{n} \exp\Big(-\int_0^{t_i} w_i^{(u)}(s) d\Lambda_0(s)\Big) \ . \end{aligned} \tag{3.24}$$

Now a profile likelihood argument well known in ordinary survival analysis (see e.g. Klein and Moeschberger (2003)) can be used as follows: for fixed $\boldsymbol{\theta}$, the maximizer of the log-likelihood with respect to the baseline hazard is a step function with increment

$$\hat{\lambda}_0(t_i) = \frac{1}{\sum_{j \in R(t_i)} w_j^{(u)}(t_i)} \ . \tag{3.25}$$

Replacing this maximizer (3.25) back into the likelihood (3.24) yields, up to a constant,

$$L^*(\boldsymbol{\theta}) = \frac{\prod_{i\in\mathcal{D}_1} w_i^{(1)} \prod_{i\in\mathcal{D}_2} w_i^{(2)}(t_i) \prod_{i\in\mathcal{D}_u} w_i^{(u)}(t_i)}{\prod_{i\in\mathcal{D}} \sum_{j\in R(t_i)} w_j^{(u)}(t_i)},$$

i.e (3.24). Define

$$\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i) = \frac{\sum_{j\in R(t_i)} \mathbf{Z}_j w_j^{(1)}}{\sum_{j\in R(t_i)} w_j^{(u)}(t_i)},\ \overline{\mathbf{Z}}_{\boldsymbol{\eta}_2}(t_i) = \frac{\sum_{j\in R(t_i)} \mathbf{Z}_j w_j^{(2)}(t_i)}{\sum_{j\in R(t_i)} w_j^{(u)}(t_i)},$$
$$\overline{\mathbf{Z}}_{\boldsymbol{\xi}}(t_i) = \frac{\sum_{j\in R(t_i)} \mathbf{B}(t_i) w_j^{(2)}(t_i)}{\sum_{j\in R(t_i)} w_j^{(u)}(t_i)}.$$

Similar to Appendix A, $\overline{\mathbf{Z}}_{\eta_1}(t) + \overline{\mathbf{Z}}_{\eta_2}(t)$ gives a consistent estimate of the conditional expectation of Z given $T = t$ under the model (3.23).

The score functions are given by

$$\begin{aligned}
\frac{\partial \log L^*(\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_k} &= \sum_{i\in\mathcal{D}_k} \mathbf{Z}_i + \sum_{i\in\mathcal{D}_u} \mathbf{Z}_i \pi_{ki} - \sum_{i\in\mathcal{D}} \overline{\mathbf{Z}}_{\eta_k}(t_i),\ k = 1, 2, \\
\frac{\partial \log L^*(\boldsymbol{\theta})}{\partial \boldsymbol{\xi}} &= \sum_{i\in\mathcal{D}_2} \mathbf{B}(t_i) + \sum_{i\in\mathcal{D}_u} \mathbf{B}(t_i) \pi_{2i} - \sum_{i\in\mathcal{D}} \overline{\mathbf{Z}}_{\xi}(t_i),
\end{aligned}$$

with $\pi_{ki} = \pi_k(t_i|\mathbf{Z}_i)$. Define

$$\begin{aligned}
\mathbf{V}_{\boldsymbol{\eta}_1}(t_i) &= \frac{\sum_{j\in R(t_i)} \mathbf{Z}_j \mathbf{Z}_j^\top w_j^{(1)}}{\sum_{j\in R(t_i)} w_j^{(u)}(t_i)} - \overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i) \overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)^\top, \\
\mathbf{V}_{\boldsymbol{\eta}_2}(t_i) &= \frac{\sum_{j\in R(t_i)} \mathbf{Z}_j \mathbf{Z}_j^\top w_j^{(2)}(t_i)}{\sum_{j\in R(t_i)} w_j^{(u)}(t_i)} - \overline{\mathbf{Z}}_{\boldsymbol{\eta}_2}(t_i) \overline{\mathbf{Z}}_{\boldsymbol{\eta}_2}(t_i)^\top, \\
\mathbf{V}_{\boldsymbol{\xi}}(t_i) &= \frac{\sum_{j\in R(t_i)} \mathbf{B}(t_i) \mathbf{Z}_j^\top w_j^{(2)}(t_i)}{\sum_{j\in R(t_i)} w_j^{(u)}(t_i)} - \overline{\mathbf{Z}}_{\boldsymbol{\xi}}(t_i) \overline{\mathbf{Z}}_{\boldsymbol{\xi}}(t_i)^\top, \\
\mathbf{V}_B(t_i) &= \frac{\sum_{j\in R(t_i)} \mathbf{B}(t_i) \mathbf{B}(t_i)^\top w_j^{(2)}(t_i)}{\sum_{j\in R(t_i)} w_j^{(u)}(t_i)} - \overline{\mathbf{Z}}_{\boldsymbol{\xi}}(t_i) \overline{\mathbf{Z}}_{\boldsymbol{\xi}}(t_i)^\top.
\end{aligned}$$

Similar to the Appendix A,

$$\mathbf{V}_{\eta_1}(t) + \mathbf{V}_{\eta_2}(t) - \big(\overline{Z}_{\eta_1}(t) + \overline{Z}_{\eta_2}(t)\big)\big(\overline{Z}_{\eta_1}(t) + \overline{Z}_{\eta_2}(t)\big)^\top$$

gives a consistent estimate of the conditional variance of $\mathbf{Z}$ given $T = t$ under the model (3.23).

The information matrix is given by

$$\mathbf{I}(\boldsymbol{\theta}) = -\frac{\partial^2 \log L^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \mathbf{I}_{knw}(\boldsymbol{\theta}) + \mathbf{I}_{unk}(\boldsymbol{\theta}),$$

where

$$\mathbf{I}_{knw}(\boldsymbol{\theta}) = \sum_{i \in \mathcal{D}_{knw}} \begin{pmatrix} \mathbf{V}_{\boldsymbol{\eta}_1}(t_i) & -\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)\overline{\mathbf{Z}}_{\boldsymbol{\eta}_2}(t_i)^\top & -\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)\overline{\mathbf{Z}}_{\boldsymbol{\xi}}(t_i)^\top \\ -\overline{\mathbf{Z}}_{\boldsymbol{\eta}_2}(t_i)\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)^\top & \mathbf{V}_{\boldsymbol{\eta}_2}(t_i) & \mathbf{V}_{\boldsymbol{\xi}}(t_i) \\ -\overline{\mathbf{Z}}_{\xi}(t_i)\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)^\top & \mathbf{V}_{\boldsymbol{\xi}}(t_i)^\top & \mathbf{V}_B(t_i) \end{pmatrix}$$

and

$$\begin{aligned} \mathbf{I}_{unk}(\boldsymbol{\theta}) = & \sum_{i \in \mathcal{D}_u} \begin{pmatrix} \mathbf{V}_{\boldsymbol{\eta}_1}(t_i) & -\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)\overline{\mathbf{Z}}_{\boldsymbol{\eta}_2}(t_i)^\top & -\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)\overline{\mathbf{Z}}_{\boldsymbol{\xi}}(t_i)^\top \\ -\overline{\mathbf{Z}}_{\boldsymbol{\eta}_2}(t_i)\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)^\top & \mathbf{V}_{\boldsymbol{\eta}_2}(t_i) & \mathbf{V}_{\boldsymbol{\xi}}(t_i) \\ -\overline{\mathbf{Z}}_{\boldsymbol{\xi}}(t_i)\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)^\top & \mathbf{V}_{\boldsymbol{\xi}}(t_i)^\top & \mathbf{V}_B(t_i) \end{pmatrix} \\ & - \sum_{i \in \mathcal{D}_u} \pi_{1i}\pi_{2i} \begin{pmatrix} \mathbf{Z}_i\mathbf{Z}_i^\top & -\mathbf{Z}_i\mathbf{Z}_i^\top & -\mathbf{Z}_i\mathbf{B}(t_i)^\top \\ -\mathbf{Z}_i\mathbf{Z}_i^\top & \mathbf{Z}_i\mathbf{Z}_i^\top & \mathbf{Z}_i\mathbf{B}(t_i)^\top \\ -\mathbf{B}(t_i)\mathbf{Z}_i^\top & \mathbf{B}(t_i)\mathbf{Z}_i^\top & \mathbf{B}(t_i)\mathbf{B}(t_i)^\top \end{pmatrix}. \end{aligned}$$

A rearrangement yields

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbf{I}_{tot}(\boldsymbol{\theta}) - \mathbf{I}_{miss}(\boldsymbol{\theta}),$$

with

$$\mathbf{I}_{tot}^{(\boldsymbol{\theta})} = \sum_{i \in \mathcal{D}} \begin{pmatrix} \mathbf{V}_{\boldsymbol{\eta}_1}(t_i) & -\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)\overline{\mathbf{Z}}_{\boldsymbol{\eta}_2}(t_i)^\top & -\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)\overline{\mathbf{Z}}_{\boldsymbol{\xi}}(t_i)^\top \\ -\overline{\mathbf{Z}}_{\boldsymbol{\eta}_2}(t_i)\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)^\top & \mathbf{V}_{\boldsymbol{\eta}_2}(t_i) & \mathbf{V}_{\boldsymbol{\xi}}(t_i) \\ -\overline{\mathbf{Z}}_{\xi}(t_i)\overline{\mathbf{Z}}_{\boldsymbol{\eta}_1}(t_i)^\top & \mathbf{V}_{\boldsymbol{\xi}}(t_i)^\top & \mathbf{V}_B(t_i) \end{pmatrix}$$

and

$$\mathbf{I}_{miss}(\boldsymbol{\theta}) = \sum_{i \in \mathcal{D}_u} \pi_{1i}\pi_{2i} \begin{pmatrix} \mathbf{Z}_i\mathbf{Z}_i^\top & -\mathbf{Z}_i\mathbf{Z}_i^\top & -\mathbf{Z}_i\mathbf{B}(t_i)^\top \\ -\mathbf{Z}_i\mathbf{Z}_i^\top & \mathbf{Z}_i\mathbf{Z}_i^\top & \mathbf{Z}_i\mathbf{B}(t_i)^\top \\ -\mathbf{B}(t_i)\mathbf{Z}_i^\top & \mathbf{B}(t_i)\mathbf{Z}_i^\top & \mathbf{B}(t_i)\mathbf{B}(t_i)^\top \end{pmatrix}.$$

Similar to the work of Louis (1982) in the context of Fisher information in case of missing values and estimation using the EM-algorithm, $\mathbf{I}_{tot}(\boldsymbol{\theta})$ and $\mathbf{I}_{miss}(\boldsymbol{\theta})$ can be interpreted as the information in the case of complete data and the loss of information due to the missing data, respectively.

4

Dynamic prediction by landmarking in competing risks

Abstract

We propose an extension of the landmark model of van Houwelingen (2007) as a new approach to the problem of dynamic prediction in competing risks with time-dependent covariates. We fix a set of landmark time points t_{LM} within the follow-up interval. For each of these landmark time points t_{LM} we create a landmark data set by selecting individuals at risk at t_{LM}; the value of the time-dependent covariate in each landmark data set is fixed at the value at t_{LM}. We assume Cox proportional hazard models for the cause-specific hazards and we consider smoothing the (possibly) time-dependent effect of the covariate for the different landmark data sets. Fitting this model is possible within the standard statistical software. We illustrate the features of the landmark modeling on a real data set on bone marrow transplantation.

4.1 Introduction

Prediction is of crucial importance in clinical practice. Prediction models are used as a basis for treatment decisions and to communicate prognosis to patients.

Numerous prediction models have been proposed in the statistical and medical literature for a wide variety of diseases (see, for example, Wilson et al. (1998); Ravdin et al. (2001)). These are designed to render prediction probabilities of the event of interest over time, where the starting point is some pre-defined clinically important point in the event history of the patient, such as birth (age at onset), diagnosis, or start of primary treatment. The vast majority of these models are only used (or can only be used) from that starting point, but in clinical management prediction is equally relevant at later times in the follow-up. Between the starting point and the time of prediction, information on clinical events that may influence the prognosis of the endpoint of interest has become available. In statistical models, this type of information is incorporated through time-dependent covariates. The predictions that were obtained at the start need to be updated to include the time-dependent information. This prediction from later points in time is called dynamic prediction. Dynamic prediction is not new (Christensen et al., 1983; Madsen et al., 1983; Klein et al., 1993; Arjas and Eerola, 1993), but is the topic of active recent research (van Houwelingen, 2007; van Houwelingen and Putter, 2008; Proust-Lima and Taylor, 2009) (see, also, Cortese, G., Gerds, T. A. and Andersen, P. K., Comparison of prediction models for competing risks with time-dependent covariates, Research report 2011, Department of Biostatistics, University of Copenhagen). One way of approaching such dynamic predictions is through multi-state models (Putter et al., 2007), where the clinically relevant events define states, and transitions from one state to another are defined and modeled. Under the Markov assumption it is possible to obtain dynamic prediction probabilities explicitly (Aalen and Johansen, 1978). Under more realistic assumptions, such explicit calculations are not possible, but simulation may be used to approximate them (Dabrowska, 1995; Fiocco et al., 2008).

Given the often time-consuming and indirect way of obtaining dynamic predictions, statistical researchers have proposed new, more direct ways of getting these probabilities. One of these new approaches is landmarking (Anderson et al., 1983). The method was proposed for dynamic prediction by van Houwelingen (2007), and it is useful in particular in the presence of either time-dependent covariates or covariates with time-varying effects. Van Houwelingen and Putter (van Houwelingen and Putter, 2008) have proposed landmarking as an alternative for multi-state modeling if the interest is in obtaining dynamic prediction probabilities for a single endpoint of interest. They compared the multi-state and the landmarking approaches to prediction, discussing their advantages and disadvantages. Increasing interest has been shown in the clinical applications of this method (Zamboni et al., 2010; McCarthy and Hahn, 2011; Beyersmann et al., 2011) or extensions of it (Parast et al., 2011; Gran et al., 2010). For a comprehensive overview of the existing methods for dynamic prediction and of landmarking

in particular, refer to the recent book of van Houwelingen and Putter (2012).

In this paper we consider the situation of multiple competing endpoints, and our aim is to extend the landmarking approach to competing risks. The same problem was addressed by Cortese and Andersen (2010) and by van Houwelingen and Putter (2012). As one of several strategies of dealing with time-dependent covariates in competing risks Cortese and Andersen (Cortese and Andersen, 2010) considered landmarking for dynamic prediction in competing risks. They applied landmarking at a small number of pre-defined relevant time points; this approach is limited to dynamic prediction at these landmark time points. They made no attempt to construct comprehensive models that would enable dynamic prediction of the cumulative incidences of the different causes of failure at time points other than these landmark time points. Van Houwelingen and Putter (van Houwelingen and Putter, 2012) did obtain supermodels for the cause-specific hazards but did not use these to obtain dynamic prediction of the cumulative incidences of the causes of interest and concentrated on the combined endpoint.

The paper is organized as follows: in Section 2 the data are introduced and results are shown of an exploratory data analysis based on a traditional cause-specific hazards approach. In Section 3 the landmark approach to competing risks is introduced and applied to the data. Section 4 concludes the paper with a discussion.

4.2 Exploratory data analysis

In this paper we develop a dynamic prediction model for competing risks aimed at predicting events of interest at later time points based on the complete history of the patient up to relevant, intermediate time points. Our data consists of 5582 chronic myelogenous leukaemia (CML) patients, registered at the European Group for Blood and Marrow Transplantation (EBMT) who received allogeneic stem cell transplantation (SCT) between 1997 and 2003. Events recorded during the follow-up of these patients were: acute graft versus host disease (aGvHD), relapse (Rel) and non-relapse mortality (NRM). Development of aGvHD represents a major complication to transplants, especially to CML patients, being associated with considerable morbidity and increased mortality, but also with decreased probability of leukemia recurrence. The time of onset of aGvHD, which by definition occurs within the first 100 days post-transplant, was recorded in our data, as well as the maximum grade of aGvHD ever reached; here, only grade 2 or higher were considered as aGvHD event. In reality, the highest grade of aGvHD may be achieved (shortly) after the onset of aGvHD, but since the grade at onset is often unknown and otherwise too difficult to trace back, this issue is conveniently ignored in our analysis; we act as if the highest grade of aGvHD was actually attained at the onset of aGvHD. Although aGvHD occurs in the

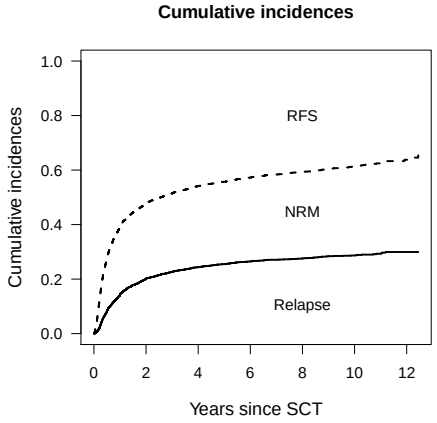
first 100 days after SCT and even though it may be resolved after some time we assume that it may have a permanent, possibly time-varying effect on relapse and NRM. We shall distinguish between *low* grade aGvHD, corresponding to grade 2, and *high* grade aGvHD, corresponding to grade 3 or higher. This distinction is clinically relevant, since it is known that high grade aGvHD's have a larger and more immediate impact on NRM. Prognostic information at time of SCT are: year of transplantation and EBMT risk score, a prognostic index based on available marker information at baseline, known to be predictive of both relapse and NRM. The latter covariate is divided into three risk groups, denoted as low risk, medium risk and high risk. The frequencies of the values of these covariates are shown in Table 4.1. The clinical purpose of our approach is to obtain a dynamic prognostic model of relapse free survival and cumulative incidences of Rel and NRM, given the history of aGvHD and the baseline covariates.

Table 4.1: Prognostic factors for all patients.

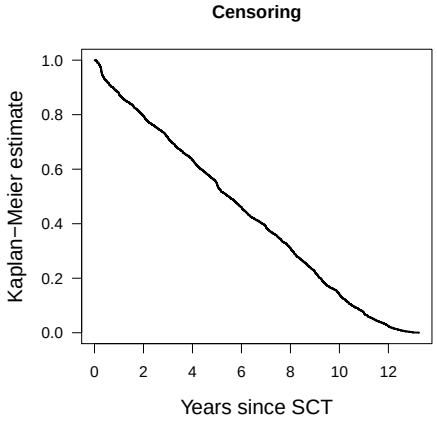
Prognostic factor	Category	n (%)
Risk score	Low	2361 (42%)
	Medium	2663 (48%)
	High	558 (10%)
Year of SCT	1997	773 (14%)
	1998	956 (17%)
	1999	1004 (18%)
	2000	956 (17%)
	2001	669 (12%)
	2002	658 (12%)
	2003	566 (10%)

Figure 4.1a shows a stacked plot of the estimated cumulative incidences of time to relapse and time to non-relapse mortality; it is clear that the situation of the patients is quite stable after 5 years. The graph of the censoring distribution shows that the median follow-up is reached at about 5.5 years. Figure 4.1b shows a plot of the censoring distribution.

To get an impression when low and high grade aGvHD occur in our data, a stacked plot of the estimated cumulative incidence functions of time to low and high grade aGvHD, respectively, is shown in Figure 4.2. The time scale goes to four months after SCT due to the fact that by definition this intermediate event occurs quite early in the follow-up interval (within 100 days after SCT).



(a)



(b)

Figure 4.1: Cumulative incidences of competing events relapse and NRM (a), the censoring distribution (b).

To explore the potentially time-varying effects of the time-dependent factors low and high grade aGvHD, let

$$Z_l(t) = \mathbf{1}\{\text{occurrence of low grade aGvHD before time } t\},$$
$$Z_h(t) = \mathbf{1}\{\text{occurrence of high grade aGvHD before time } t\}$$

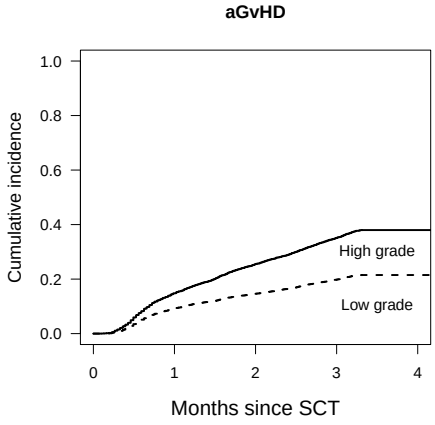


Figure 4.2: Cumulative incidence functions of time until aGvHD.

be two binary time-dependent covariates which refer to having experienced low or high grade aGvHD, respectively, before time t. These covariates are equal to 0 for all individuals at the time origin, and their values might change to 1 at the time of occurrence of low or high grade aGvHD, respectively. In that case, they remain time-constant equal to 1 over the remaining follow-up interval. Note that $Z_l(t)$ and $Z_h(t)$ are mutually exclusive, in the sense that they cannot take simultaneously the value 1. An exploratory Cox regression analysis revealed that the effects of $Z_l(t)$ and $Z_h(t)$ on the cause-specific hazards of relapse and NRM varied over time, especially for NRM. To distinguish between short-term and long-term effects in our analysis, we define subsequently the time-dependent binary covariates:

$$\begin{aligned} Z_l^{VR}(t) &= \mathbf{1}\{\text{occurrence of low grade aGvHD within one month before } t\}, \\ Z_l^{R}(t) &= \mathbf{1}\{\text{occurrence of low grade aGvHD between 6 and one months before } t\}, \\ Z_l^{P}(t) &= \mathbf{1}\{\text{occurrence of low grade aGvHD more than 6 months before } t\}. \end{aligned}$$

The superscripts are abbreviations of "very recent", "recent" and "past", respectively. Similarly, define $Z_h^{VR}(t)$, $Z_h^{R}(t)$ and $Z_h^{P}(t)$ for high grade aGvHD. By definition, we have

$$Z_k(t) = Z_k^{VR}(t) + Z_k^{R}(t) + Z_k^{P}(t),\ k \in \{l,\ h\}.$$

We fitted a Cox model on the cause-specific hazards of relapse and non-relapse

mortality with these time-dependent covariates and the two baseline covariates. Year of transplantation (centered around 2000) is included as a continuous covariate. A backward selection procedure based on the likelihood ratio test was used to test whether $Z_l^{VR}(t)$, $Z_l^R(t)$ and $Z_l^P(t)$ could be replaced by $Z_l(t)$ (and similarly for $Z_h(t)$) both for relapse and NRM. This procedure showed no significant difference among the effects of $Z_l^{VR}(t)$, $Z_l^R(t)$ and $Z_l^P(t)$ on the cause-specific hazard of relapse, therefore resulting in inclusion of only $Z_l(t)$ in the model for the cause-specific hazard of relapse. The results in Table 4.2 confirm some known facts about CML patients. Occurrence of aGvHD has a highly significant, protective effect on the risk of relapse; for high grade aGvHD, one month after its occurrence, this protective effect even increases followed by a non-significant decrease after 6 months. This finding could be real or an artefact caused by an immortal time bias due to the fact that the time of onset of aGvHD may be before the time at which the highest grade is reached. Another contributing factor could be the fact that in general patients with low grade aGvHD are not treated, while patients with high grade aGvHD are treated with immunosuppression. The risk of non-relapse mortality is increased by the occurrence of aGvHD; the highest detrimental effect is seen immediately after the occurrence of high grade aGvHD. Later year of transplantation seems to be associated with higher relapse and lower NRM rates, both effects being close to significance level of 0.05. Higher risk score increases the risk of both causes of failure.

Table 4.2: Estimated parameters for time-dependent effects of aGvHD.

Prognostic factor	Relapse		NRM	
	$\widehat{\beta}$	SE($\widehat{\beta}$)	$\widehat{\beta}$	SE($\widehat{\beta}$)
Year	0.029	0.015	-0.027	0.013
Risk score				
Low risk				
Medium risk	0.297	0.058	0.372	0.055
High risk	1.038	0.085	0.895	0.077
Low grade aGvHD	-0.428	0.072		
Very recent			1.028	0.135
Recent			0.550	0.100
Past			0.649	0.099
High grade aGvHD				
Very recent	-0.365	0.359	2.794	0.083
Recent	-1.153	0.241	1.922	0.080
Past	-0.848	0.167	1.208	0.114

Checking the validity of the proportionality assumption for the time-fixed covariates revealed that the effect of EBMT risk score varies considerably over time. Stratification by risk score would improve the modeling of cause-specific hazards and would be preferable to obtain predictions, but this is not pursued here since the emphasis of this paper is on the use of landmarking for prediction purposes.

In a traditional approach, prediction of relapse-free survival and cumulative incidences of relapse and of non-relapse mortality would require a joint model comprising a model for event time (time to relapse and time to non-relapse mortality) which incorporates time-dependent and baseline covariates, and a model for the time-dependent covariates incorporating the baseline covariates: Table 4.3 shows the effects of the baseline covariates estimated from a Cox proportional hazards model on the cause-specific hazards of low and high grade aGvHD, respectively. In this particular model, being in the highest risk score increases the rate of low and high grade aGvHD with more than 50%, while year of transplantation is slightly protective. Interpretation of the results in Table 4.3 in terms of cumulative incidences is not straightforward. The presence of lagged covariates, $Z_k^{VR}(t)$, $Z_k^R(t)$ and $Z_k^P(t)$, $k \in \{l,\ h\}$, in this comprehensive joint model makes prediction far more difficult, because the multi-state model defined by the clinical events low and high grade aGvHD, relapse and NRM would be non-Markovian. The motivation for this last statement is that after the occurrence of an aGvHD event, the rate of relapse and NRM depend on when (in the history) the aGvHD occurred.

Table 4.3: Cause-specific hazard ratios (HR) of the prognostic factors and the corresponding 95% confidence intervals (95% CI) for the competing events low and high grade aGvHD.

Prognostic factor	Category	Low grade aGvHD HR (95% CI)	High grade aGvHD HR (95% CI)
Year		0.93 (0.90-0.96)	0.92 (0.88-0.95)
Risk score	Low risk	1	1
	Medium risk	1.16 (1.03-1.31)	1.47 (1.28-1.71)
	High risk	1.53 (1.26-1.86)	2.04 (1.65-2.53)

4.3 Dynamic prediction based on the landmark model

In this section we extend the landmark approach of van Houwelingen (2007) to competing risks. There are recent advancements in this direction; Parast et al. (2011) give non-parametric estimators to prediction probabilities for a fixed landmark time point in the context of semi-competing risks and Cortese and Andersen (2010) use landmarking to estimate the effects of time-dependent covariates on the cause-specific hazards of competing events, their principal aim being estimation rather than dynamic prediction. For this purpose, these authors only fit models at the landmark time points and do not go to supermodels as proposed in van Houwelingen (2007).

Our goal is to develop a dynamic prediction model for relapse-free survival and for the cumulative incidence of relapse and non-relapse mortality based on the time-dependent covariates $Z_k^{VR}(t)$, $Z_k^R(t)$ and $Z_k^P(t)$, $k \in \{l,\ h\}$, and the time-fixed covariates year of transplantation and risk score. Since the intermediate clinical events occur within the first year (see Figure 4.2), the target period for initiating dynamic prediction could be anywhere in the first year. Our aim is to predict 5 years ahead from some time s within the first year post-transplant. In other words, we want to obtain dynamic models for relapse-free survival and for the cumulative incidences of relapse and non-relapse mortality for a window with a fixed width of 5 years from anywhere in the first year post-transplant. Section 4.3.1 gives the general theory, while Section 4.3.2 contains an application to the EBMT data.

4.3.1 Landmarking and competing risks

Suppose that data are available from n individuals each of whom can experience one of J types of failure, which we term $1, \ldots, J$, respectively, or can be subject to a noninformative censoring. Let $\tilde{T}$ denote the time of failure, C the censoring time, and D the cause of failure. Let $Z(\cdot)$ denote a p-vector of covariates, which could be measured at baseline or be time-dependent. The observed data for individual i is $(T_i, \Delta_i, Z_i(\cdot))$, where $T_i = \min(\tilde{T}_i, C_i)$ is the earliest of failure and censoring time, $\Delta_i = \mathbf{1}\{\tilde{T}_i < C_i\} \cdot D_i$ is the cause of failure in case of failure and 0 in case of censoring and where $Z_i(\cdot)$ denotes the covariates of individual i observed until T_i, for $i = 1, \ldots, n$. The usual requirement of conditional independence of $(\tilde{T}, D)$ and C, given $Z(\cdot)$, is assumed to be true here as well. Data from different individuals are supposed to be independent.

We are interested in the dynamic prediction of survival and of the cumulative incidences of cause j, $j = 1, \ldots, J$. More precise, our aim is to estimate the survival probability and the cumulative incidences of cause j, $j = 1, \ldots, J$ at

time $s+w$, respectively, conditional on surviving event-free at a certain time s and on $Z(s)$, that is

$$\begin{aligned} S_{\mathrm{LM}}(s+w|\ Z(s),\ s) &= P(T > s+w|\ T > s,\ Z(s)) \\ F_{j,\mathrm{LM}}(s+w|\ Z(s),\ s) &= P(T \leq s+w, D=j|\ T > s,\ Z(s)),\ j=1,\ldots,J, \end{aligned} \tag{4.1}$$

respectively.

The landmark approach to dynamic prediction consists of two steps: the construction of a landmark data set and the development of models for the competing endpoints based on that landmark data set. The first step requires the choice of a set of landmark time points t_{LM} within an interval $[s_0; s_1]$ and the length of the prediction interval, w. For each landmark time point t_{LM}, a new data set is built as a subset of the initial data set, referred as the t_{LM}-landmark data set. This t_{LM}-landmark data set comprises only the subjects at risk at t_{LM}, for which the events occurring after the horizon time as defined by $t_{\mathrm{hor}} = t_{\mathrm{LM}} + w$ are administratively censored at t_{hor}. The time-dependent covariates are fixed at their current value at t_{LM}, that is $Z(t_{\mathrm{LM}})$. The result of this data construction procedure is a large data set, obtained by stacking the individual t_{LM}-landmark data sets. In the following, we shall replace t_{LM} by s for the sake of convenience.

The second step of the landmark approach consists of obtaining models for each of the cause-specific hazards within a prediction interval $[s; s+w]$ and combining these into a supermodel which dictates prediction in any time period of length w starting anywhere in $[s_0; s_1]$.

To this goal, we define models for the cause-specific hazards, conditional on survival beyond time s and covariates $Z(s)$, that is for

$$\lambda_j(t|\ Z(s),\ s) = \lim_{\delta \to 0} \frac{1}{\delta} \cdot P(t \leq T \leq t+\delta, D=j|\ T \geq t,\ Z(s)),\ j=1,\ldots,J,$$

for $s \leq t \leq s+w$, based on the s-landmark data set, and we use these models to obtain estimates of the quantities in (4.1) through the relations

$$\begin{aligned} S(s+w|\ Z(s),\ s) &= \exp\Big(-\int_s^{s+w} \lambda_j(u|\ Z(s),\ s)du\Big), \\ F_j(s+w|\ Z(s),\ s) &= \int_s^{s+w} \lambda_j(u|\ Z(s),\ s) \cdot S(u|\ Z(s),\ s)du. \end{aligned} \tag{4.2}$$

The simplest model we could consider for the conditional cause-specific hazards on each s-landmark data set would be

$$\lambda_j(t|\ Z(s),\ s) = \lambda_{j0}(t|s)\exp(\beta_j(s)Z(s)),\ j=1,\ldots,J, \tag{4.3}$$

where $\lambda_{j0}(t|s)$, $j = 1, \ldots, J$, are unspecified baseline hazards, and $\beta_j(s)$, $j = 1, \ldots, J$, are unknown regression coefficients that are unique for each s-landmark data set. Note that (4.3) specifies a time-fixed Cox model, where the time-dependent covariates are taken at their values at s and hence β_j may depend on s but not on t. Fitting this model for each landmark point separately would lead to the estimation of the desired dynamic prediction probabilities in (4.1), but would ignore the overlap of subjects in the landmark data sets. We can expect that the coefficients $\beta_j(s)$ depend on s in a smooth way. We can bring more structure into the analysis by modeling the regression parameters $\beta_j(s)$ as functions of s:

$$\beta_j(s) = f(s; \beta^{(j)}),\ j = 1, \ldots, J, \tag{4.4}$$

where $\beta^{(j)} = (\beta_{j1}, \ldots, \beta_{jk})$ is a p_j-vector of regression parameters and $f(\cdot)$ is a parametric function of s. This choice could include, for instance, polynomial or splines. We gather all the regression coefficients of the models (4.3)-(4.4) into β_{LM}. Fitting this model with the Breslow partial likelihood for those tied observations is equivalent to maximizing the pseudo-partial log-likelihood

$$\text{ipl}(\beta_{\text{LM}}) = \sum_{j=1}^{J} \sum_{i} d_{ij} \left(\sum_{s:\ s<t_i\leq s+w} \left[Z_i(s)^\top \beta_j(s) - \ln \sum_{t_k:\ t_i \leq t_k \leq s+w} e^{Z_k(s)\beta_j(s)} \right] \right), \tag{4.5}$$

where $d_{ij} = 1\{D_i = j\}$ is the event indicator of patient i. It can be fitted to the data by means of standard software, provided that the software allows for delayed entry at s, using the stacked data set, containing all the landmark data sets with stratification on the landmark. Repeated observations of the same subject automatically lead to the presence of many ties. Models (4.3)-(4.4) can be used to inspect whether the regression coefficients depend on the landmark. However, such a fit cannot be used directly to test the statistical significance of the components $\beta^{(j)}$, $j = 1, \ldots, J$, since the data of the same patient are used repeatedly in the different landmark strata. The correct standard errors can by obtained by taking into account the "clustering" of the data and using the sandwich estimators proposed in Lin and Wei (1989).

After fitting the model, Breslow type estimators of the conditional baseline hazards are available, given by

$$\widehat{\lambda}_{j0}(t_i|s) = \frac{1}{\sum_{t_k:\ s\leq t_i \leq t_k < s+w} \exp(Z_k(s)^\top \widehat{\beta}_j(s))},\ j = 1, \ldots, J.$$

The estimated baseline hazards $\widehat{\lambda}_{j0}(t|s)$ can be expected to vary continuously

with s through $\beta_j(s)$ via (4.4). We can model this dependence directly through

$$\lambda_{j0}(t|s) = \lambda_{j0}(t)\exp(\gamma_j(s)),\ j = 1, \ldots, J, \tag{4.6}$$

where $\gamma_j(s)$, $j = 1, \ldots, J$, are some parametric functions of s with the restriction $\gamma(s_0) = 0$ to guarantee identifiability. More specifically, assume that

$$\gamma_j(s) = g(s; \gamma^{(j)}),\ j = 1, \ldots, J, \tag{4.7}$$

where $\gamma^{(j)} = (\gamma_{j1}, \ldots, \gamma_{jr})$ is a r_j-vector of regression parameters and $g(\cdot)$ is a parametric function of s. We gather the parameters from (4.6)-(4.7) in a vector denoted by γ. The model (4.3), (4.4), (4.6), (4.7), which we shall refer to as the landmark supermodel, can be fitted directly by applying a simple Cox model to the stacked data set, again provided that the software allows for delayed entry at s. Again, repeated observations of the same subject automatically lead to the presence of many ties. Fitting the model with the Breslow partial likelihood for those tied observations is equivalent to maximizing a different pseudo-likelihood, namely

$$\text{ipl}^*(\beta_{\text{LM}}, \gamma) = \sum_{j=1}^{J}\sum_i d_{ij} \ln\left(\frac{\sum_{s:\ s\le t_i\le s+w}\exp(Z_i(s)^\top\beta_j(s) + \gamma_j(s))}{\sum_{s:\ s\le t_i\le s+w}\sum_{t_k:\ t_k\ge t_i}\exp\left(Z_k(s)^\top\beta_j(s) + \gamma_j(s)\right)}\right). \tag{4.8}$$

The estimators of the corresponding baseline hazards are given by

$$\widehat{\lambda}^*_{j0}(t_i) = \frac{\#(s \le t_i \le s+w,\ D_i = j)}{\sum_{s:\ s\le t_i\le s+w}\sum_{t_k:\ s\le t_i\le t_k<s+w}\exp(Z_k(s)^\top\widehat{\beta}_j(s) + \widehat{\gamma}_j(s))},\ j = 1, \ldots, J.$$

Again, standard errors of the regression parameters can be obtained by sandwich estimators.

Let $\widehat{\Lambda}^*_{j0}(t) = \sum_{t_i\le t}\widehat{\lambda}^*_{j0}(t_i)$ be the cumulative cause-specific baseline hazard of cause j, $j = 1, \ldots, J$. Then the estimated dynamic prediction probabilities are given by substituting the estimators of β_{LM}, γ and λ_{j0} into (4.2), resulting in

$$\widehat{S}_{\text{LM}}(s+w|Z(s), s) = \exp\Big(-\sum_{j=1}^{J} e^{Z(s)\widehat{\beta}_j(s)+\widehat{\gamma}_j(s)}\left[\widehat{\Lambda}^*_{j0}(s+w) - \widehat{\Lambda}^*_{j0}(s-)\right]\Big) \tag{4.9}$$

and

$$\widehat{F}_{j,\text{LM}}(s+w|Z(s), s) = \sum_{s<t_i\le s+w}\widehat{\lambda}^*_j(t_i|Z(s))\widehat{S}_{\text{LM}}(t_i - |Z(s), s),\ j = 1, \ldots, J. \tag{4.10}$$

4.3.2 Application to the EBMT data

In this section we apply the theory described in Section 4.3.1 to the EBMT data where we shall initiate prediction anywhere in the interval $[0, 1]$ for a prediction interval of width $w = 5$. Competing endpoints are relapse (cause 1) and NRM (cause 2), while the covariate vector $Z(\cdot)$ comprises the prognostic covariates at baseline and low and high grade aGvHD with their counterparts, very recent, recent and past, respectively. We set up a grid of 13 landmark (prediction) time points $t_{\text{LM}} = 0, 1, \ldots, 12$ months.

The frequencies of the outcomes in each of the landmark data sets across the first year post-transplant are shown in Figure 4.3. The combination of high relapse and NRM rates and modest censoring in the first year post-transplant (see Figure 4.1) explains that the relative size of the alive/censoring part increases from 45% at $t_{\text{LM}} = 0$ to 70% at $t_{\text{LM}} = 12$ months.

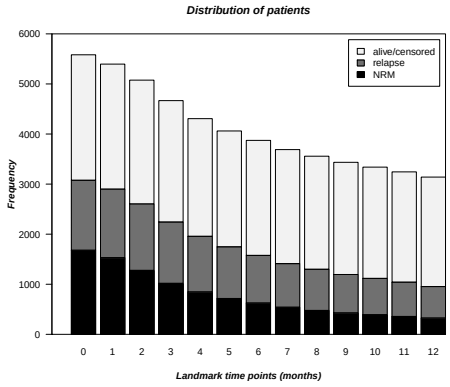


Figure 4.3: Frequencies of outcomes for each of the landmark data sets.

Figure 4.4 shows the frequencies of the values of the time-dependent covariates in each of the landmark data sets. For a given s, $s \in \{1, 2, \ldots, 12\}$ months, the corresponding bar comprises the frequencies of the relapse-free survivors patients at s who either have not developed aGvHD yet or who are subjected to low or high grade aGvHD exclusively in one of the intervals $[s,\ s-1]$ (very recent), $[s-1,\ s-6]$ (recent) or $[0,\ s-6]$ (past), when applicable; this last possibility corresponds to a change in value from 0 to 1 of one of the $Z_k^{VR}(s)$, $Z_k^R(s)$ and $Z_k^P(s)$, $k \in \{l,\ h\}$, respectively. White areas at s correspond to patients who developed low (bottom part) or high (upper part) grade aGvHD within one month before s (very recent),

therefore spanning from $s = 1$ to $s = 4$ only, in agreement with Figure 4.2. Light grey areas correspond to patients who developed low (bottom part) or high (upper part) grade aGvHD within one month and six months before s (recent), therefore spanning from $s = 2$ to $s = 9$ only; these include patients counted in white areas at earlier landmark time points who survived relapse-free at time s. Dark grey areas correspond to patients who developed low (bottom part) or high (upper part) grade aGvHD within more than six months before s (past), therefore spanning from $s = 7$ to $s = 12$ only; these include patients counted in light grey areas at earlier landmark time points who survived relapse-free at the corresponding s.

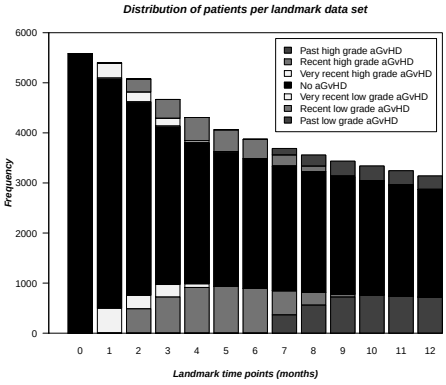


Figure 4.4: Frequencies of the values of the time-dependent covariates in each of the landmark data sets.

We fitted the ipl* - model to our data, with

$$\beta_j(s) = \beta_{j0} + \beta_{j1}s + \beta_{j2}s^2,\ j = 1, 2,$$

and

$$\gamma_j(s) = \gamma_{j1}s + \gamma_{j2}s^2,\ j = 1, 2.$$

For each of the competing end points $j = 1$ and $j = 2$, a backward selection procedure was used, starting from a model with all time-fixed covariates effects described by quadratic terms, where Wald tests were used to test whether the linear and quadratic terms could be removed. Interaction terms between $Z_k^{VR}(s)$, $Z_k^R(s)$ and $Z_k^P(s)$, $k \in \{l,\ h\}$, on the one hand, and landmark time points on the

other hand, were not tested because they would make the model overly complicated. From the model obtained after this initial selection procedure, a further backward selection procedure was used again using Wald tests to replace $Z_k^{VR}(t)$, $Z_k^R(t)$ and $Z_k^P(t)$ by $Z_k(t)$, $k \in \{l,\ h\}$. This resulted in the final supermodel reported in Table 4.4. For NRM, the interaction between year of transplantation and landmark time points was found to be non-significant, and was therefore not included in the model of NRM. The linear and quadratic terms of the EBMT risk score effects were significant for both causes of failure. Similarly to the analysis in Section 2, indication of when low and high grade aGvHD occurred in the past exhibits non-significant effects on relapse; only $Z_l(s)$ and $Z_h(s)$ show significant effects on relapse. In contrast, Wald tests showed differential effects of $Z_k^{VR}(s)$, $Z_k^R(s)$ and $Z_k^P(s)$, $k \in \{l,\ h\}$, on the cause-specific hazard of NRM and therefore they are kept in the model. A plot of the regression parameters of the EBMT risk score which vary with s together with their 95% confidence intervals is given in Figure 4.5. The effect of year on relapse varies with s; it decreases from 0.044 at $s = 0$ to 0 at $s = 1$. The detrimental effect of higher risk score shows a decreasing trend for later times s; the behaviour in the right tail might reflect the artefact of selecting healthier patients (those who are event-free before s, for late s) made intrinsically by the model. As expected, occurrence of aGvHD before s is beneficial for preventing relapse. We see that the earlier high grade aGvHD occurred before s, the higher the risk of NRM. The increase in the estimated regression coefficients for $Z_l^{VR}(s)$, $Z_l^R(s)$ and $Z_l^P(s)$ is unexpected. Note that especially for NRM the estimated effects of $Z_k^{VR}(s)$, $Z_k^R(s)$ and $Z_k^P(s)$, $k \in \{l,\ h\}$, are attenuated compared to those based on the time-dependent Cox models of Table 4.2. The explanation for this is that the estimates based on the supermodel are weighted averages of the possibly time-varying effects over the follow-up period $[s; s + 5]$ of the Cox model. The parameters γ_j connecting the baseline parameters at the different landmark time points are more difficult to be interpreted by themselves, but they can be interpreted in connection with the baseline hazards.

In Figure 4.6 we show the estimated cumulative baseline hazards $\widehat{\Lambda}_{j0}^*(t)$ and the exponent of the linear combination of the baseline parameters, $\exp(\gamma_j(s))$. The rapidly decreasing trend with increasing s in Fig. 4.6 (b) and the concave shape in Fig. 4.6 (a), especially for non-relapse mortality, leads us - via (4.6) and (4.10) - to expect a marked decrease in the dynamic prediction probabilities for the cumulative incidence function of non-relapse mortality with increasing s.

Table 4.4: Estimated regression parameters of the landmark supermodel for relapse and non-relapse mortality.

Covariate	Relapse		NRM	
	$\widehat{\beta}$	SE($\widehat{\beta}$)	$\widehat{\beta}$	SE($\widehat{\beta}$)
Year of transplantation				
Constant	0.044	0.016	-0.021	0.016
s	-0.070	0.033		
s^2	0.026	0.031		
Risk score				
Low risk				
Medium risk				
Constant	0.344	0.063	0.538	0.061
s	-0.287	0.121	-0.585	0.219
s^2	0.061	0.118	0.437	0.211
High risk				
Constant	1.171	0.099	1.216	0.091
s	-0.959	0.233	-1.552	0.356
s^2	0.352	0.225	1.201	0.330
Low grade aGvHD	-0.370	0.080		
Very recent			0.205	0.063
Recent			0.531	0.077
Past			0.616	0.113
High grade aGvHD	-0.873	0.155		
Very recent			1.392	0.062
Recent			1.369	0.079
Past			1.100	0.130
Baseline parameters				
γ_1	-0.173	0.095	-3.224	0.185
γ_2	-0.429	0.088	1.429	0.167

In Figure 4.7 we show the stacked estimated prediction probabilities for a patient transplanted in 2003. The lower curve represents $\widehat{F}_{1,\mathrm{LM}}$, the conditional cumulative incidence of relapse before $s+5$ years, given no relapse or NRM up to time s and no aGvHD. The distance between the top curve and the lower curve represents $\widehat{F}_{2,\mathrm{LM}}$, the conditional cumulative incidence of non-relapse mortality at $s+5$ years given no event up to time s and no aGvHD, while the distance between 1 and the upper curve represents $1-\widehat{F}_{1,\mathrm{LM}}-\widehat{F}_{2,\mathrm{LM}}$, the conditional probability

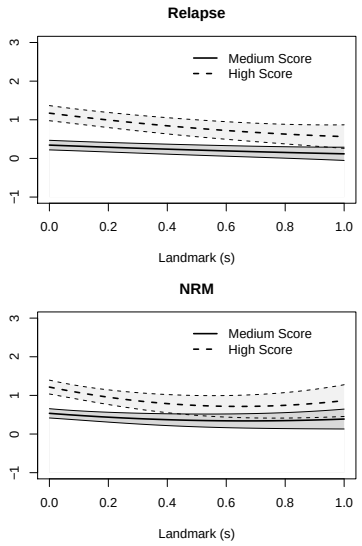


Figure 4.5: Time-varying regression coefficients $\beta(s)$ of EBMT risk score and associated pointwise 95% confidence intervals implied by the landmark supermodel of Table 4.

of being alive without relapse at $s + 5$, given no relapse or NRM up to time s and no aGvHD. By looking simultaneously at all the plots in this figure, we see that surviving relapse-free and having no aGvHD the first year after SCT greatly improves relapse-free survival after 5 years. As expected, the higher the risk score, the higher the cumulative incidence of failure of both causes and the smaller the relapse-free survival probability. Interestingly, the differences between the dynamic 5-years width cumulative incidences for the three risk scores are much smaller at the end of the prediction period ($s = 1$) than at the beginning ($s = 0$). Figures 4.8 and 4.9 clearly show that (especially high grade) aGvHD increases the conditional cumulative incidence of non-relapse mortality. The gradual decrease with later s of the cumulative incidences is a natural consequence of conditional probabilities estimated solely on individuals event-free at s. The reason why the time axes are truncated in Figures 4.8 and 4.9 in the presence of $Z_k^{VR}(s)$, $Z_k^R(s)$ and $Z_k^P(s)$, $k \in \{l,\ h\}$, is that these time-dependent covariates can only be non-zero in a limited time window (see Figure 4.4). The landmark model will allow us to do prediction in the presence of these covariates also outside these time

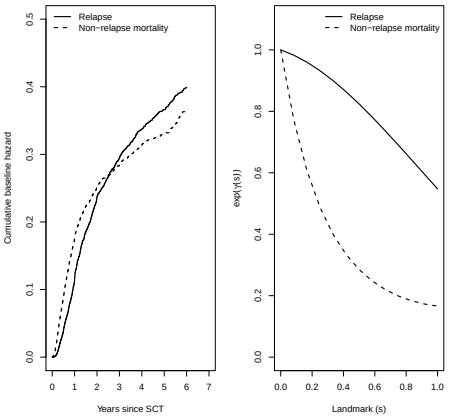


Figure 4.6: (a) Estimated cumulative baseline hazards; (b) Estimated γ's on exponential scale.

windows, but these predictions will not be meaningful in practice.

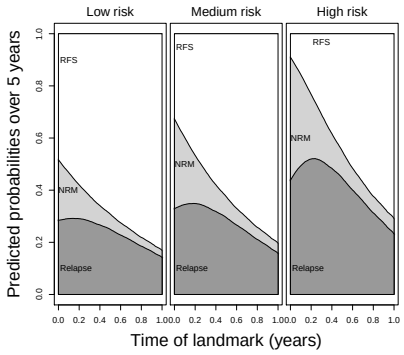


Figure 4.7: The estimated 5-years fixed width predictive cumulative incidences of relapse and NRM for a patient transplanted in 2003 with no aGvHD, for each of the levels of EBMT risk score.

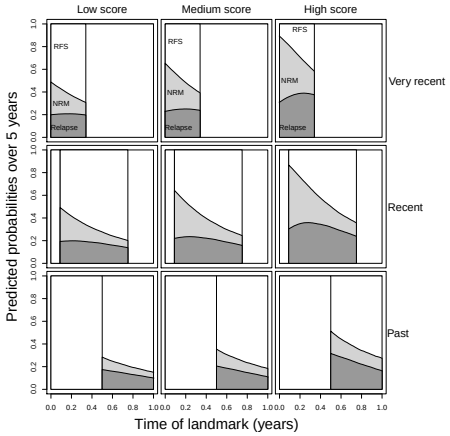


Figure 4.8: The estimated 5-years fixed width predictive cumulative incidences of relapse and NRM for a patient transplanted in 2003 with low grade aGvHD, for each of the levels of EBMT risk score.

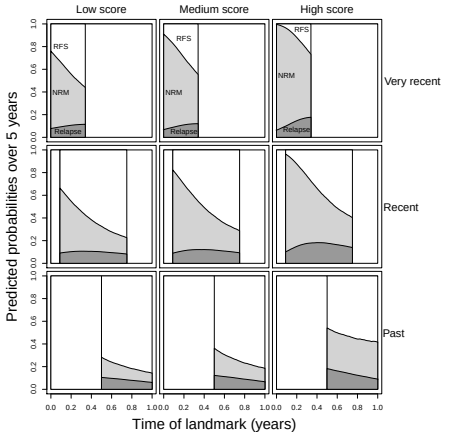


Figure 4.9: The estimated 5-years fixed width predictive cumulative incidences of relapse and NRM for a patient transplanted in 2003 with high grade aGvHD, for each of the levels of EBMT risk score.

Figure 4.10 shows an example of dynamic fixed width prediction, based on the landmark supermodel, for a patient transplanted in 2003, at different values of the EBMT risk score, who experienced high grade aGvHD at 1 month post-transplant. This patient initially follows the predicted probabilities of the Figure 4.7, until one month, when he/she switches to the predictions from the "very recent" cell of Figure 4.9. One month later, the aGvHD is no longer very recent, so he/she switches again, now to the predictions from the "recent" cell, until the seventh month, when the aGvHD is no longer recent. From that time on, the predictions from the "past" cell of Figure 4.9 are followed.

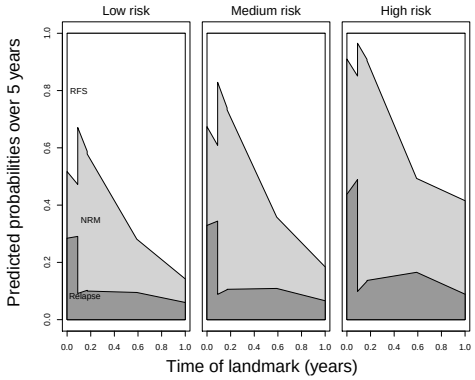


Figure 4.10: The estimated 5-years-fixed width predictive cumulative incidences for a patient transplanted in 2003 with high grade aGvHD at 1 month after SCT.(a) Low risk; (b) Medium risk; (c) High risk.

4.4 Discussion

In the practice of prediction, measurements are ascertained at baseline and patients are followed over time until the occurrence of a clinical outcome of interest. But this approach does not take into account the fact that the risk of the event of interests may change over time, depending on the evolution of the patient. Therefore, it is important to take all the available information into account when the intention is to predict future clinical events of interest at later points in time. In this paper we extended landmarking and in particular landmark supermodels to dynamic prediction in the presence of competing risks. We used these landmark supermodels based on the cause-specific hazards to obtain long term

dynamic prediction probabilities of the cumulative incidences of the causes of failure, accounting for baseline information and intermediate clinical events. The proposed method when applied per landmark is robust against violations of the hazard assumption in (4.3). The use of supermodels as proposed in this paper have the advantage of gaining efficiency but possibly at the cost of introducing model assumptions such as (4.4) and (4.7) which might not be met. The proposed landmark supermodel provides a prediction rule at baseline using only covariate information available at baseline. If time-dependent covariates are available, the landmark supermodel provides an updated prediction rule at the landmark time point s using covariate information collected up to s. Such a model is used for prognostic purposes without requiring a model for the distribution of the time-dependent covariate(s). The landmark approach is applicable for any choice of s and w, given that the time-dependent covariate values are known at the specific landmark time-points.

Interpretation of the regression coefficients of the landmark supermodel is not straightforward. This is already the case for ordinary survival analysis with a single endpoint, but even more so in the context of competing risks, because the present approach is based on the cause-specific hazards and there is no one-to-one correspondence to covariate effects on the cause-specific hazards and the cumulative incidences (Putter et al., 2007), which makes it hard to summarize the covariate effect on the cumulative scale. Moreover, it is hard to identify the time-varying effect on the cumulative incidence function for a specific covariate when regression is based on cause-specific hazards. Koller et al. (2011) made the point that for etiological research regression models based on the cause-specific hazards would usually be the most appropriate approach, while regression models based on the subdistribution hazards (Fine and Gray, 1999) would be most appropriate if prediction is the aim. These would argue for a dynamic prediction landmark model based on the Fine and Gray approach (Fine and Gray, 1999). This is a topic for future research; landmark supermodels based on the Fine and Gray approach are not straightforward because it is not clear whether or not to include in the s-landmark data set individuals with a competing event before the landmark time point s and what to use as value of their time-dependent covariates.

We have shown how the clinical event aGvHD can be used to provide a prognostic tool that can be updated for each new landmark time point. One limitation in our analysis is that time of onset of aGvHD and time of attaining highest grade are not (necessarily) the same but are taken as such in this analysis (because we do not have any more information); therefore, patients are inappropriately classified only by their final grade of aGvHD and the time they spend on "waiting" to reach a higher grade of aGvHD is incorrectly credited to higher grade of aGvHD. This could have lead to a bias the direction of which is difficult to predict in our specific setting as it is unclear how much time a patient spent 1. to reach the

stage of onset of aGvHD and 2. to reach higher grade aGvHD from the onset of aGvHD. A correct approach would require that the time to onset of aGvHD and the time from onset to higher grade aGvHD be known and accounted for.

5

Dynamic pseudo-observations: a robust approach to dynamic prediction in competing risks

Abstract

In this paper, we propose a new approach to the problem of dynamic prediction of survival data in the presence of competing risks as an extension of the landmark model for ordinary survival data of van Houwelingen (2007). The key feature of our method is the introduction of dynamic pseudo-observations constructed from the prediction probabilities at different landmark prediction times. They specifically address the issue of estimating covariate effects directly on the cumulative incidence scale in competing risks. A flexible generalized linear model based on these dynamic pseudo-observations and a generalized estimation equations approach to estimate the baseline and covariate effects will result in the desired dynamic predictions and robust standard errors. Our approach has a number of attractive features. It focuses directly on the prediction probabilities of interest, avoiding in this way complex modeling of cause-specific hazards or subdistribution hazards. As a result, it is robust against departures from these omnibus models. From a computational point of view an advantage of our approach is that it can be fitted with existing statistical software and that a variety of link

functions and regression models can be considered, once the dynamic pseudo-observations have been estimated. We illustrate our approach on a real data set of chronic myeloid leukemia patients after bone marrow transplantation.

5.1 Introduction

In medical studies, dynamic prediction of time-to-event data has recently received a lot of attention in terms of statistical development (van Houwelingen, 2007; van Houwelingen and Putter, 2008; Proust-Lima and Taylor, 2009; Cortese and Andersen, 2010; Parast et al., 2011; van Houwelingen and Putter, 2012), though clinical applications are seriously lagging behind with some happy exceptions (Zamboni et al., 2010; Sabatier et al., 2012). The dynamic aspect refers to how prognosis changes over the course of time as information on clinical events and/or measurements of biomarkers becomes available during follow-up.

In this paper we address the problem of dynamic prediction in the context of competing risks as an extension of the landmark model for ordinary survival data of van Houwelingen (2007). The advantages of landmarking for dynamic prediction are that the method is very direct and robust against misspecification of the proportional hazards assumption. These advantages are also valuable in competing risks situations especially in conjunction with the Fine and Gray model (Fine and Gray, 1999; Cortese and Andersen, 2010).

Currently, the use of landmarking in competing risks has been addressed in a number of papers (Cortese and Andersen, 2010; Cortese et al., 2013; Nicolaie et al., 2013a). In particular, Nicolaie et al. (2013a) have proposed supermodels which yield the dynamic cause-specific probabilities over an interval of prediction time points. In that approach, dynamic prediction probabilities were based on landmark models for the cause-specific hazards which were specified by proportional hazards models. However, for dynamic prediction primary interest is in the conditional cause-specific cumulative incidence, that is the probability of dying before some future time t from a specific cause in the presence of other risks, conditional on being alive at the prediction time. It is well-known that there is no one-to-one relation between covariate effects on the cause-specific hazards and on the cumulative incidence scale. Recently, a number of direct regression models have been proposed to assess covariate effects on cumulative incidences (Fine and Gray, 1999; Andersen et al., 2003; Klein and Andersen, 2005; Scheike et al., 2008); they primarily address the necessity of capturing the possibly time-varying covariate effects directly on the cumulative incidence scale. The aim of the present paper is to explore the possibility of combining the landmark paradigm and direct modeling the cumulative incidence function using pseudo-observations. We propose supermodels based on what we call *dynamic pseudo-observations* associated with cumulative incidences, that is pseudo-observations updated at each

prediction time point. Similar ideas have been proposed in Cortese et al. (2013), but for a different purpose, namely estimation of prediction error of dynamic prediction probabilities. These supermodels are intended to directly yield the desired prediction probabilities of cumulative incidences of the event of interest in a fixed, prespecified prediction time point, avoiding the burden of complex modeling of the complete survival process jointly with the covariate process.

Our paper is organized as follows: in Section 5.2 we introduce and discuss our approach. In Section 5.2.1 we lay out the notation. In Section 5.2.2 we define dynamic pseudo-observations and illustrate some of their asymptotic properties. In Section 5.2.3 we describe regression models based on pseudo-observations and we show how to obtain dynamic prediction probabilities from these models. In Section 5.3 we apply the approach to data from European Group for Bone and Marrow Transplantation (EBMT). Section 5.4 concludes the paper with a discussion.

5.2 Dynamic prediction

5.2.1 Notation

Suppose that data are available from n individuals each of whom can experience one of J types of failure, which we term $1, \ldots, J$, respectively, or can be subject to noninformative censoring within a time interval $[0, \tau]$. Let $\tilde{T}$ denote the time of failure, C the censoring time, and D the cause of failure. Let $\mathbf{Z}(\cdot)$ denote a p-vector of covariates, which could be measured at baseline or be time-dependent; this could be either internal or external covariates. $\mathbf{Z}(\cdot)$ is shorthand for the entire covariate process, which is assumed to be observed without error over the time interval(s) for which the individual is at risk. The observed data for individual i is $\{T_i, \Delta_i, \mathbf{Z}_i(\cdot)\}$, where $T_i = \min(\tilde{T}_i, C_i)$ is the earliest of failure and censoring time, and $\Delta_i = \mathbf{1}(\tilde{T}_i < C_i) \cdot D_i$ is the cause of failure in case of failure and 0 in case of censoring. Data from different individuals are assumed to be independent. Let S be the survival function of $\widetilde{T}$ and G the survival function of C.

Let $s < t$ be two time points and define the cumulative incidence of event j, conditional on being event-free at time s, by

$$F_j(t|s) = P(T \leq t, D = j | T > s), \text{ for } j = 1, \ldots, J.$$

We denote by $t_1 < t_2 < \ldots$ the times at which events occur irrespective of the cause. Let $d_j(t_k)$ be the number of individuals who die at time t_k from cause j and $d(t_k) = \sum_{j=1}^{J} d_j(t_k)$ be the number of individuals who die at time t_k from any cause. Let $r(t_k)$ be the number of individuals at risk just prior to time t_k. Let $\widehat{F}_j(\cdot|s)$ be the non-parametric estimator of the conditional probability $F_j(\cdot|s)$,

as given by

$$\widehat{F}_j(t|s) = \sum_{s<t_k\leq t} \widehat{S}(t_k - |s)\frac{d_j(t_k)}{r(t_k)} \ , \tag{5.1}$$

where

$$\widehat{S}(t-|s) = \prod_{s<t_l\leq t} \{1 - \frac{d(t_l)}{r(t_l)}\}$$

is the Kaplan-Meier estimate of the conditional survival function given no event before time s. Consistency of $\widehat{F}_j(\cdot|s)$ as an estimator of $F_j(\cdot|s)$ follows from the same arguments as in Aalen and Johansen (1978) under the assumption of independence of $\widetilde{T}$ and C.

5.2.2 Dynamic pseudo-observations in competing risks

Let $[s_1, s_K]$ be the interval within which we want to obtain dynamic prediction probabilities and denote by w a fixed prediction window width. Our aim is to obtain estimators for the dynamic fixed width prediction probabilities $P\{T \leq s + w, D = j|\mathbf{Z}(s), T > s\}$ for varying $s \in [s_1, s_K]$ and fixed prediction width w. Let us fix a landmark time point s in $[s_1, s_K]$. In the following, we refer to the corresponding landmark data set as $\mathcal{L}_s$, obtained by selecting only the individuals who are at risk at time s, and fixing the time-dependent covariates at their current value at s, that is $\mathbf{Z}(s)$. Denote by n_s the sample size of this landmark data set $\mathcal{L}_s$.

Define the dynamic pseudo-observation within $\mathcal{L}_s$ for $\mathbf{1}(T \leq s+w, D = j)$ for individual i at risk in s as

$$\hat{\theta}^j_{isw} = n_s\widehat{F_j}(s + w|s) - (n_s - 1)\widehat{F_j}^{(-i)}(s + w|s) \ , \tag{5.2}$$

where $\widehat{F_j}^{(-i)}$ is the Aalen-Johansen estimator (6.12) based on the sample of size $n_s - 1$ obtained by eliminating individual i from $\mathcal{L}_s$, for $j = 1, \ldots, J$ and $i = 1, \ldots, n_s$ (see also Cortese et al. (2013)). We stress that $\hat{\theta}^j_{isw}$ is defined if and only if T_i exceeds s, i.e. for an individual i at risk at s, irrespective of whether a failure or censoring occurs before time $s + w$. These dynamic pseudo-observations can be computed using existing statistical software (Klein et al., 2008). Note that for complete data, i.e. data with no censoring, in $\mathcal{L}_s$ the relationships $\widehat{\theta}^j_{isw} = \mathbf{1}(T_i \leq s + w, D_i = j)$ hold, for $i = 1, \ldots, n_s$. In the following we shall focus on a fixed cause of failure j and fixed prediction width w and to simplify notation we therefore suppress j and w in the notation of $\widehat{\theta}^j_{isw}$ and replace it by $\widehat{\theta}_{is}$.

Following Graw et al. (2009), we will impose the following conditions:

(C1) The censoring time C is independent of $\{T, D, \mathbf{Z}(\cdot)\}$;

(C2) Any prediction time point s satisfies $s + w < \tau$ with $G(\tau) > 0$ and $S(\tau) > 0$.

Some asymptotic properties of dynamic pseudo-observations which are important in the next section, when regression models based on dynamic pseudo-observations are considered, are gathered in the following proposition; see Appendix B for a brief proof.

Proposition 5.2.1 *Assume that (C1) and (C2) hold. Then the following properties hold:*

(P1) $\widehat{\theta}_{is}$ *is asymptotically independent of* $\widehat{\theta}_{ls}$ *for individuals* $i \neq l$ *as* n *tends to infinity.*

(P2) $\widehat{\theta}_{is}$ *is asymptotically independent of* $\widehat{\theta}_{ls'}$ *for individuals* $i \neq l$ *and landmark time points* $s \neq s'$ *as* n *tends to infinity.*

(P3) $E\{\widehat{\theta}_{is} | \boldsymbol{Z}_i(s),\ T_i > s\}$ *equals asymptotically its theoretical counterpart* $E\{\boldsymbol{1}(T_i \leq s + w, D_i = j) | \boldsymbol{Z}_i(s),\ T_i > s\}$ *as* n *tends to infinity.*

5.2.3 Specification of models

In this section we consider regression models based on the dynamic pseudo-observations. In Subsection 5.2.3 we consider models for fixed s, while Section 5.2.3 will be devoted to supermodels combining dynamic pseudo-observations for a collection of landmark time points.

Models for fixed landmark time points

Let us fix s. For complete data, i.e., when no censoring occurs, we give a derivation in Appendix A that shows how dynamic prediction can be achieved based on (observable) binomial variable $Y(s) = \mathbf{1}(T \leq s + w, D = j)$, using standard modeling approaches for these variables. In the presence of censoring, we do not always observe the binomial random variable $Y(s) = \mathbf{1}(T \leq s + w, D = j)$. For this reason, we replace the (unobservable) $Y_i(s)$ by the dynamic pseudo-observations $\widehat{\theta}_{is}$ based on the sample $\mathcal{L}_s$; the $\widehat{\theta}_{is}$ are meant to mimic the behaviour of $Y_i(s)$, in the sense that they approximate a $0 - 1$ variable and they are asymptotically independent, as stated in Proposition 5.2.1. Let

$$\mu_{is} = E\{Y_i(s) | \mathbf{Z}_i(s),\ T_i > s\}. \tag{5.3}$$

We shall postulate a generalized linear model on the binomial expectations μ_{is} of the form

$$g[E\{Y_i(s) | \mathbf{Z}_i(s),\ T_i > s\}] = \boldsymbol{\beta}^\top(s) \mathbf{Z}_i^*(s)\ , \tag{5.4}$$

for a given link function g, where $\boldsymbol{\beta}(s) = \{\beta_0(s), \beta_1(s), \ldots, \beta_p(s)\}$ and $\mathbf{Z}^*(s) = \{1,\ \mathbf{Z}^\top(s)\}^\top$, so that $\beta_0(s)$ stands for the intercept. In the following, since we

consider a fixed value of s we shall suppress the dependence of $\boldsymbol{\beta}(s)$ on s in the notation.

Since we can approximate μ_{is} by $E\{\widehat{\theta}_{is}|\mathbf{Z}_i(s),\ T_i > s\}$, as stated in Proposition 5.2.1, our estimate of $\boldsymbol{\beta}$ is based on the quasi-score equations (see McCullagh and Nelder, 1999), where $Y_i(s)$ is replaced by the pseudo-observations:

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n_s} \mathbf{U}_i(\boldsymbol{\beta}) = \sum_{i=1}^{n_s} \frac{\partial \mu_{is}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \cdot \frac{1}{\mu_{is}(1-\mu_{is})} \cdot (\widehat{\theta}_{is} - \mu_{is}) = 0 \ , \tag{5.5}$$

with μ_{is} as defined in (5.3). Note that this model was used in Klein and Andersen (2005) for prediction in the competing risks framework for the special case where $s = 0$.

With specific choices of the link function g, equation (5.5) can again be simplified. For the logit link function $g(x) = \log \frac{x}{1-x}$, (5.5) simplifies to

$$\sum_{i=1}^{n_s} \mathbf{Z}_i^* \cdot (\widehat{\theta}_{is} - \mu_{is}) = 0 \ , \tag{5.6}$$

and for the cloglog link function $g(x) = \log\{-\log(1-x)\}$ to

$$\sum_{i=1}^{n_s} \mathbf{Z}_i^* \cdot \frac{\log(1-\mu_{is})}{\mu_{is}} \cdot (\widehat{\theta}_{is} - \mu_{is}) = 0.$$

Model (5.4) with the cloglog link function is equivalent to

$$\Lambda_{ij}^*(s+w|s) = \exp\left(\boldsymbol{\beta}^\top(s)\mathbf{Z}_i^*(s)\right),$$

where $\Lambda_{ij}^*(s+w|s) = \int_s^{s+w} \lambda_{ij}^*(u|s)du$, with λ_{ij}^* the subdistribution hazard of cause j of subject i. Thus, it has clear similarities with the Fine-Gray model on the subdistribution hazard of cause j: covariate $\mathbf{Z}_i(s)$ in (5.4) plays the role of the time-independent covariate in the Fine-Gray model. However, the important distinction is that the Fine-Gray model is being used here only at a single time point, $s+w$, conditionally on no event yet at s. Since only the functional relation between covariates and conditional cumulative incidence function at $s+w$ is used in model (5.4), and no proportional hazards assumption of the Fine-Gray model on the subdistribution hazard over the entire follow-up, the estimates $\boldsymbol{\beta}$ are robust against departures from the proportionality assumption of the subdistribution hazards. This could however be at a cost of loss of efficiency if the proportionality assumption of the Fine and Gray model is true. A similar distinction holds between our estimating equations in (5.5), which uses pseudo-observations of the conditional cumulative incidence functions at a single time point $s+w$, and those

in Klein and Andersen (2005), who use several pseudo-observations, evaluated at different prediction widths; see also the Discussion section. Yet another choice of link function is $g(x) = \log(x)$, which has similarities with absolute risk regression (Gerds et al., 2012). See Gerds et al. (2012) for a discussion on the choice of link functions and on the interpretation of regression coefficients in these models.

Define the sandwich estimate by

$$\{\mathbf{I}(\widehat{\boldsymbol{\beta}})\}^{-1} \cdot \widehat{\text{var}}\{\mathbf{U}(\widehat{\boldsymbol{\beta}})\} \cdot \{\mathbf{I}(\widehat{\boldsymbol{\beta}})\}^{-1} , \tag{5.7}$$

where

$$\mathbf{I}(\boldsymbol{\beta}) = \frac{1}{n_s} \sum_{i=1}^{n_s} \frac{\partial \mu_{is}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \cdot V_i^{-1}(\boldsymbol{\beta}) \cdot \Big\{ \frac{\partial \mu_{is}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big\}^\top$$

with

$$V_i(\boldsymbol{\beta}) = \mu_{is}(\boldsymbol{\beta}) \cdot \{1 - \mu_{is}(\boldsymbol{\beta})\}$$

and

$$\widehat{\text{var}}\{\mathbf{U}(\boldsymbol{\beta})\} = \frac{1}{n_s} \sum_{i=1}^{n_s} \mathbf{U}_i(\boldsymbol{\beta}) \cdot \{\mathbf{U}_i(\boldsymbol{\beta})\}^\top.$$

We will impose the following additional condition:

(C3) g is invertible and its inverse g^{-1} is continuously differentiable at each of the $\boldsymbol{\beta}^\top \mathbf{Z}_i$.

Proposition 5.2.2 *Assume model (5.4) is correctly specified. Under conditions (C1), (C2) and (C3), the solution $\widehat{\boldsymbol{\beta}}$ to (5.5) is consistent and asymptotically normal for estimating the parameter $\boldsymbol{\beta}$ of model (5.4):*

$$\sqrt{n_s}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

where the asymptotic variance $\boldsymbol{\Sigma}$ can be consistently estimated by (5.7).

See Appendix B for a brief proof of Proposition 5.2.2.

Let $\widetilde{\mathbf{Z}}(s)$ be the covariate vector for a new patient at prediction time s and consider $F_j(s + w|s, \widetilde{\mathbf{Z}}(s)) = P(T \leq s + w, D = j|T > s, \widetilde{\mathbf{Z}}(s))$. Consider the estimate given by

$$\widehat{F}_j(s + w|s,\ \widetilde{\mathbf{Z}}(s)) = g^{-1}(\widehat{\boldsymbol{\beta}}^\top \widetilde{\mathbf{Z}}^*(s)) , \tag{5.8}$$

obtained by plugging in the estimated $\widehat{\boldsymbol{\beta}}$, the solution to (5.5), in (5.4) and where $\widetilde{\mathbf{Z}}^*(s) = (1,\ \widetilde{\mathbf{Z}}(s)^\top)^\top$.

Proposition 5.2.3 *Under conditions (C1), (C2) and (C3), a consistent, asymptotically normal distributed estimator of* $F_j(s+w|s, \widetilde{\boldsymbol{Z}}(s))$ *is given by (5.8). The variance of (5.8) is estimated consistently by*

$$\left\{\frac{dg^{-1}(x)}{dx}\right\}^2_{|x=\widehat{\boldsymbol{\beta}}^\top \widetilde{\boldsymbol{z}}^*}(s) \cdot (\widetilde{\boldsymbol{Z}}^*(s))^\top \cdot \widehat{var}(\widehat{\boldsymbol{\beta}}) \cdot \widetilde{\boldsymbol{Z}}^*(s) \ , \tag{5.9}$$

with $\widehat{var}(\widehat{\boldsymbol{\beta}})$ *as derived from Proposition 5.2.2.*

See Appendix B for a brief proof of Proposition 5.2.3.

Supermodels for dynamic pseudo-observations

In Section 5.2.3 we have considered models based on dynamic pseudo-observations $\widehat{\theta}_{is}$ obtained from a landmark data set $\mathcal{L}_s$ for a fixed landmark point s. These separate models can be used to estimate $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(s)$ and to obtain estimates of the prediction probabilities $F_j(s+w|s, \widetilde{\mathbf{Z}}(s))$ for several different values of s. As in landmark prediction models for ordinary survival (van Houwelingen, 2007), we would expect $\widehat{\boldsymbol{\beta}}(s)$ to vary smoothly with s. This idea can be exploited by combining dynamic pseudo-observations $\widehat{\theta}_{is}$ from different landmark data sets. To this end, define a set of landmark time points $0 \leq s_1 < \ldots < s_K$, such that $s_K + w < \tau$, and construct the corresponding landmark data sets $\mathcal{L}_k := \mathcal{L}_{s_k}$; each of these comprises the $n_k := n_{s_k}$ individuals at risk at time s_k only. Within each $\mathcal{L}_k$ we fix the time-dependent covariates at their current values at s_k, that is we use $\mathbf{Z}(s_k)$.

For each prediction point s_k we estimate the cumulative incidence function of cause j at $s_k + w$ based on the complete landmark data set $\mathcal{L}_k$ of size n_k and the cumulative incidence function of cause j based on the sample of size $n_k - 1$ obtained by deleting the ith observation from $\mathcal{L}_k$. We then define the dynamic pseudo-observation $\widehat{\theta}^j_{i,s_k,w}$ of individual i at time $s_k + w$ as in (6.13) and denote it shorthand by $\widehat{\theta}_{ik}$. When only right censoring is present, individual i with failure or censoring time t_i will be represented in landmark data sets $\mathcal{L}_k$ for all k such that $s_k < t_i$. For landmark points $s_k \geq t_i$ the individual will not be in the risk set. Let $\mathcal{S}_i \subset \{1, \ldots, K\}$ denote the set of indices k of the landmark time points such that individual i is at risk at s_k and let l_i be the last index in $\mathcal{S}_i$. Define the l_i-vector of dynamic pseudo-observations of individual i by

$$\widehat{\boldsymbol{\theta}}_i = \{\widehat{\theta}_{ik},\ k \in \mathcal{S}_i\}.$$

Our goal is to specify a regression model for

$$\mu_{is} = E\{Y_i(s) \,|\, \mathbf{Z}_i(s), T_i > s\}.$$

Thus, our target of inference, in the terminology of Kurland and Heagerty (2005), is a partly conditional mean model, conditioning on being alive (Pepe et al., 1999). This implies that our regression model for $\mu_{ik} = \mu_{i,s_k}$ conditions on being alive at s_k; the cohort of survivors at s_k might comprise survivors, dead or censored patients at $s_k + w$. For a link function g, assume

$$g(\mu_{ik}) = \boldsymbol{\beta}^\top(s_k)\mathbf{Z}_i^*(s_k) \ , \tag{5.10}$$

where $\mu_{ik} = \mu_{i,s_k}$ with μ_{is} as defined in (5.3), $\boldsymbol{\beta}(s) = \{\beta_0(s), \beta_1(s), \ldots, \beta_p(s)\}$ and $\mathbf{Z}^*(s) = \{1, \ \mathbf{Z}(s)^\top\}^\top$, so that $\beta_0(s)$ stands for the intercept. To model the time-dependent behaviour of $\boldsymbol{\beta}(s)$ across $s \in [s_1, s_K]$, we can employ a linear model for the lth component of $\boldsymbol{\beta}(s)$

$$\beta_l(s) = \boldsymbol{\beta}_l^\top \mathbf{h}_l(s), \ \ s \in [s_1, s_K] \ , \tag{5.11}$$

where $\mathbf{h}_l(s)$ is a suitable set of basis functions and $\boldsymbol{\beta}_l$ is a vector of parameters, for $l = 0, \ldots, p$. Different covariates $Z_l(s)$ may use different time functions $\mathbf{h}_l(s)$, $l = 0, \ldots, p$. Define $\boldsymbol{\beta}$ to be the vector with length q containing all $\boldsymbol{\beta}_l$ vectors. Then the vector $\boldsymbol{\beta}(s)$ can be written as $\mathbf{H}(\mathrm{s})\boldsymbol{\beta}$, with $\mathbf{H}(s)$ a $(p+1) \times q$ matrix containing the basis functions. We shall refer to the model (5.10)-(5.11) as the landmark supermodel.

For a link function g and independence working correlation, a linear quasi-score equation for regression parameter vector $\boldsymbol{\beta}$ is given by

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{U}_i(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \cdot \mathbf{V}_i^{-1} \cdot (\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\mu}_i) = \ 0 \ , \tag{5.12}$$

with $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}) = \{\mu_{i1}(\boldsymbol{\beta}), \ldots, \mu_{il_i}(\boldsymbol{\beta})\}^\top$ and $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\beta})$ an $l_i \times l_i$ diagonal matrix with elements $V_{ik}(\boldsymbol{\beta}) = \mu_{ik}(\boldsymbol{\beta}) \cdot \{1 - \mu_{ik}(\boldsymbol{\beta})\}$.

Define the sandwich estimator by

$$\{\mathbf{I}(\widehat{\boldsymbol{\beta}})\}^{-1} \cdot \widehat{\mathrm{var}}\{\mathbf{U}(\widehat{\boldsymbol{\beta}})\} \cdot \{\mathbf{I}(\widehat{\boldsymbol{\beta}})\}^{-1} \ , \tag{5.13}$$

where

$$\mathbf{I}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \cdot V_i^{-1}(\boldsymbol{\beta}) \cdot \Big\{ \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big\}^\top$$

with

$$\widehat{\mathrm{var}}\{\mathbf{U}(\boldsymbol{\beta})\} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{U}_i(\boldsymbol{\beta}) \cdot \{\mathbf{U}_i(\boldsymbol{\beta})\}^\top.$$

Given correctly specified regression models on the means μ_{ik} and on the $\boldsymbol{\beta}(s)$, the above generalized estimating equations (GEE) approach leads to consistent

estimators of parameters $\boldsymbol{\beta}$, as specified in the next

Proposition 5.2.4 *Assume models (5.10) and (5.11) are correctly specified. Under conditions (C1), (C2) and (C3) and assuming independence working correlation, the solution $\widehat{\boldsymbol{\beta}}$ to (5.12) is consistent and asymptotically normal for estimating $\boldsymbol{\beta}$ of models (5.10) and (5.11):*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

where the asymptotic variance-covariance matrix $\boldsymbol{\Sigma}$ can be consistently estimated by (5.13).

A short proof of Proposition 5.2.4 is given in Appendix B and relies on showing that the estimating equations in (5.12) are asymptotically unbiased. Kurland and Heagerty (2005, p. 247) argue that with non-independence working correlation, the solution to (5.12) is no longer guaranteed to be consistent. The reason is that for non-independence working correlation additional random terms multiplying $(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\mu}_i)$ appear in the estimating equations coming from the inverse of $\mathbf{V}_i$, which may disrupt the (asymptotic) unbiasedness of these estimating equations.

Let $\widetilde{\mathbf{Z}}(s)$ be the covariate vector for a new patient at prediction time s and consider $F_j(s + w|s, \widetilde{\mathbf{Z}}(s)) = P(T \leq s + w, D = j|T > s, \widetilde{\mathbf{Z}}(s))$. An estimate of this quantity is obtained by calculating $\widehat{\boldsymbol{\beta}}(s) = \mathbf{H}(s)\widehat{\boldsymbol{\beta}}$, for $\widehat{\boldsymbol{\beta}}$ the solution to (5.12), and s and setting

$$\widehat{F}_j(s + w|\ s, \widetilde{\mathbf{Z}}(s)) = g^{-1}\{\widehat{\boldsymbol{\beta}}(s)^\top \widetilde{\mathbf{Z}}^*(s)\}\ , \tag{5.14}$$

where $\widetilde{\mathbf{Z}}^*(s) = (1,\ \widetilde{\mathbf{Z}}^\top(s))^\top$.

Proposition 5.2.5 *Assume (5.10) and (5.11) are correctly specified. Under conditions (C1), (C2) and (C3), the estimator given by (5.14) is a consistent estimator of $F_j(s + w|s, \boldsymbol{Z}(s))$ for any $s \in [s_1, s_K]$ and its variance is estimated consistently by*

$$\left\{\frac{dg^{-1}(x)}{dx}\right\}^2_{|x=\widehat{\boldsymbol{\beta}}(s)^\top \widetilde{\boldsymbol{z}}^*(s)} \cdot (\widetilde{\boldsymbol{Z}}^*(s))^\top \cdot \boldsymbol{H}(s) \cdot \widehat{var}(\widehat{\boldsymbol{\beta}}) \cdot \boldsymbol{H}(s)^\top \cdot \widetilde{\boldsymbol{Z}}^*(s)\ , \tag{5.15}$$

with $\widehat{var}(\widehat{\boldsymbol{\beta}})$ as derived from Proposition 5.2.4.

See Appendix B for a brief proof of Proposition 5.2.5.

5.3 Exploratory data analysis

5.3.1 Data

Our data is the same as in Nicolaie et al. (2013a) and consists of 5582 chronic myelogenous leukaemia (CML) patients, registered at the EBMT who received allogeneic stem cell transplantation (SCT) between 1997 and 2003. We also use the same time-dependent covariates (acute graft versus host disease (aGvHD), which is a major complication in SCT, being associated with increased mortality and decreased relapse rates) and endpoints relapse and non-relapse mortality (NRM). It was shown in cause-specific hazards analysis (Nicolaie et al. (2013a)) that aGvHD is significantly associated with the two endpoints; aGvHD was further classified into *low* grade aGvHD, corresponding to grade 2 or lower, and *high* grade aGvHD, corresponding to grade 3 or higher. Let

$$\begin{aligned} Z_{\text{low}}(t) &= \mathbf{1}(\text{occurrence of low grade aGvHD before time } t) \text{ and} \\ Z_{\text{high}}(t) &= \mathbf{1}(\text{occurrence of high grade aGvHD before time } t) \end{aligned}$$

be binary time-dependent covariates which refer to having experienced low or high grade aGvHD, respectively, before time t. Prognostic information at baseline includes: year of SCT, a continuous covariate which was centered around 2000 and divided by 10 in our analysis, and the EBMT risk score, a prognostic index based on available marker information at baseline, known to be predictive of both relapse and NRM. The latter covariate is divided into three risk groups, denoted as low risk (set as reference value in the analysis), medium risk and high risk. The frequencies of the values of these covariates are shown in Table 5.1.

Table 5.1: Prognostic factors for all patients.

Prognostic factor	Category	n (%)
Risk score	Low	2361 (42%)
	Medium	2663 (48%)
	High	558 (10%)
Year of SCT	1997	773 (14%)
	1998	956 (17%)
	1999	1004 (18%)
	2000	956 (17%)
	2001	669 (12%)
	2002	658 (12%)
	2003	566 (10%)

A stacked plot of the estimated cumulative incidences of relapse and NRM, shown in Figure 5.1a, suggests that the majority of events occur within the first 5 years post-transplant. The estimated cumulative incidence functions of low and high grade aGvHD, respectively, displayed in Figure 5.1b, show a steep increase up to approximately first four months post-transplant only, due to the fact that by definition this intermediate event occurs quite early in the follow-up interval (within 100 days after SCT).

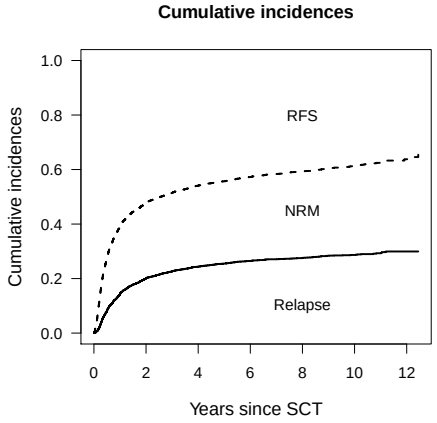
The clinical purpose of our approach is to obtain a dynamic prognostic model of conditional cumulative incidence functions of relapse and NRM at pre-specified prediction time points, given the history of aGvHD and the baseline covariates and given that no terminal event has occurred yet. To facilitate comparison with the landmark dynamic prediction model based on cause-specific hazards, we shall use the same grid of 13 landmark time points consisting of the first 12 months post-transplant including 0, namely $s = 0, 1/12, \ldots, 1$ years and the same prediction width $w = 5$ years as in Nicolaie et al. (2013a).

5.3.2 Dynamic pseudo-observations

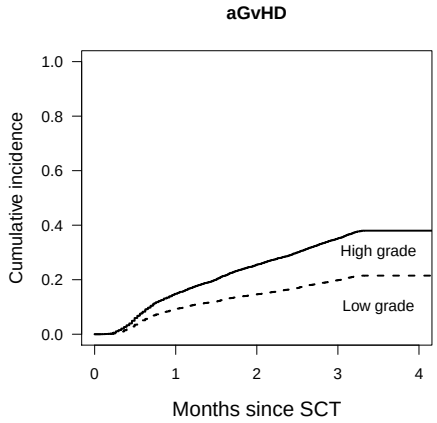
In Appendix C, we show examples of what dynamic pseudo-observations may look like for the entire grid of landmark time points under four different scenarios. The two main messages are: 1. $\widehat{\theta}_{is}^{j}$ resembles $\mathbf{1}(T_i \leq s+w, D_i = j)$ and 2. the closer in time are the landmark time points s and $s^{'}$, the stronger the correlation between the dynamic pseudo-observations $\widehat{\theta}_{is}^{j}$ and $\widehat{\theta}_{is^{'}}^{j}$ is. More specifically, in our data correlation decreases from an average of 0.995 (for landmark time points s and $s^{'}$ 1 month apart) to 0.912 (for s and $s^{'}$ 12 months apart) for relapse and from an average of 0.996 (for landmark time points 1 month apart) to 0.903 (for landmark time points 12 month apart) for NRM.

5.3.3 Dynamic prediction by landmarking using dynamic pseudo-observations

In this section we shall apply the theory described in Section 5.2 to the EBMT data where we initiate prediction anywhere in the interval $[0, 1]$ years for a fixed prediction interval width of $w = 5$ years. Competing endpoints are relapse (cause 1) and NRM (cause 2), while the covariate vector $\mathbf{Z}(\cdot)$ comprises the prognostic covariates at baseline and low and high grade aGvHD. We set up a grid of 13 landmark (prediction) time points $s = 0, 1/12, \ldots, 1$ years.



(a)



(b)

Figure 5.1: Stacked cumulative incidences of the competing events relapse and NRM (a), stacked cumulative incidence functions of low and high grade aGvHD (b).

Separate analyses per landmark time point

Here we follow the approach described in Section 5.2.3, separating the analysis per landmark time point s, for all s. We show results based on the logit link function, i.e., we fit the GLM model (5.6) on the dynamic pseudo-observations of each of the two competing events to model the effects on outcome of the baseline covariates and time-dependent covariates fixed at their current value in s. The resulting regression parameter estimates together with 95% confidence intervals based on the sandwich estimators of their standard errors as given by (5.7) are reported in Table 5.2 for a coarsened subset of the landmark time points.

Looking simultaneously at the estimates, the effect of year of SCT appears to be important in the beginning ($s = 0$) and quickly looses importance afterwards. At $s = 0$ more recent year of SCT (Table 5.2) implies higher 5-years risk of relapse and lower risk of NRM. In Figure 5.2 we display as error bars the estimated regression coefficients of intercept $\widehat{\beta}_0(s)$ (corresponding to the reference value, low risk, of EBMT risk score) and of intercept plus effects of medium and high risk score, implied by these separate landmark models, along with associated 95% confidence intervals. Both for relapse and NRM, the three error bars in Figure 5.2 develop in parallel which suggests that the effects of medium and high risk scores remain stable over time for both competing events. Figure 5.2 also shows that overall prognosis of NRM improves over time, illustrated by the steep decrease of the intercept; this is in accordance with Figure 5.1a, which shows a steep increase of the cumulative incidence of NRM in the first year which quickly levels off afterwards. Similar plots of year of SCT and low/high grade aGvHD are shown in Figures 5.7, 5.8 and 5.9. Dynamic prediction results from these models using (5.8) are shown in Figure 5.3 and discussed in conjunction with the supermodels of the next section.

Supermodels

In this section we apply the approach of Section 5.2.3, again using the logit link function. To apply our approach we used independence working correlation (see our remark just below Proposition 5.2.4). We fitted a landmark supermodel (5.10)-(5.11) on $\widehat{\theta}_{is}$ with covariate vector $\mathbf{Z}(\cdot)$ and regression parameters for each covariate $Z_l(\cdot)$, $l = 1, \ldots, p$, of the form $\beta_l(s) = \beta_{l0} + \beta_{l1}s + \beta_{l2}s^2$, $s \in [0, 1]$, and g the logit link function. For each of the competing events, a backward selection procedure was used, starting from a model with all time-fixed covariates effects described by quadratic terms, where Wald tests were used to test whether the linear and quadratic terms could be removed. This resulted in the final supermodel reported in Table 5.3.

For the two competing causes, the interaction between year of transplantation and landmark time points was found to be significant, and was therefore

Table 5.2: Estimated regression parameters of the stratified landmark models for relapse and non-relapse mortality.

	Relapse				
	0 months	3 months	6 months	9 months	12 months
Covariate	$\widehat{\beta}$ (SE)	$\widehat{\beta}$ (SE)	$\widehat{\beta}$ (SE)	$\widehat{\beta}$ (SE)	$\widehat{\beta}$ (SE)
Intercept	-1.186 (0.051)	-0.945 (0.056)	-1.041 (0.063)	-1.173 (0.067)	-1.321 (0.073)
Year of SCT	0.570 (0.170)	0.230 (0.190)	0.090 (0.214)	0.000 (0.234)	-0.080 (0.255)
Risk score					
Low risk					
Medium risk	0.138 (0.069)	0.214 (0.075)	0.174 (0.083)	0.129 (0.090)	0.125 (0.099)
High risk	0.544 (0.107)	0.828 (0.129)	0.734 (0.155)	0.722 (0.176)	0.684 (0.196)
Low grade aGvHD		-0.547 (0.094)	-0.515 (0.101)	-0.481 (0.112)	-0.444 (0.121)
High grade aGvHD		-1.599 (0.167)	-1.225 (0.175)	-1.083 (0.195)	-1.201 (0.231)
	NRM				
Intercept	-1.241 (0.050)	-2.019 (0.069)	-2.352 (0.086)	-2.604 (0.100)	-2.751 (0.113)
Year of SCT	-0.620 (0.161)	-0.160 (0.208)	-0.040 (0.253)	0.020 (0.300)	-0.030 (0.339)
Risk score					
Low risk					
Medium risk	0.529 (0.066)	0.436 (0.083)	0.353 (0.101)	0.429 (0.120)	0.365 (0.137)
High risk	1.090 (0.101)	0.846 (0.134)	0.772 (0.173)	0.796 (0.211)	0.802 (0.240)
Low grade aGvHD		0.657 (0.093)	0.748 (0.109)	0.611 (0.130)	0.595 (0.147)
High grade aGvHD		1.830 (0.106)	1.538 (0.130)	1.272 (0.157)	1.023 (0.189)

Table 5.3: Estimated regression parameters of the landmark super model for relapse and non-relapse mortality.

Covariate	Relapse		NRM	
	$\widehat{\beta}$	SE($\widehat{\beta}$)	$\widehat{\beta}$	SE($\widehat{\beta}$)
Intercept				
Constant	-1.160	0.027	-1.156	0.029
s	0.839	0.126	-3.603	0.165
s^2	-1.072	0.129	2.080	0.183
Year of transplantation				
Constant	0.530	0.126	-0.591	0.125
s	-1.165	0.657	1.604	0.718
s^2	0.553	0.678	-1.079	0.766
Risk score				
Low risk				
Medium risk	0.166	0.022	0.431	0.025
High risk	0.725	0.039	0.880	0.042
Low grade aGvHD				
Constant	-0.490	0.030	0.168	0.102
s			2.032	0.461
s^2			-1.738	0.436
High grade aGvHD				
Constant	-1.305	0.054	1.916	0.129
s			-0.416	0.579
s^2			-0.524	0.545

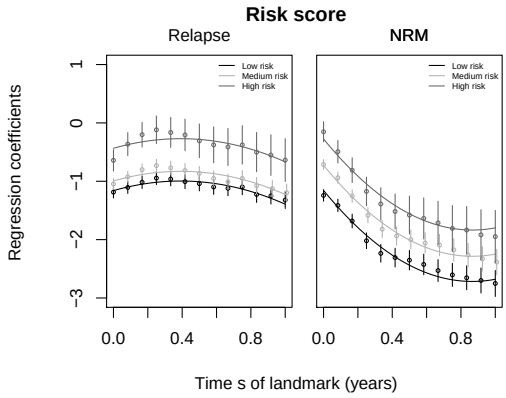


Figure 5.2: Regression coefficients of intercept (Low risk) and of intercept plus effects of medium risk and high risk, together with associated 95% confidence intervals implied by the stratified landmark models. The solid lines show the same regression coefficients implied by the landmark supermodel of Section 5.3.3.

included in the models: for relapse the effect decreases from 0.530 at $s = 0$ to almost 0 at $s = 1$, while for NRM it increases from -0.591 at $s = 0$ to almost 0 at $s = 1$. Figure 5.7 presents regression coefficients of year of SCT based on the separate models (error bars) and on the supermodel (solid lines) for the two competing risks and the estimated 95% confidence intervals; it suggests that the two approaches agree on quantifying the trend of this covariate. Moreover, approximately the same behavior was found in the supermodel on cause-specific hazards in Nicolaie et al. (2013a). The linear and quadratic terms of the risk score effects were non-significant for both causes of failure, and therefore not included in the models (see Figure 5.2). For relapse, interactions between occurrence of aGvHD and landmark time points were not found to be significant, and therefore not included in the model. In contrast, Wald test showed for NRM significant interactions between $Z_{\text{low}}(s)$ and $Z_{\text{high}}(s)$ and landmark time points s and therefore they are kept in the model. In Figures 5.8 and 5.9 we plot the regression coefficients of $Z_{\text{low}}(s)$ and $Z_{\text{high}}(s)$ based on the separate models (error bars) and the supermodel (solid lines) for the two competing risks. Even though the effects of low grade and high grade aGvHD may vary across landmarks, occurrence of aGvHD before s consistently decreases the conditional probability of relapse within 5 years and increases the conditional probability of NRM within 5 years, as found in Nicolaie et al. (2013a). Increasing the number of landmark time points in the prediction interval did not have noteworthy effects (data not shown): it resulted in comparable, slightly more accurate estimates of the

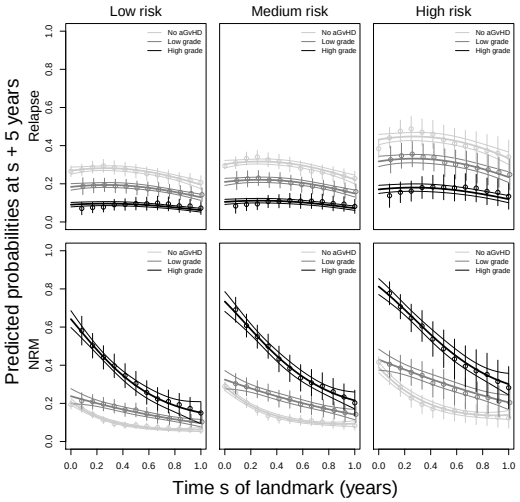


Figure 5.3: Dynamic prediction of probability of relapse and NRM at $s+5$ years and associated 95% confidence intervals for different landmark time points s, for a patient transplanted in 2003. The error bars are based on the separate models and the solid lines are based on the supermodel.

regression coefficients of the landmark supermodel, implying slightly narrower confidence intervals of the dynamic prediction estimates.

In Figure 5.3 we show the estimated prediction probabilities of experiencing relapse and NRM, respectively, before 5 years after the prediction time points for all prediction time points between SCT and 1 year after SCT, resulting from the separate models (error bars) and from the supermodel (smoothed lines) with pointwise 95% confidence intervals using (5.9) and (5.15), respectively, for a patient transplanted in 2003. The two approaches agree with respect to the estimated dynamic prediction probabilities. However, the pointwise confidence intervals for prediction from the supermodel are narrower than those based on the separate models, which can be explained by a higher degree of accuracy obtained by combining in one model information from all individual landmark data sets. Looking simultaneously at all the plots in Figure 5.3, we see that, as expected, the higher the risk score, the higher the conditional cumulative incidence of both relapse and NRM. Figure 5.3 clearly shows that (especially high grade) aGvHD increases the conditional cumulative incidence of NRM, while it decreases the conditional cumulative incidence of time to relapse. The gradual decrease with later s of the conditional cumulative incidences is a natural consequence of conditional probabilities estimated solely on individuals event-free at s.

Example R code implementing our methods is available in Appendix E. The majority of the computing time of our approach is in the computation of the dynamic pseudo-observations (few tens of seconds in our application). Once these have been computed both model fitting and prediction can be done in a matter of seconds.

5.4 Discussion

In this paper we proposed an alternative approach to dynamic prediction of time-to-event data with competing risks using dynamic pseudo-observations. Our aim is to directly estimate the conditional probability of failing due to a given cause within a given time window ($\leq s + w$) conditionally given failure-free at a prediction point $s \geq 0$. We used the dynamic pseudo-observations as extensions of "static" pseudo-observations to obtain landmark supermodels for these probabilities. The use of dynamic pseudo-observations for dynamic prediction of (conditional) cumulative incidences has a number of practical advantages. The first is that the approach can be implemented using standard statistical software. After having obtained the dynamic pseudo-observations, regression estimates may be obtained with relative ease using generalized estimation equations (GEE); standard statistical software like PROC GENMOD in SAS or the `geepack` package (Højsgaard et al., 2006) in R may be used for fitting the GEE. The dynamic pseudo-observations may be obtained in R using the packages `dynpred` (van Houwelingen and Putter, 2012) and `pseudo` (Klein et al., 2008). The second practical advantage of our approach is that it can deal with both internal (endogenous) and external (exogenous) time-dependent covariates; see Kalbfleisch and Prentice (2002, Chapter 6) and Cortese and Andersen (2010) for a discussion on internal versus external time-dependent covariates. Standard modeling procedures like the Fine-Gray model do not allow internal time-dependent covariates for the prediction of cumulative incidences, see for instance Latouche et al. (2005); Beyersmann and Schumacher (2008). In contrast, landmarking can incorporate internal time-dependent covariates for prediction because it avoids joint modeling of the internal time-dependent covariates and the endpoints (van Houwelingen, 2007; van Houwelingen and Putter, 2008; Cortese and Andersen, 2010). A third advantage of the current approach in comparison with landmark supermodels based on cause-specific hazards (Nicolaie et al., 2013a) is that the dynamic pseudo-observations may be used to directly model the conditional cumulative incidences. As a result, the regression models allow a direct interpretation in terms of the cumulative incidence(s) of the event(s) of interest. In our application the two approaches led to comparable estimated prediction probabilities.

An important distinction between the traditional use of (static) pseudo ob-

servations and our use of dynamic pseudo-observations is that we only use a *single* dynamic pseudo-observation per subject for each landmark time point. Traditionally, for a fixed prediction time point ($s = 0$) pseudo-observations for $\mathbf{1}(T \leq t, D = j)$ are calculated and used for several time points t, either at a grid (Klein and Andersen, 2005) or at all event time points. These approaches exploit the proportional hazards assumption on either the logit or the cloglog scale, the latter being equivalent with the Fine and Gray (1999) model. In our dynamic prediction setting it would also be possible to use several prediction widths rather than a single one. Different models could be fitted for each width separately. Combining these into a single supermodel is also possible, but would require the proportional hazards assumption. By using dynamic pseudo-observations at a single time point $s+w$ for each landmark time point s we avoid the proportional hazards assumption, which makes our approach robust to deviations from such assumptions. The disadvantage of using only a single dynamic pseudo-observation at each landmark time point is a possible loss of efficiency if the proportional hazards assumption holds.

An issue with the present pseudo-observations approach is that the correlation structure of dynamic pseudo-observations is ignored in the working correlation used in the supermodels; an independence working correlation yields consistent estimators, but may lead to loss of efficiency. Simulation studies performed in Andersen and Klein (2007) suggest that efficiency is indeed influenced by the choice of correlation structure in the working correlation matrix, but the influence is not great. We are currently investigating this issue in a simulation study. It may be possible to account for correlation, but additional work is needed to determine scenarios where consistency of estimators may be achieved.

A cautionary remark is in order when the dynamic prediction model is to be applied to an external population. If in the new population the probability of the competing risk is much different than in the original population on which the dynamic prediction model was developed, then this could distort the dynamic predictions for the event of interest. A common situation where this could happen is when interest is in disease-specific survival and where death due to other causes is a competing risk. If the dynamic prediction model is developed in a young population, for instance, then application in an older population with higher risk of death due to other causes (old age) may be problematic because as a result one would expect a lower probability of the event of interest. We don't expect such issues to play a role in our specific application because the competing risk of relapse, non-relapse mortality, is mainly disease-specific, consisting of direct treatment-related mortality and mortality due to aGvHD.

Appendix A: Specification of models for complete data

Suppose that the data are complete, that is, censoring does not occur. In this case, the indicators $Y_i(s) = \mathbf{1}(T_i \leq s + w, D_i = j)$ are observed for all subjects i in $\mathcal{L}_s$. The observed values of $Y_i(s)$, denoted by $y_i(s)$, form a sample of n_s independent binomial variables. We have $E\{Y_i(s)|\mathbf{Z}_i(s), T_i > s\} = P\{Y_i(s) = 1|\mathbf{Z}_i(s), T_i > s\}$ and let

$$\mu_i(s) = E\{Y_i(s)|\mathbf{Z}_i(s), T_i > s\}.$$

We postulate a generalized linear model on the binomial expectations $\mu_i(s)$ of the form

$$g\{\mu_i(s)\} = \boldsymbol{\beta}^\top(s)\mathbf{Z}_i^*(s) \ ,$$

for a given link function g, where $\boldsymbol{\beta}(s) = \{\beta_0(s), \beta_1(s), \ldots, \beta_p(s)\}$ and $\mathbf{Z}^*(s) = \{1, \ \mathbf{Z}(s)^\top\}^\top$, so that $\beta_0(s)$ stands for the intercept. In the following, since we consider a fixed value of s we shall suppress the dependence on s of the notation.

For the case of complete data the analysis is completely standard and follows a generalized linear model (GLM). The regression parameters $\boldsymbol{\beta} = \boldsymbol{\beta}(s)$ can be estimated by a maximum likelihood approach; the likelihood would be given by the product of binomial probabilities

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n_s} p_i^{y_i} \cdot (1 - p_i)^{1-y_i} \ , \tag{5.16}$$

with $p_i = p_i(\boldsymbol{\beta}) = P(Y_i = 1|\mathbf{Z}_i) = E(Y_i|\mathbf{Z}_i) = \mu_i(\boldsymbol{\beta}) =: g^{-1}(\boldsymbol{\beta}^\top\mathbf{Z}_i^*)$, where g^{-1} stands for the inverse of g. Using $\mu_i(\boldsymbol{\beta})$ instead of p_i, the log-likelihood $\ell(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$ is given by

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n_s} \left[y_i \log \mu_i(\boldsymbol{\beta}) + (1 - y_i) \log\{1 - \mu_i(\boldsymbol{\beta})\} \right] \ ,$$

and the score equations by

$$\begin{aligned} \sum_{i=1}^{n_s} \frac{\partial}{\partial \boldsymbol{\beta}} \ell(\boldsymbol{\beta}) &= \sum_{i=1}^{n_s} \frac{y_i}{\mu_i} \cdot \frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \frac{1 - y_i}{1 - \mu_i} \cdot \frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ &= \sum_{i=1}^{n_s} \frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \cdot \frac{1}{\mu_i(1 - \mu_i)} \cdot (y_i - \mu_i) = 0 \ , \end{aligned} \tag{5.17}$$

where $\frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \{\frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1}, \ldots, \frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_p}\}^\top$ is the vector of partial derivatives of $\mu_i(\boldsymbol{\beta})$

with respect to $\boldsymbol{\beta}$.

The asymptotic variance of $\widehat{\boldsymbol{\beta}}$, the solution to (5.17), can be estimated by the inverse of the matrix

$$\sum_{i=1}^{n_s} \Big\{ \frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big\}^\top \cdot \frac{1}{\mu_i(1-\mu_i)} \cdot \frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}.$$

It is seen that equation (5.17) indeed follows the score equations of a GLM on a binary variable, that is

$$\sum_{i=1}^{n_s} \frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \cdot \frac{1}{\mathrm{var}(y_i)} \cdot \big(y_i - \mu_i\big) = 0 \ , \tag{5.18}$$

where $\frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ are the rows of the matrix $\mathrm{d}\mu = \big\{ \frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_r} \big\}_{n_s \times (p+1)}$. With specific choices for the link function g, equation (5.18) may be simplified to

$$\sum_{i=1}^{n_s} \mathbf{Z}_i^* \cdot \big(y_i - \mu_i\big) = 0 \ ,$$

for the logit link function, $g(x) = \log \frac{x}{1-x}$, or to

$$\sum_{i=1}^{n_s} \mathbf{Z}_i^* \cdot \frac{\log(1-\mu_i)}{\mu_i} \cdot \big(y_i - \mu_i\big) = 0 \ ,$$

for the cloglog link function, $g(x) = \log\{-\log(1-x)\}$.

Appendix B: Proofs of Propositions

Proof of Proposition 5.2.1

The proof follows directly from Lemma 2 of Graw et al. (2009). Note that assumption $(C2)$ guarantees $n_s \to \infty$ as $n \to \infty$.

Proof of Proposition 5.2.2

The proof of consistency relies on similar arguments as those in Theorem 2 of Graw et al. (2009). The asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ follows from the asymptotic unbiasedness of $U_i(\boldsymbol{\beta})$ in (5.5) and the results of Liang and Zeger (1986).

Proof of Proposition 5.2.3

The asymptotic behaviour of $\widehat{F}_j(s+w|\ s,\ \widetilde{\mathbf{Z}}(s))$ could be derived by means of the delta-method using the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$.

Proof of Proposition 5.2.4

Consider the estimating equations

$$\tilde{\mathbf{U}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \cdot \mathbf{V}_i^{-1} \cdot (\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0 \ , \tag{5.19}$$

with $\boldsymbol{\mu}_i$ and $\mathbf{V}_i$ as defined in (5.12), and $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{i,l_i})$ is the vector with true (possibly unobservable) outcomes $Y_{ik} = \mathbf{1}(T_i \leq s_k + w, D_i = j)$. By the assumption that the regression models are correctly specified, we have $E\{Y_{ik} \,|\, Z_i(s_k), T_i > s_k\} = \mu_{ik}$. Some care should be taken here because of the fact that μ_{ik} are not only functions of $\boldsymbol{\beta}_l$, but of s as well via (5.11). However, $\mathbf{h}_l(s)$, as deterministic functions of s contribute to the asymptotic behavior of $\mathbf{U}_i(\boldsymbol{\beta})$ only as scaling factors. In the absence of censoring, by including only the Y_{ik}'s for which $T_i > s_k$, we are implicitly fitting

$$\sum_{i=1}^{n} \sum_{k=1}^{K} \frac{\partial \mu_{ik}}{\partial \boldsymbol{\beta}} \cdot \frac{1}{\mu_{ik}(1-\mu_{ik})} \cdot A_{ik} \cdot (Y_{ik} - \mu_{ik}) = 0 \ ,$$

where $A_{ik} = 1$ if $T_i > s_k$ and 0 otherwise. This coincides with the equation $U(\beta^A) = 0$ in the middle of page 247 of Kurland and Heagerty (2005), for the special case of the logit link function (see Equation (5.6)), for which Kurland and Heagerty (2005) argue that the estimating equations are unbiased. We have additional missingness due to censoring, but because of condition (C1) these (potential) outcomes are missing completely at random, and hence the estimation equations (5.19) are still asymptotically unbiased. Finally, replacing $\mathbf{Y}_i$ by $\widehat{\boldsymbol{\theta}}_i$ will retain the asymptotic unbiasedness of (5.12), by Proposition 5.2.1. This concludes the proof of Proposition 5.2.4.

Proof of Proposition 5.2.5

The asymptotic behaviour of $\widehat{F}_j(s+w|s,\ \widetilde{\mathbf{Z}}(s))$ follows from the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ using the delta-method. Indeed,

$$\begin{aligned}\widehat{\text{var}}\{\widehat{F}_j(s+w|\ s,\widetilde{\mathbf{Z}}(s))\} &= \widehat{\text{var}}[g^{-1}\{\widehat{\boldsymbol{\beta}}(s)^\top\widetilde{\mathbf{Z}}^*(s)\}] \\ &= \Big[\frac{d}{d\widehat{\boldsymbol{\beta}}}g^{-1}\{\widehat{\boldsymbol{\beta}}(s)^\top\widetilde{\mathbf{Z}}^*(s)\}\Big]^\top\cdot\widehat{\text{var}}(\widehat{\boldsymbol{\beta}}) \\ &\cdot \Big[\frac{d}{d\widehat{\boldsymbol{\beta}}}g^{-1}\{\widehat{\boldsymbol{\beta}}(s)^\top\widetilde{\mathbf{Z}}^*(s)\}\Big]\end{aligned}$$

The expression (5.15) follows immediately if we notice that

$$\begin{aligned}\frac{d}{d\widehat{\boldsymbol{\beta}}}g^{-1}\{\widehat{\boldsymbol{\beta}}(s)^\top\widetilde{\mathbf{Z}}^*(s)\} &= \Big\{\frac{\mathrm{d}g^{-1}(x)}{\mathrm{d}x}\Big\}_{|x=\widehat{\boldsymbol{\beta}}(s)^\top\widetilde{\mathbf{z}}^*(s)}\cdot\frac{\partial}{\partial\widehat{\boldsymbol{\beta}}}\{\widehat{\boldsymbol{\beta}}(s)^\top\widetilde{\mathbf{Z}}^*(s)\} \\ &= \Big\{\frac{\mathrm{d}g^{-1}(x)}{\mathrm{d}x}\Big\}_{|x=\widehat{\boldsymbol{\beta}}(s)^\top\widetilde{\mathbf{z}}^*(s)}\cdot\widetilde{\mathbf{Z}}^*(s)\cdot H(s).\end{aligned}$$

Appendix C: Illustration of properties of dynamic pseudo-observations

In Figure 5.4 we illustrate what the dynamic pseudo-observations may look like for the entire grid of landmark time points, for four patients chosen from the data. "Patient A" (first column) was censored at 5.88 years, "patient B" (second column) experienced relapse at 5.07 years, "patient C" (third column) experienced non-relapse mortality at 5.50 years, while "patient D" was censored at 0.76 years. The upper row presents the dynamic pseudo-observations for relapse, $\widehat{\theta}^1_{is}$, for these four patients, while the bottom row presents the dynamic pseudo-observations for non-relapse mortality, $\widehat{\theta}^2_{is}$. Dynamic pseudo-observations are defined at all prediction time points for the first three individuals, because their event takes place after the last landmark time point, while for the last patient we only have 10 dynamic pseudo-observations for both endpoints relapse and non-relapse mortality. The main message of these figures is that $\widehat{\theta}^j_{is}$ resembles $\mathbf{1}(T_i \leq s+w, D_i = j)$ (represented by the dotted lines in Figure 5.4); for patient A, for instance, the pseudo-observations for relapse jump from approximately 0 to approximately 1 at the first landmark time point s for which relapse took place before $s+w$. We observe that for individuals at risk at $s+w$ the dynamic pseudo-observations tend to be negative, with an increasing trend with increasing s; this phenomenon is explained by the fact that omitting the individual i in $\widehat{F}^{(-i)}_j$, irrespective of their status (censoring or death), lowers the risk set causing the discrepancy between $\widehat{F}_j$ and $\widehat{F}^{(-i)}_j$ to increase with increasing s. In case of patient failing, their dynamic pseudo-observations corresponding to the cause in question jump above 1 while the dynamic pseudo-observations corresponding to the competing

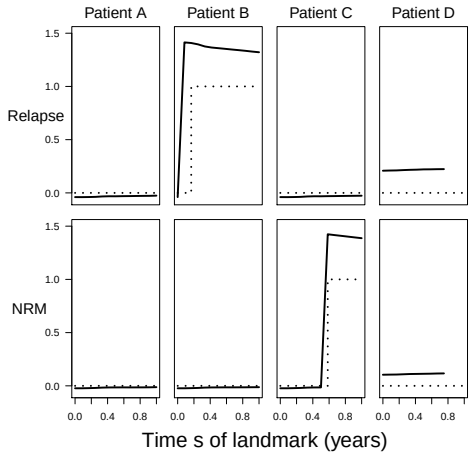


Figure 5.4: Dynamic pseudo-observations for relapse and non-relapse mortality for four example patients. "Patient A" (first column) was censored at 5.88 years, "patient B" (second column) experienced relapse at 5.07 years, "patient C" (third column) experienced non-relapse mortality at 5.50 years, while "patient D" was censored at 0.76 years.

cause remain negative, increasing with increasing s. In case of an early censored patient, their dynamic pseudo-observations increase at the very first event time point which corresponds to a failure due to the cause in question (see "patient D") succeeding their censoring time point. These trends were also observed in Andersen and Perme (2010).

Figures 5.5 and 5.6 show some of the variation and co-variation of the dynamic pseudo-observations $\widehat{\theta}^1_{is}$ and $\widehat{\theta}^2_{is}$, respectively, at a coarser selection of landmark time points: 0 years, 0.25 years, 0.5 years, 0.75 years and 1 year. The closer in time are the landmark time points, the stronger the correlation between two individual sets of dynamic pseudo-observations is, as seen in the upper-diagonal plots. For $s^{'} < s$, the correlation of the indicators $\mathbf{1}(T_i \leq s^{'} + w, D = j)$ and $\mathbf{1}(T_i \leq s + w, D = j)$ in $\mathcal{L}_s$ (i.e. given $T_i > s$) is given by

$$\sqrt{\frac{F_j(s' + w|s)}{1 - F_j(s' + w|s)} \cdot \frac{1 - F_j(s + w|s)}{F_j(s + w|s)}}.$$

This implies that for $s^{'}$ close to s,

$$\text{corr}(\widehat{\theta}^{j}_{is^{'}}, \widehat{\theta}^{j}_{is}) \approx 1 - \frac{1}{2}(s - s^{'}) \frac{F^{'}_{j}(s+w|s)}{F_j(s+w|s)\{1 - F_j(s+w|s)\}} \ ,$$

where $F^{'}(s+w|s)$ stands for the derivative of $F(t|s)$ with respect to t, evaluated at $t = s + w$. The diagonal plots show histograms of the dynamic pseudo-observations for each of the selected landmark time points. The subdiagonal plots show scatter-plots of any two sets of the selected dynamic pseudo-observations; the superdiagonal plots show some correlations. Most striking are the points visible in the upper-left corner of each of the subdiagonal plots. For a particular subdiagonal plot corresponding to two landmark time points s and $s^{'}$, these are individuals at risk at both landmark time points and failing due to the event type in question at a time point between the two prediction time points $s + w$ and $s^{'} + w$.

Appendix D: Selected figures

Appendix E: Example R code

The following code is a template for implementing our method in the R software; the statistical models are those used in our paper for the analysis of the EBMT data (see Section 5.2.3).

```
library(pseudo)
library(geepack)

########################################################
###     building of the stacked landmark data set    ###
########################################################

LMs  <- seq(0, 12, by =1) # time in the data is in months
LMdata <-  NULL
for (i in seq(along = LMs)) {
```

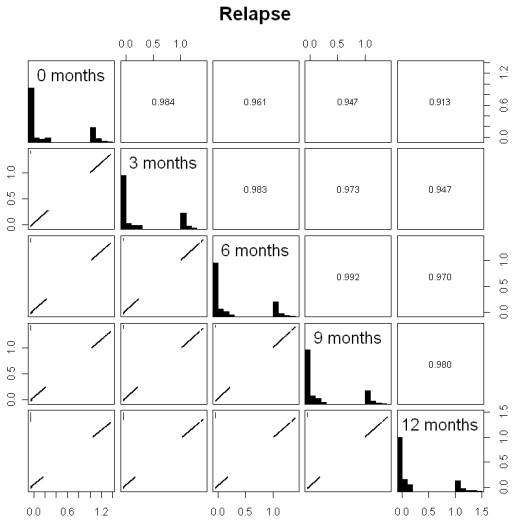


Figure 5.5: Scatterplot of dynamic pseudo-observations associated with the cumulative incidence of relapse. Diagonals show histograms of dynamic pseudo-observations; the numbers in the upper-diagonal plots are the correlations between the dynamic pseudo-observations at different landmark time points.

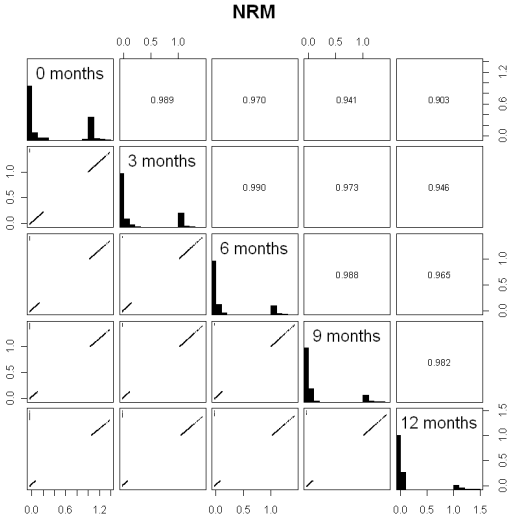


Figure 5.6: Scatterplot of pseudo-observations associated with the cumulative incidence of non-relapse mortality. Diagonals show histograms of dynamic pseudo-observations; the numbers in the upper-diagonal plots are the correlations between the dynamic pseudo-observations at different landmark time points.

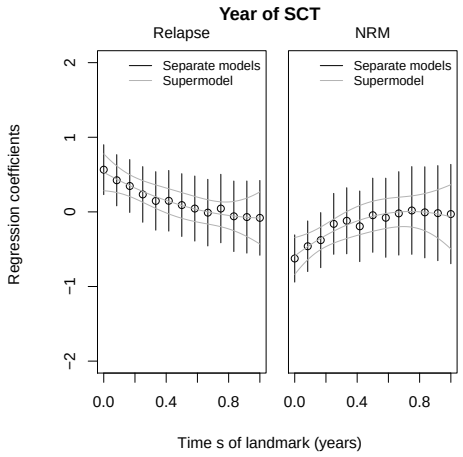


Figure 5.7: Regression coefficients of year of SCT (centered at 2000, scaled by factor 10) and associated 95% confidence intervals implied by the separate landmark models (error bars) and by the landmark supermodel (solid lines).

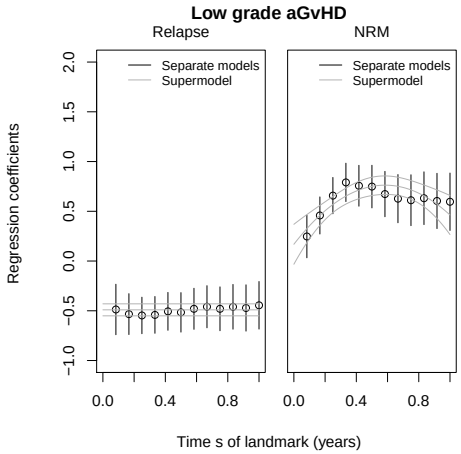


Figure 5.8: Regression coefficients of the low grade aGvHD and associated 95% confidence intervals implied by the separate landmark models (error bars) and by the landmark supermodel (solid lines).

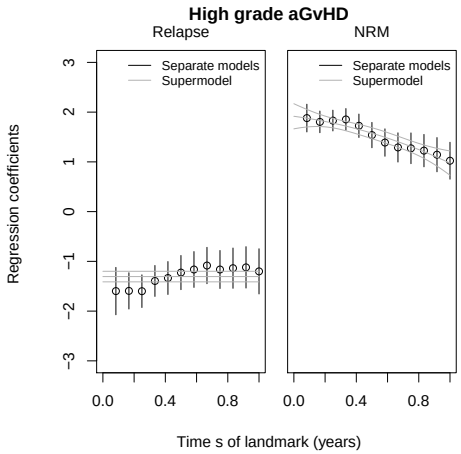


Figure 5.9: Regression coefficients of the high grade aGvHD and associated 95% confidence intervals implied by the separate landmark models (error bars) and by the landmark supermodel (solid lines).

```
 LM    <- LMs[i] # current landmark time point
 datai <- data[data$ci > LM, ] # select subjects at risk
 # low and high grade aGvHD at landmark
 #(a.t is time, a.sc grade, 1=low, 2=high)
 #low grade aGvHD:
 alo   <- as.numeric(datai$a.t <= LM)*as.numeric(datai$a.sc == 1)
 #high grade aGvHD:
 ahi   <- as.numeric(datai$a.t <= LM)*as.numeric(datai$a.sc == 2)
 # pseudo-observations are calculated using pseudoci from pseudo package
 dfri  <- data.frame(id = datai$id, alo = alo, ahi = ahi,
          year = datai$year/10, score = datai$score,
          # time to event
          time = datai$ci,
          # type of event
          status = datai$ci_s,
          pse1 = pseudoci(datai$ci, datai$ci_s,
                 tmax = 60 + LM)$pseudo$cause1,
          pse2 = pseudoci(datai$ci, datai$ci_s,
                 tmax = 60 + LM)$pseudo$cause2,
          LM = rep(LM, nrow(datai)))
 LMdata  <- rbind(LMdata, dfri)
}

############################################################
###   fitting separate landmark models              ####
############################################################

# do this for (i in seq(along = LMs))

datap <- LMdata[LMdata$LM == i, ]
fit <- geese(pse1 ~ year + score + alo + ahi, data = datap,
             id = id, scale.fix = TRUE, family = gaussian,
             jack = TRUE, mean.link = "logit",
             corstr = "independence", var = "binomial")

############################################################
###  fitting the landmark supermodel                ####
############################################################

############################################################
###    1. prepare the regression coefficients        ###
```

```
##########################################################

f0    <- function(t) 1
f1    <- function(t) t/12 # change of time scale to years
f2    <- function(t) (t/12)^2

LMdata$year.t0 <- LMdata$year
LMdata$year.t1 <- LMdata$year * f1(LMdata$LM) #interaction with s
LMdata$year.t2 <- LMdata$year * f2(LMdata$LM) #interaction with s^2

LMdata$sc1 <- as.numeric(LMdata$score == 2)
LMdata$sc2 <- as.numeric(LMdata$score == 3)

LMdata$alo.t0 <- LMdata$alo
LMdata$alo.t1 <- LMdata$alo * f1(LMdata$LM)
LMdata$alo.t2 <- LMdata$alo * f2(LMdata$LM)

LMdata$ahi.t0 <- LMdata$ahi
LMdata$ahi.t1 <- LMdata$ahi * f1(LMdata$LM)
LMdata$ahi.t2 <- LMdata$ahi * f2(LMdata$LM)

LMdata$LM1 <- f1(LMdata$LM)
LMdata$LM2 <- f2(LMdata$LM)

##########################################################
#### 2. fitting the landmark super models           ####
##########################################################

# Final model for Rel (selection procedure not shown)
fit1 <- geese(pse1 ~ year.t0 + year.t1 + year.t2
                        + sc1 + sc2
                        + alo.t0  + ahi.t0
                        + LM1 + LM2 , data = LMdata, id = id,
                        scale.fix = TRUE, family = gaussian,
                        jack = TRUE, mean.link = "logit",
                        corstr = "independence", var = "binomial")
# Final model for NRM
fit2 <- geese(pse2 ~ year.t0 + year.t1 + year.t2
                        + sc1 + sc2
                        + alo.t0 + alo.t1 + alo.t2
                        + ahi.t0 + ahi.t1 + ahi.t2
```

```
                        + LM1 + LM2 , data = LMdata, id = id,
                        scale.fix = TRUE, family = gaussian,
                        jack = TRUE, mean.link = "logit",
                        corstr = "independence", var = "binomial")

##########################################################
#### 3. estimate dynamic predictions (supermodels)  ####
##########################################################

# Time points at which we want predictions
tt <- seq(0,12,length = 101)

expit <- function(x) exp(x)/(1 + exp(x))

# Dynamic prediction for relapse
# example for patient with medium risk score,
# transplanted in 2003 (standardized value 0.3), high grade aGvHD

coef <- fit1$beta
# will contain the dynamic predictions:
dynpred.Rel <- rep(NA, length(tt))

for (i in 1:length(tt)) {
  linpred <- coef[["(Intercept)"]] + coef[["year.t0"]]*0.3
              + coef[["year.t1"]]*0.3 * f1(tt[i])
              + coef[["year.t2"]]*0.3 * f2(tt[i])
              + coef[["sc1"]] + coef[["ahi.t0"]]
              + coef[["LM1"]]*f1(tt[i]) + coef[["LM2"]]*f2(tt[i])
  dynpred.Rel[i] <- expit(linpred)
}

# Dynamic prediction for NRM, same patient

coef <- fit2$beta
# will contain the dynamic predictions:
dynpred.NRM <- rep(NA, length(tt))

for (i in 1:length(tt)) {
  linpred <- coef[["(Intercept)"]] + coef[["year.t0"]]*0.3
             + coef[["year.t1"]]*0.3 * f1(tt[i])
             + coef[["year.t2"]]*0.3 * f2(tt[i])
```

```
              + coef[["sc1"]] + coef[["ahi.t0"]]
              + coef[["ahi.t1"]]*f1(tt[i]) + coef[["ahi.t2"]]*f2(tt[i])
              + coef[["LM1"]]*f1(tt[i]) + coef[["LM2"]]*f2(tt[i])
  dynpred.NRM[i] <- expit(linpred)
}
```

6

Comparison of dynamic prediction models

Abstract

In this paper, we consider the problem of prediction accuracy in survival data with competing risks when the target of estimation is the dynamic prediction probabilities of the terminal events. We evaluate the properties of a number of methods for dynamic prediction in competing risks using simulated data. We generate several scenarios for the transition intensities leading to multi-state models with multiple endpoints and intermediate states, under the Markov assumption or subjected to departures from it. This technique conveniently mirrors competing risks modeling in the presence of time-dependent covariates. We compare modeling approaches which either are based on comprehensive modeling or which are focused directly on the dynamic prediction probabilities.

6.1 Introduction

Dynamic prediction models for survival data with competing risks have recently gained growing interest in terms of theoretical developments (Cortese and Andersen, 2010; Parast et al., 2011; Nicolaie et al., 2013a,b; Cortese et al., 2013). The task of dynamic prediction is challenging because (1) the presence of time-dependent covariates implies a complicated mathematical form of the relation

among variables and dynamic prediction probabilities; (2) different aspects of covariates, which are relevant to the response, might destroy the Markov assumption typically used in modeling. In this work, we provide a comparison among several approaches which yield dynamic prediction probabilities of competing terminal events. The first approach is based on a Markov multi-state model which includes, besides the competing terminal events, intermediate states corresponding to the different stages in the development of the time-dependent covariates. The remaining approaches use the landmark method (van Houwelingen, 2007; van Houwelingen and Putter, 2012); they share the specific feature of being directly targeted to the modeling of the dynamic prediction probabilities. The first landmark method consists of the approach of Nicolaie et al. (2013a) based on modeling the cause-specific hazards and the second landmark method consists of the approach of Nicolaie et al. (2013b) based on modeling dynamic pseudo-observations of the cause-specific cumulative incidences, each of them using the current information at different landmark time points. We consider two types of such landmark models: separate landmark models, obtained from separate landmark data sets, and supermodels, obtained by combing information from different landmark data sets. Related work is comprised in a previous paper of Cortese et al. (2013), which evaluated dynamic prediction models for cause-specific cumulative incidences in the presence of an internal time-dependent covariate, including separate landmark models for cause-specific hazards or for subdistribution hazards.

The aim of this paper is to evaluate among these competing risks models which can be applied best, in the framework of Markov or non-Markov multi-state settings, to evaluate the dynamic prediction probability of experiencing a terminal event at a certain point in time using all available information until that point.

Under the Markov assumption, it is natural to expect that a joint analysis of the time-dependent covariates and survival data, as the Markov multi-state model does, would provide more efficient estimates. When the Markov assumption does not hold it is natural to expect that the pragmatic and robust approaches based on landmarking have lower bias at the cost of a higher variability when compared to misspecified models, like a Markovian multi-state approach. The difference is due to the direct modeling which is robust against departures from the Markov assumption. In this paper, we will show that for dynamic prediction it is convenient to postulate a model on the probabilities of interest and to collect only the necessary information to estimation.

Our paper is organized as follows: in Section 6.2 the simulation scenarios are introduced. In Section 6.3 the dynamic prediction problem is introduced. Section 6.4 presents different approaches to this problem. Section 6.5 presents the simulation results. Final comments and directions for future research are

given in Section 6.6.

6.2 Data generation

We simulated data for $n = 500, 1000, 2500$ individuals, each of whom can fail from one of two causes. Individuals are followed over a period of maximally 14 years; random right-censoring, which is independent of survival time, occurred uniformly between 5 years and 14 years. Assume $\mathbf{Z} = (Z_1, Z_2)$ is a vector of two binary baseline covariates chosen such that $P(Z_1 = 0) = 0.7$ and $P(Z_2 = 0) = 0.5$. We generated data as random samples drawn from a multi-state model $\mathbf{X}(t)_{t\geq 0}$, as shown in Figure 6.1, whose state space comprises 4 states: an initial state, denoted by state 1, an intermediate state, denoted by 2, and two terminal events which act as competing risks, denoted by states 3 and 4, respectively. In the multi-state model there are five transitions possible across these states, whose transition intensities, defined by

$$\lambda_{gh}(t|\ \mathcal{F}_{t_-},\ \mathbf{Z}) = \lim_{\Delta t\to 0} \frac{P(X(t+\Delta t) = h|X(t) = g, \mathbf{Z}, \mathcal{F}_{t_-})}{\Delta t},$$

are modeled by

$$\lambda_{gh}(t|\ \mathcal{F}_{t_-},\ \mathbf{Z}) = \lambda_{gh,0}(t) \exp(\boldsymbol{\beta}_{gh}(t)^\top \mathbf{Z})\ , \tag{6.1}$$

where $\lambda_{gh,0}(t)$ is the baseline transition intensity of the transition from g to h, $\boldsymbol{\beta}_{gh}(t) = (\boldsymbol{\beta}_{gh,1}(t), \boldsymbol{\beta}_{gh,2}(t))$ is a vector of (possibly, time-varying) transition-specific regression parameters for the transition from g to h, for $g, h \in \{1, 2, 3, 4\}$, and $\mathcal{F}_{t_-}$ stands for the process history up to time t, that is $\mathcal{F}_{t_-} = \{\mathbf{X}(u) : 0 \leq u < t\}$. Later (see (6.2)) we will consider the (non-Markov) case where λ_{23} and λ_{24} depend on $\mathcal{F}_{t_-}$.

We chose piecewise constant baseline transition-specific intensities with a cut-off point at $t = 5$; using matrix notation, $\lambda_{gh,0}(t)$ is the (g, h) element of the baseline transition intensity matrix denoted by $Q(t)$, where

$$Q(t) \equiv Q_1 = \begin{pmatrix} -0.35 & 0.15 & 0.15 & 0.1 \\ 0 & -0.35 & 0.25 & 0.1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

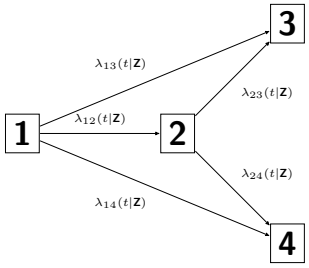


Figure 6.1: The multi-state model used in the data generation.

for $t \in [0, 5)$ and

$$Q(t) \equiv Q_2 = \begin{pmatrix} -0.15 & 0.1 & 0.05 & 0.1 \\ 0 & -0.3 & 0.25 & 0.15 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

for $t \in [5, 14]$. The diagonal terms $Q_{gg}(t)$ are defined as $Q_{gg}(t) = -\sum_{h \neq g} Q_{gh}(t)$. The choice of $Q(t)$ is loosely based on the European Organization for Research and Treatment of Cancer (EORTC) trial 10854 breast cancer data (van der Hage (2001)), also used in Putter et al. (2006) and in van Houwelingen and Putter (2012), with local recurrence as intermediate event (state 2) and distant metastasis and death as endpoints (states 3 and 4, respectively).

In terms of regression coefficients, the effect of Z_1 is taken to be constant, equal to 0.75 for each transition, that is

$$\beta_{gh,1}(t) = 0.75 \text{ for all transitions } g \to h \ ,$$

while the effect of Z_2, also taken to be the same for all transitions, is possibly time-varying, that is

$$\beta_{gh,2}(t) = \begin{cases} c_1 & \text{if } t \in [0, 2.5) \\ c_2 & \text{if } t \in [2.5, 14] \end{cases},$$

for all possible combinations (c_1, c_2), where we take $c_1 = \pm 0.5$, and for $c_1 = -0.5$ we take $c_2 \in \{-1, -0.5, 0\}$ and for $c_1 = 0.5$ we take $c_2 \in \{0, 0.5, 1\}$.

Data from different individuals are supposed to be independent.

Under this setting, we deal with a Markov multi-state model, which we refer to as the "true" Markov model and which will be used to generate various data sets in our simulations; note that its initial distribution is degenerate in state 1. Notably here, several transition probabilities of the true Markov model are referential because they will be used to assess the accuracy of our methods in estimating several dynamic prediction probabilities. More exactly, the targeted transition probabilities are $P_{gh}(s, t|\mathbf{Z} = \mathbf{0}) = P\big(X(t) = h|X(s) = g, \mathbf{Z} = \mathbf{0}\big)$ at various pre-specified time points s. They are referred to as the "true" dynamic prediction probabilities, and are given by

$$\begin{aligned} P_{gh}^{(M)}(s, t|\mathbf{Z} = \mathbf{0}\big) &= \big[\exp\{(t-s)\cdot Q_1\}\big]_{gh}, \text{ for } 0 \le s < t \le 5\ , \\ P_{gh}^{(M)}(s, t|\mathbf{Z} = \mathbf{0}\big) &= \big[\exp\{(5-s)\cdot Q_1\}\cdot\exp\{(t-5)\cdot Q_2\}\big]_{gh}, \text{ for } 0 \le s \le 5 < t\ , \\ P_{gh}^{(M)}(s, t|\mathbf{Z} = \mathbf{0}\big) &= \big[\exp\{(t-s)\cdot Q_2\}\big]_{gh}, \text{ for } 5 \le s < t \le 14\ , \end{aligned}$$

where, for a matrix W, notation $\big[\exp\{W\}\big]_{gh}$ stands for the (g, h) element of the matrix $\exp\{W\}$ (see Horn et Johnson (1991)).

We will also be interested to study the performance of different dynamic prediction methods under departures from the Markov assumption, allowing transition probabilities to the terminal states to depend on the time to reach state 2. To accomplish this we modify the cause-specific hazards model in (6.1) such that

$$\lambda_{2h}(t|\ \mathcal{F}_{t-},\ \mathbf{Z}) = \lambda_{2h,0}(t)\exp(\beta_{2h}(t)^\top\mathbf{Z} + \xi_{2h}\cdot T_2)\ , \tag{6.2}$$

where T_2 stands for the time to reach state 2 from state 1 and ξ_{2h} is the corresponding regression coefficient for transition $2 \to h$, $h \in \{3, 4\}$. Possible combinations (ξ_{23}, ξ_{24}) considered in our set-up are $\xi_{23} = -0.5$ and $\xi_{24} \in \{\pm 0.5, 0\}$.

Under this new setting, we deal with a non-Markov multi-state model, which we refer to as the "true" non-Markov model and which will be used to generate various data sets in our simulations; note that its initial distribution is degenerate in state 1. Again, several transition probabilities of the true non-Markov model are of reference because they will be used to asses the accuracy of our methods

in estimating several dynamic prediction probabilities. These are the transition probabilities to the terminal states $h = 3, 4$ at various pre-specified time points s, which we denote by $P_{gh}^{(nM)}(s, t|\mathbf{Z} = \mathbf{0})$. They cannot be estimated exactly; instead, they can be approximated by a Monte Carlo method. To this goal, during the data generation process from this non-Markov model, before the censoring is applied we counted for each of the simulated data set for each of the relevant combinations of $g, h, s, t, \mathbf{Z}$ how many satisfied $X(s) = g, \mathbf{Z}$ (denominator) and how many satisfied $X(s) = g, \mathbf{Z}, X(t) = h$ (nominator). The Monte Carlo conditional probabilities were finally obtained by adding the numerators and denominators over all simulated data sets and taking the ratio.

6.3 Dynamic prediction

Our problem can be formulated in two equivalent ways, as follows.

The multi-state formulation

We are interested in the dynamic prediction probability of experiencing a terminal event of a given type h, $h \in \{3, 4\}$, i.e. we want to model and estimate the probability of experiencing a terminal event of type h by time t, conditional on being event-free at a certain time s and possibly on a set of values for prognostic factors $\mathbf{Z}$ of a patient, that is

$$P(X(t) = h|\mathbf{Z}, \{X(u), u \leq s\}, X(s) \in \{1, 2\}) \ , \tag{6.3}$$

where $h \in \{3, 4\}$.

The competing risks formulation

It is worth noting that the multi-state process $X(t)$ can be reformulated as a competing risks process with a time-dependent covariate. Using the standard competing risks notation, we denote by T the event time variable, that is the time spent by the multi-state process $X(t)$ to reach one of the two competing, terminal events 3 or 4 from state 1, irrespective of the trajectory taken to reach it (irrespective whether it previously reached state 2 or not), and by D the corresponding type of terminal event, that is $D \in \{3, 4\}$. The transition of $X(\cdot)$ from state 1 to state 2 at some time s can be conveniently interpreted in this new framework as the change in status, at time s, of a binary time-dependent covariate $Z_3(\cdot)$ from 0 (before s) to 1 (from time s onwards). More exactly, given that the subject is event-free at time s, $Z_3(s) = 0$ if and only if $X(s) = 1$ and $Z_3(t) = 1$ for all $t \geq s$ if and only if $X(s) = 2$. Our target can be reformulated as

to model and estimate the conditional (on $T > s$) cumulative incidence function of cause h at time t, given no event by time s and given the current status of covariates at time s, that is

$$P(T \leq t,\ D = h|\mathbf{Z}(s), T > s)\ , \tag{6.4}$$

where $\mathbf{Z}(s) = (Z_1, Z_2, Z_3(s))$ and $h \in \{3, 4\}$.

In the remainder of the paper, we will specify in each context which formulation is preferred.

6.4 Methods

There are different ways to approach the modeling and estimation of the dynamic prediction probabilities (6.3) or (6.4). In the following, we will discuss three different perspectives on this problem.

6.4.1 Landmarking based on cause-specific hazards

General methods

We adopt the competing risks formulation and use the landmark approach described in Nicolaie et al. (2013a). To this goal, define a set of landmark time points $0 \leq s_1 < \ldots < s_K$. We want to model and estimate dynamic prediction probabilities at each $s \in [0, s_K]$ and for a fixed width prediction window w (such that $s_K + w < 14$), that is we want to infer over intervals of the form $[s, s + w]$ for varying s, $s \in [0, s_K]$. We build the landmark data sets corresponding to the selected grid of landmark time points.

For a fixed s, we postulate a Cox proportional hazards model on each conditional (on $T > s$) cause-specific hazard, that is

$$\lambda_h(t|\ \mathbf{Z}(s),\ s) = \lambda_{h0}(t|s) \exp(\phi_h(s)^\top \mathbf{Z}(s)),\ t \in [s, s + w], \tag{6.5}$$

where $\lambda_{h0}(t|s)$ is the conditional (on $T > s$) cause-specific baseline hazard of cause h, $\phi_h(s)$ is a vector of unknown regression coefficients, for $h = 3, 4$. We refer to (6.5) as to the separate landmark model; its parameters can be estimated by fitting the model to the landmark data set corresponding to s, where we impose administrative censoring at $s + w$. The dynamic prediction probabilities (6.4) can be estimated by

$$\widehat{P}^{\text{sep}}_{h,\text{CS}}(s + w|\mathbf{Z}(s), s) = \sum_{s<t_i\leq s+w} \widehat{\lambda}_h(t_i|\mathbf{Z}(s), s)\widehat{S}_{\text{LM}}(t_i - |\mathbf{Z}(s), s),\ h = 3, 4, \tag{6.6}$$

where

$$\widehat{\lambda}_h(t_i|\ \mathbf{Z}(s),s) = \widehat{\lambda}_{h0}(t_i|s)\exp(\widehat{\phi}_h(s)^\top \mathbf{Z}(s)), \qquad t_i > s,$$

is the estimated conditional cause-specific hazard at the event time t_i,

$$\widehat{S}_{\text{LM}}(s+w|\mathbf{Z}(s),s) = \exp\Big(-\sum_{h=3}^{4} e^{\widehat{\phi}_h(s)^\top \mathbf{Z}(s)}\widehat{\Lambda}_{h0}(s+w|s)\Big)$$

and $\widehat{\Lambda}_{h0}(t|s) = \sum_{t_i \le t}\widehat{\lambda}_{h0}(t_i|s)$ is the estimated conditional (on $T > s$) cumulative cause-specific baseline hazard of cause h at time t, $h = 3, 4$. Note that the estimated baseline hazards depend on the landmark time point s, because a separate model is fitted at each s.

Further, we include dynamic prediction methods based on cause-specific hazards that combine information from all landmark data sets. For $s \in [0, s_K]$, we postulate model (6.5) and we model the regression coefficients as follows

$$\phi_h(s) = f(s;\phi^{(h)}), \tag{6.7}$$

where $\phi^{(h)} = (\phi^{(h1)}, \ldots, \phi^{(hp_h)})$ is a p_h-vector of regression parameters for $h = 3, 4$ and $f(\cdot)$ is a parametric function of s, and the cause-specific baseline hazards $\lambda_{h0}(t|s)$, conditional on $T > s$, is an unspecified function of t, that is, for each cause h, each landmark time point s has a separate cause-specific baseline hazard, $h = 3, 4$.

Combining models (6.5) and (6.7) leads to the stratified supermodels on the conditional (on $T > s$, for varying s) cause-specific hazards, which are denoted now by $\lambda_h^\sharp(t|\mathbf{Z}(s),s)$, $h = 3, 4$ and can be estimated by fitting the stratified supermodels to the data set obtained by stacking the landmark data sets, where we impose administrative censoring at $s + w$, for varying s (see van Houwelingen (2007) and Nicolaie et al. (2013a)). The dynamic prediction probabilities (6.4) can be estimated by

$$\widehat{P}^{\text{str}}_{h,\text{CS}}(s+w|\mathbf{Z}(s),s) = \sum_{s<t_i\le s+w} \widehat{\lambda}_h^\sharp(t_i|\mathbf{Z}(s),s)\widehat{S}_{\text{LM}}(t_i - |\mathbf{Z}(s),s),\ s \in [0,s_K], \tag{6.8}$$

$h = 3, 4$, where

$$\begin{aligned}
\widehat{\lambda}_h^\sharp(t_i|\mathbf{Z}(s),s) &= \widehat{\lambda}_{h0}^\sharp(t_i|s)\exp(\widehat{\phi}_h(s)^\top \mathbf{Z}(s)),\\
\widehat{S}_{\text{LM}}(s+w|\mathbf{Z}(s),s) &= \exp\Big(-\sum_{h=3}^{4} e^{\widehat{\phi}_h(s)^\top \mathbf{Z}(s)}\widehat{\Lambda}_{h0}^\sharp(s+w|s)\Big),
\end{aligned}$$

$s \in [0, s_K]$ and $\widehat{\Lambda}_{h0}^\sharp(t|s) = \sum_{t_i \le t}\widehat{\lambda}_{h0}^\sharp(t_i|s)$ is the estimated cumulative cause-

specific baseline hazard of cause h at time t, $h = 3, 4$, specific to landmark time point s.

We obtain another dynamic prediction model if we postulate

$$\lambda_{h0}(t|s) = \lambda_{h0}(t) \exp(\gamma_h(s)), \tag{6.9}$$

where $\gamma_h(s)$ are some parametric functions of s for $h = 3, 4$, with the restriction $\gamma(0) = 0$ to guarantee identifiability. We assume that

$$\gamma_h(s) = g(s; \gamma^{(h)}) \ , \tag{6.10}$$

where $\gamma^{(h)} = (\gamma^{(h1)}, \ldots, \gamma^{(hr_h)})$ is a r_h-vector of regression parameters and $g(\cdot)$ is a parametric function of s, for $h = 3, 4$.

Combining models (6.5), (6.7) and (6.9)–(6.10) leads to the so-called supermodels on the conditional (on $T > s$, for varying s) cause-specific hazards, which are denoted now by $\lambda_h^*(t|\mathbf{Z}(s), s)$, $h = 3, 4$, and can be estimated by fitting the supermodels to the data set obtained by stacking the landmark data sets, where we impose administrative censoring at $s + w$, for varying s (see van Houwelingen (2007) and Nicolaie et al. (2013a)). The dynamic prediction probabilities (6.4) can be estimated by

$$\widehat{P}^{\text{sup}}_{h,\text{CS}}(s + w|\mathbf{Z}(s), s) = \sum_{s < t_i \leq s+w} \widehat{\lambda}^*_h(t_i|\mathbf{Z}(s), s) \widehat{S}_{\text{LM}}(t_i - |\mathbf{Z}(s), s), \ s \in [0, s_K], \tag{6.11}$$

$h = 3, 4$, where

$$\begin{aligned} \widehat{\lambda}^*_h(t_i|\mathbf{Z}(s), s) &= \widehat{\lambda}_{h0}(t) \exp(\widehat{\phi}_h(s)^\top \mathbf{Z}(s) + \widehat{\gamma}_h(s)), \\ \widehat{S}_{\text{LM}}(s + w|\mathbf{Z}(s), s) &= \exp\Big(- \sum_{h=3}^{4} e^{\widehat{\phi}_h(s)^\top \mathbf{Z}(s) + \widehat{\gamma}_h(s)} \big[\widehat{\Lambda}^*_{h0}(s + w) - \widehat{\Lambda}^*_{h0}(s-)\big]\Big), \end{aligned}$$

$s \in [0, s_K]$ and $\widehat{\Lambda}^*_{h0}(t) = \sum_{t_i \leq t} \widehat{\lambda}^*_{h0}(t_i)$ is the estimated cumulative cause-specific baseline hazard of cause h at time t, $h = 3, 4$.

Model building

We fix the prediction window width w at 5 years and prediction time points $s_k \in \{0, 1, \ldots, 5\}$ years. We computed predictions over $[s_k, s_k + 5]$. Building of the separate and (stratified) supermodels is based on a covariate selection procedure, which is described in the following. Besides, for building the (stratified) supermodel we select a denser grid of prediction time points, that is, the sequence from 0 to 5 of equally spaced values with an increment of 0.2.

First, for a fixed s_k, the separate landmark model at s_k is given by

$$\lambda_h(t|\ \mathbf{Z}(s_k),\ s_k) = \lambda_{h0}(t|s_k)\exp(\phi_{h1}(s_k)Z_1 + \phi_{h2}(s_k)Z_2 + \phi_{h3}(s_k)Z_3(s_k)),$$

$t \in [s_k, s_k + 5]$, with the restriction $\phi_{h3}(0) = 0$. We removed from the model those covariates for which the corresponding subset of individuals at s_k belongs to only one subgroup as defined by that covariate. For $h = 3, 4$, we computed $\widehat{P}^{\text{sep}}_{h,\text{CS}}(s_k{+}5|Z_{1i}, Z_{2i}, Z_{i3}(s_k) = 0, s_k)$ using (6.6) for individuals i with $Z_{i3}(s_k) = 0$ and $\widehat{P}^{\text{sep}}_{h,\text{CS}}(s_k + 5|Z_{1i}, Z_{2i}, Z_{3i}(s_k) = 1, s_k)$ for individuals i with $Z_{i3}(s_k) = 1$.

For building the (stratified) supermodels, we chose for each covariate $f_h(s) = \phi^{(h1)} + \phi^{(h2)}s + \phi^{(h3)}s^2$ and $\gamma_h(s) = \gamma^{(h1)}s + \gamma^{(h2)}s^2$, $h = 3, 4$. For each of the competing endpoints $h = 3$ and $h = 4$, a backward selection procedure was used, starting from a model with all time-fixed covariates effects described by quadratic terms, where Wald tests were used to test whether the linear and quadratic terms $\phi^{(h2)}$, $\phi^{(h3)}$, $\gamma^{(h1)}$ and $\gamma^{(h2)}$ could be removed. This resulted in a final (stratified) supermodel based on which we computed $\widehat{P}^{\text{M}}_{h,\text{CS}}(s_k + 5|Z_{1i}, Z_{2i}, Z_{i3}(s_k) = 0, s_k)$ based on (6.8) and (6.11) for individuals i with $Z_{i3}(s_k) = 0$ and $\widehat{P}^{\text{M}}_{h,\text{CS}}(s_k + 5|Z_{1i}, Z_{2i}, Z_{3i}(s_k) = 1, s_k)$ for individuals i with $Z_{i3}(s_k) = 1$, where M = str and M = sup.

6.4.2 Models based on dynamic pseudo-observations

General methods

We adopt the competing risks formulation and we use the landmark approach described in Nicolaie et al. (2013b). To this goal, define a set of landmark time points $0 \leq s_1 < \ldots < s_K$. We want to model and estimate dynamic prediction probabilities at each $s \in [0, s_K]$ and for a fixed width prediction window w (such that $s_K + w < 14$), that is we want to infer over intervals of the form $[s, s + w]$ for varying s, $s \in [0, s_K]$; note that this time we do not impose administrative censoring at the horizon $s{+}w$. Denote by D_s the landmark data set corresponding to s and by n_s its sample size.

Let $s < t$ be two time points and define the cumulative incidence of event h, conditional on being event-free at time s, by

$$F_h(t|s) = P(T \leq t, D = h|T > s), \qquad \text{for } h = 3, 4.$$

We denote by $t_1 < t_2 < \ldots$ the distinct times at which events occur irrespective of the cause. Let $d_h(t_k)$ be the number of individuals who die at time t_k from cause h and $d(t_k) = \sum_{h=3}^{4} d_h(t_k)$ be the number of individuals who die at time t_k from any cause. Let $r(t_k)$ be the number of individuals at risk just prior to time t_k. Let $\widehat{F}_h(\cdot|s)$ be the non-parametric estimator of the conditional probability

$F_h(\cdot|s)$, as given by

$$\widehat{F}_h(t|s) = \sum_{s<t_k\leq t} \widehat{S}(t_k - |s)\frac{d_h(t_k)}{r(t_k)} \ , \tag{6.12}$$

where

$$\widehat{S}(t - |s) = \prod_{s<t_l\leq t} \{1 - \frac{d(t_l)}{r(t_l)}\}$$

is the Kaplan-Meier estimate of the conditional survival function given no event before time s.

Define the dynamic pseudo-observation within D_s for $\mathbf{1}\{T \leq s + w, D = h\}$ for individual i at risk at s by

$$\hat{\theta}_{is}^h = n_s\hat{F}_h(s + w|s) - (n_s - 1)\hat{F}_h^{(-i)}(s + w|s), \tag{6.13}$$

where $\hat{F}_h^{(-i)}(t|s)$ is the non-parametric Aalen-Johansen estimator of $F_h(t|s)$ based on the sample of size $n_s - 1$ obtained by eliminating individual i from D_s, for $h = 3, 4$ and $i = 1, \ldots, n_s$.

For a fixed s and a cause h, the separate landmark approach consists of specifying a generalized linear model on the expectation $\mu_i^{(h)}(s) = \mathbf{E}[Y_i(s)|\mathbf{Z}_i(s), T_i > s]$ of the indicators $Y_i(s) = \mathbf{1}\{T_i \leq s + w, D_i = h\}$ for individual i at risk at s, that is

$$g(\mu_i^{(h)}(s)) = \beta_{(h)}^\top(s)\mathbf{Z}_i^*(s), \tag{6.14}$$

for a given link function g, where $\beta_{(h)}(s) = (\beta_0^{(h)}(s), \beta_1^{(h)}(s), \ldots, \beta_p^{(h)}(s))$ and $\mathbf{Z}^*(s) = (1,\ \mathbf{Z}^\top(s))^\top$, so that $\beta_0^{(h)}(s)$ stands for the intercept, for $h = 3, 4$. We shall refer to model (6.14) as the separate landmark model based on the dynamic pseudo-observations.

Under mild conditions, the vector of regression parameters $\beta_{(h)}(s)$ could be estimated consistently by using generalized estimating equations; for details see Nicolaie et al. (2013b). Let $\widehat{\beta}_{(h)}(s)$ be the estimator of $\beta_{(h)}(s)$. The dynamic prediction probabilities (6.4) can be estimated by

$$\widehat{P}_{h,\mathrm{PS}}^{\mathrm{sep}}(s + w|\mathbf{Z}(s), s) = g^{-1}(\widehat{\beta}_{(h)}^\top(s)\widetilde{\mathbf{Z}}^*(s)),\ h = 3, 4, \tag{6.15}$$

where $\widetilde{\mathbf{Z}}^*(s) = (1,\ \widetilde{\mathbf{Z}}(s))$.

Now we want to combine information from different D_s. For $s \in [0, s_K]$, we postulate model (6.14) and we model the time-dependent $\beta_{(h)}(s)$ such that

$$\beta_{(h)}(s) = f(s; \beta^{(h)}), \tag{6.16}$$

where $f(\cdot)$ is a parametric linear function of s. Define $\boldsymbol{\beta_{(h)}}$ to be the vector containing all regression vectors.

Combining models (6.14) and (6.16) leads to the supermodel on the dynamic pseudo-observations. Under mild conditions and assuming a working independence correlation across the dynamic pseudo-observations of an individual for different landmark time points, the vector of regression parameters $\boldsymbol{\beta_{(h)}}$ could be estimated consistently by using generalized estimating equations. Let $\widehat{\boldsymbol{\beta}}_{(h)}$ be the estimator of $\boldsymbol{\beta_{(h)}}$. Then the vector $\widehat{\beta}_{(h)}(s)$ can be written as $H(s)\widehat{\boldsymbol{\beta}}_{\boldsymbol{(h)}}$, with $H(s)$ a $(p+1)\times q$ matrix containing the linear components of function f. The dynamic prediction probabilities (6.4) can be estimated by

$$\widehat{P}^{\text{sup}}_{h,\text{PS}}(s+w|\mathbf{Z}(s),s) = g^{-1}(\widehat{\beta}_{(h)}(s)^\top \widetilde{\mathbf{Z}}^*(s)),\ s\in[0,s_K],\ h=3,4, \tag{6.17}$$

where $\widetilde{\mathbf{Z}}^*(s) = (1,\ \widetilde{\mathbf{Z}}(s))$.

Model building

We fix the prediction window width and prediction time points as in Section 6.4.1. Building of the separate model and of the supermodel is based on a covariate selection procedure, which is described in the following. Besides, for building the supermodel we select a wider grid of prediction time points, that is, the sequence from 0 to 5 of equally spaced values with an increment of 0.2.

First, for a fixed s_k and a fixed h, the separate landmark model at s_k is given by

$$g(\mu_i(s_k)) = \beta_1^{(h)}(s_k)Z_1 + \beta_2^{(h)}(s_k)Z_2 + \beta_3^{(h)}(s_k)Z_3(s_k),$$

with the restriction $\beta_3^{(h)}(0)=0$, for $h=3,4$. We removed from the model those covariates for which the corresponding subset of individuals at baseline belongs to only one subgroup as defined by each covariate. We computed $\widehat{P}^{\text{sep}}_{h,\text{PS}}(s_k+5|Z_{1i},Z_{2i},Z_{i3}(s_k)=0,s_k)$ for individuals i with $Z_{i3}(s_k)=0$ and $\widehat{P}^{\text{sep}}_{h,\text{PS}}(s_k+5|Z_{1i},Z_{2i},Z_{3i}(s_k)=1,s_k)$ for individuals i with $Z_{i3}(s_k)=1$, both using (6.15).

For building supermodels, we chose for each covariate $\beta_{(h)l}(s) = \beta_{(h)l1} + \beta_{(h)l2}s + \beta_{(h)l3}s^2$. For each of the competing end points $h=3$ and $h=4$, a backward selection procedure was used, starting from a model with all time-fixed covariates effects described by quadratic terms, where Wald tests were used to test whether the linear and quadratic terms could be removed. This resulted in a final supermodel based on which we computed $\widehat{P}^{\text{sup}}_{h,\text{PS}}(s_k+5|Z_{1i},Z_{2i},Z_{i3}(s_k)=0,s_k)$ for individuals i with $Z_{i3}(s_k)=0$ and $\widehat{P}^{\text{sup}}_{h,\text{PS}}(s_k+5|Z_{1i},Z_{2i},Z_{3i}(s_k)=1,s_k)$ for individuals i with $Z_{i3}(s_k)=1$, for $h=3,4$, both using (6.17).

6.4.3 Markov multi-state model

General method

We adopt the multi-state formulation. The multi-state approach relies on assuming that the multi-state process $X(t)$ is Markovian and its distribution is specified through the transition-specific hazards $\lambda_{gh}(t)$. We postulate Cox proportional hazards model on each $\lambda_{gh}(t|\mathbf{Z})$, that is

$$\lambda_{gh}(t|\ \mathcal{F}_{t-},\ \mathbf{Z}) = \lambda_{gh,0}(t)\exp(\boldsymbol{\zeta}_{gh}^{\top}\mathbf{Z}), \tag{6.18}$$

where $\lambda_{gh,0}(t)$ is an unspecified transition-specific intensity and $\boldsymbol{\zeta}_{gh}$ stands for a vector of regression parameters, for all possible transitions $g \to h$.

The transition probabilities, for $s \leq t$ and $h = 3, 4$ are given by:

$$\begin{aligned}
P_{12}(s,t|\mathbf{Z}) &= \int_s^t P_{11}(s,u-|\mathbf{Z})\lambda_{12}(u|\mathbf{Z})P_{22}(u,t|\mathbf{Z})du,\\
P_{2h}(s,t|\mathbf{Z}) &= \int_s^t P_{22}(s,u-|\mathbf{Z})\lambda_{2h}(u|\mathbf{Z})du,\\
P_{1h}(s,t|\mathbf{Z}) &= \int_s^t \big[P_{11}(s,u-|\mathbf{Z})\lambda_{1h}(u|\mathbf{Z}) + P_{12}(s,u-|\mathbf{Z})\lambda_{2h}(u|\mathbf{Z})\big]du,
\end{aligned}$$

where $P_{jj}(s,t|\mathbf{Z}) = \exp\big\{-\sum_{l>j}\int_s^t \lambda_{jl}(u|\mathbf{Z})du\big\}$, $j = 1, 2$, stand for the state occupation probabilities. The transition probability matrix

$$P(s,t|\mathbf{Z}) = (P_{gh}(s,t|\mathbf{Z}))_{g,h\in\{1,\ldots,4\}}$$

can be estimated by the Aalen-Johansen estimator (Aalen and Johansen, 1978). However, note that the effect of $\mathbf{Z}$ on $P_{1h}(s,t|\mathbf{Z})$ is not described by simple parameters. We denote the estimated dynamic prediction probabilities (6.3) obtained via this Markov multi-state approach by $\widehat{P}_{h,\mathrm{MM}}(s+w|\mathbf{Z},s)$, $h = 3, 4$.

Model building

We fix the prediction window width and prediction time points as in Section 6.4.1. We removed from the modeling of a certain transition those baseline covariates for which the corresponding subset of individuals at risk for that transition belongs to only one subgroup as defined by each covariate. We computed $\widehat{P}_{h,\mathrm{MM}}(s_k + 5|\mathbf{Z}_i, X(s_k) = 1, s_k)$ for individuals i with $X(s_k) = 1$ as $\widehat{P}_{1h}(s_k, s_k + 5|\mathbf{Z}_i)$ and $\widehat{P}_{h,\mathrm{MM}}(s_k + 5|\mathbf{Z}_i, X(s_k) = 2, s_k)$ for individuals i with $X(s_k) = 2$ as $\widehat{P}_{2h}(s_k, s_k + 5|\mathbf{Z}_i)$.

6.5 Simulation and results

Each of the methods introduced in Sections 6.4.1, 6.4.2 and 6.4.3 were applied to each of the simulated data sets. We computed the probabilities $\widehat{P}^{\text{sep}}_{h,\text{CS}}$, $\widehat{P}^{\text{str}}_{h,\text{CS}}$, $\widehat{P}^{\text{sup}}_{h,\text{CS}}$, $\widehat{P}^{\text{sep}}_{h,\text{PS}}$, $\widehat{P}^{\text{sup}}_{h,\text{PS}}$ and $\widehat{P}_{h,\text{MM}}$ at each of the prediction time points $s_k \in \{0, 1, \ldots, 5\}$ years and for a prediction window width of 5 years, for an individual i with $Z_{i1} = Z_{i2} = 0$. We reported the estimated bias and root mean squared error (RMSE) on a scale of order 10^{-2}, when the true underlying model is either Markovian or non-Markovian. Results will be shown for specific choices of scenarios. We ran 10000 simulations on each scenario.

6.5.1 True Markovian model

First, we study the case where the multi-state model is Markovian and the covariates exhibit time-fixed effects on the transition intensities. Results are shown for $c_1 = c_2 = -0.5$ and $\xi_{23} = \xi_{24} = 0$. Therefore, the distinction between the two competing events becomes apparent only at the baseline transition intensities. Table 6.1 shows the results.

In terms of model fitting, the multi-state model of Section 6.4.3 is larger than the true underlying model because the proposed model (6.18) allows $\mathbf{Z}$ to exhibit transition-specific effects. The landmark models of Sections 6.4.1 and 6.4.2 are hard to reconcile with the true underlying model because they are not comprehensive models; however, note that they assume no common effect of $\mathbf{Z}$ on the dynamics of the two competing events (see (6.5) and (6.14)). Instead, the two landmark approaches of Sections 6.4.1 and 6.4.2 differ from each other with respect to the functional relation between covariates and conditional cumulative incidence function at $s + w$ (see, e.g., (6.11) and (6.17)).

In general, the estimates are close to the true values though a slight bias is observed. The clear winner is the multi-state approach: virtually unbiased and smaller RMSE than competitors.

As expected, the most accurate estimates across the landmark models are $\widehat{P}^{\text{sep}}_{h,\text{CS}}$, $\widehat{P}^{\text{sep}}_{h,\text{PS}}$ and $\widehat{P}^{\text{sup}}_{h,\text{PS}}$, while the most efficient estimates overall are $\widehat{P}_{h,\text{MM}}$.

With the landmark model based on dynamic pseudo-observations, the estimators produced comparable RMSE, but higher bias for supermodels than for separate models. With the separate landmark model based on cause-specific hazards, the estimators produced acceptable bias and slightly smaller RMSE than the separate landmark model based on dynamic pseudo-observations. Of special note, supermodels based on cause-specific hazards come with large bias, smaller for the stratified one.

An interesting observation goes directly to the core of the issue concerning the large bias of $\widehat{P}^{\text{sup}}_{h,\text{CS}}$. We plotted the estimated cumulative baseline all-causes

Table 6.1: Estimated bias and RMSE for $c_1 = c_2 = -0.5$ and $\xi_{23} = \xi_{24} = 0$.

	$\widehat{P}_{h,\mathrm{MM}}$				$\widehat{P}^{\mathrm{sup}}_{h,\mathrm{CS}}$			
	No intermediate Event		Intermediate Event		No intermediate Event		Intermediate Event	
LM	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4
0	0.00 (3.29)	0.00 (2.93)			4.21 (5.89)	2.86 (4.70)		
1	0.03 (3.59)	0.01 (3.25)	0.05 (6.21)	-0.11 (5.55)	2.04 (4.71)	0.92 (3.91)	3.35 (7.63)	0.42 (5.58)
2	0.09 (3.99)	0.04 (3.75)	0.06 (6.24)	-0.07 (5.62)	-1.36 (4.96)	-0.38 (4.45)	-1.56 (6.74)	-0.61 (5.53)
3	0.16 (4.51)	0.10 (4.38)	0.12 (6.69)	-0.05 (6.12)	-4.28 (6.69)	-1.45 (5.34)	-6.79 (9.76)	-1.36 (6.10)
4	0.08 (5.10)	0.25 (5.40)	0.21 (7.42)	-0.10 (6.94)	-5.36 (7.84)	-2.80 (6.80)	-11.14 (13.81)	-2.28 (7.15)
5	0.09 (5.88)	0.34 (7.00)	0.22 (8.44)	-0.11 (8.12)	-3.38 (7.42)	-4.93 (9.24)	-13.91 (17.42)	-3.56 (9.25)
	$\widehat{P}^{\mathrm{str}}_{h,\mathrm{CS}}$				$\widehat{P}^{\mathrm{sep}}_{h,\mathrm{CS}}$			
	No intermediate Event		Intermediate Event		No intermediate Event		Intermediate Event	
LM	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4
0	0.05 (3.61)	0.00 (3.24)			0.02 (3.40)	0.01 (3.05)		
1	-0.29 (3.96)	-0.09 (3.57)	0.46 (6.86)	0.45 (5.57)	-0.06 (4.07)	0.00 (3.72)	-0.51 (7.61)	-0.18 (6.67)
2	-0.12 (4.41)	0.00 (4.13)	-0.82 (6.29)	0.00 (5.35)	0.04 (4.87)	0.11 (4.62)	-1.13 (7.17)	-0.45 (6.31)
3	0.23 (4.91)	0.33 (4.83)	-1.27 (6.45)	-0.55 (5.60)	0.13 (5.84)	0.33 (5.79)	-1.29 (7.71)	-0.59 (6.82)
4	0.82 (5.97)	0.75 (5.95)	-1.52 (7.82)	-0.93 (6.47)	-0.04 (6.98)	0.43 (7.43)	-0.74 (8.80)	-0.48 (7.90)
5	2.73 (8.40)	0.55 (7.72)	-2.21 (10.61)	-0.74 (8.32)	-1.14 (8.19)	0.11 (9.64)	1.17 (10.94)	-0.05 (9.61)
	$\widehat{P}^{\mathrm{sup}}_{h,\mathrm{PS}}$				$\widehat{P}^{\mathrm{sep}}_{h,\mathrm{PS}}$			
	No intermediate Event		Intermediate Event		No intermediate Event		Intermediate Event	
LM	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4
0	-0.93 (3.85)	-0.48 (3.37)			-0.70 (3.53)	-0.58 (3.12)		
1	-0.44 (4.00)	-0.62 (3.51)	-0.54 (6.97)	-0.72 (5.76)	-0.59 (4.24)	-0.54 (3.79)	-0.90 (7.92)	-0.83 (6.71)
2	-0.52 (4.65)	-0.59 (4.19)	-0.37 (6.36)	-0.51 (5.45)	-0.50 (5.10)	-0.58 (4.68)	-0.61 (7.30)	-0.55 (6.39)
3	-0.85 (5.30)	-0.54 (5.00)	-0.27 (6.82)	-0.23 (6.03)	-0.53 (6.15)	-0.62 (5.82)	-0.22 (7.78)	-0.27 (6.92)
4	-0.81 (6.58)	-0.65 (6.59)	-0.16 (7.98)	-0.07 (7.06)	-0.68 (7.46)	-0.74 (7.51)	0.09 (8.97)	0.08 (8.11)
5	0.90 (8.68)	-1.19 (9.66)	-0.04 (11.17)	0.10 (9.79)	-0.78 (8.91)	-0.90 (9.93)	0.25 (11.20)	0.49 (9.90)

hazards of the stratified supermodel and of the supermodel on cause-specific hazards in Figure 6.2 based on a representative simulated data set. We compared them with the cumulative baseline all-causes hazards obtained from the true Markovian model of this scenario. The steeper increase in the cumulative all-causes hazards obtained from the former landmark method compared to the latter landmark method reflects the fact that information on the number and type of events is used differently in the two models. An exploratory analysis in which we replaced the second degree polynomials in (6.10) by smoothing splines led to the same amount of bias. This aspect needs to be further investigated.

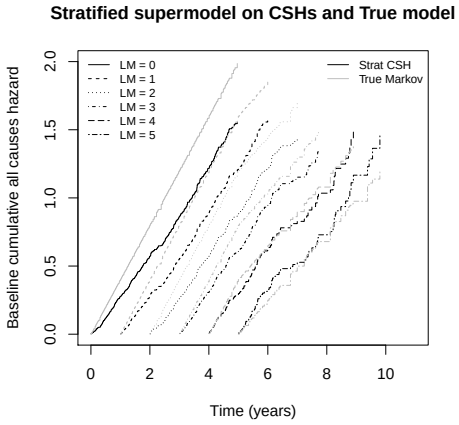
The higher variability in the estimates of dynamic prediction probabilities of cause $h = 3$ compared to those of cause $h = 4$ consistently observed for each modeling technique can be explained by the steeper decrease of the number of events of type $h = 3$ compared to those of cause $h = 4$ across the prediction intervals $[s, s + w]$, as suggested by Figure 6.3.

Secondly, we study the case where the true underlying multi-state model is Markovian and the covariates exhibit either time-fixed or time-varying effects on the transition intensities. We take $c_1 = c_2 = -0.5$ for $t \in [0, 2.5]$, $c_1 = -0.5$ and $c_2 = -1$ for $t \in [2.5, 14]$, and $\xi_{23} = \xi_{24} = 0$. Again, the distinction between the two competing events becomes apparent only at the baseline transition intensities. Table 6.2 shows the results.

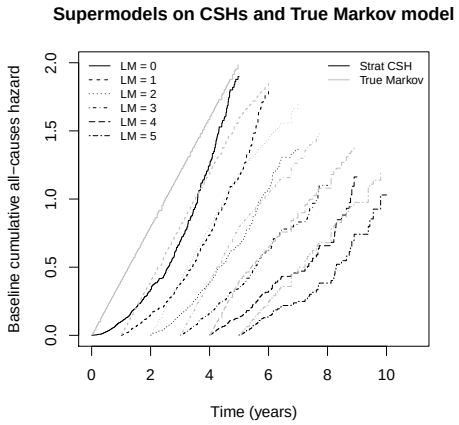
In terms of model fitting, the multi-state model of Section 6.4.3 misspecifies the true underlying model because the model (6.18) does not capture the time varying-effect $\beta_{gh,2}(t)$ of Z_2, but $\boldsymbol{\zeta}_{gh,2}$ rather estimates a time averaged effect of Z_2 over $[0, 14]$.

In terms of variability, the winner is again the multi-state model, which produces the best precision despite the fact that the fit of the model is not perfect. In terms of bias, the estimators $\widehat{P}^{\text{str}}_{h,\text{CS}}$, $\widehat{P}^{\text{sep}}_{h,\text{CS}}$, $\widehat{P}^{\text{sep}}_{h,\text{PS}}$ and $\widehat{P}^{\text{sup}}_{h,\text{PS}}$ are quite comparable, with smaller bias than the multi-state model. This phenomenon clearly shows the trade-off present with modeling by the two techniques, the (separate) landmark model on dynamic pseudo-observations producing smallest bias, while the Markovian multi-state inferring at the least waste of information. Intuitively, having less information should result in the variance increasing. Again, the $\widehat{P}^{\text{sup}}_{h,\text{CS}}$ performs dramatically in terms of bias. However, when we increased the magnitude of the time-varying effect of Z_2 (we replaced $c_2 = -1$ by $c_2 = -2$ for $t \in [2.5, 14]$ in the true model), the variability (both bias and RMSE) exceedingly increased for the multi-state approach, while the landmark models performed relatively stable (results not shown).

We were also interested to check how much our dynamic prediction techniques were influenced by the sample size. To this goal, we increased the sample size to $n = 1000$. As expected, all the methods produced slightly more accurate estimators in terms of RMSE, while bias did not change appreciably.



(a)



(b)

Figure 6.2: The all-causes cumulative baseline hazards from (a) the stratified landmark supermodel and (b) the supermodel on cause-specific hazards based on one simulated data set (in black), and the true cumulative baseline all-causes hazards (in grey), when the true underlying model is Markovian with no covariate effects.

Table 6.2: Estimated bias and RMSE for $c_1 = c_2 = -0.5$ for $t \in [0, 2.5]$, $c_1 = -0.5$ and $c_2 = -1$ for $t \in [2.5, 14]$, and $\xi_{23} = \xi_{24} = 0$.

	$\widehat{P}_{h,\text{MM}}$				$\widehat{P}^{\text{sup}}_{h,\text{CS}}$			
	No intermediate Event		Intermediate Event		No intermediate Event		Intermediate Event	
LM	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4
0	-0.68 (3.35)	-0.14 (2.96)			4.51 (6.05)	3.32 (4.97)		
1	-1.76 (3.99)	-0.81 (3.42)	-0.17 (6.27)	-0.15 (5.56)	1.31 (4.51)	0.63 (3.94)	2.69 (7.59)	0.26 (5.76)
2	-3.27 (5.17)	-1.62 (4.22)	-0.91 (6.52)	-0.39 (5.81)	-3.16 (5.84)	-1.25 (4.82)	-3.06 (7.54)	-1.20 (5.84)
3	-4.38 (6.26)	-2.40 (5.19)	-1.50 (7.22)	-0.46 (6.49)	-6.61 (8.44)	-2.76 (6.06)	-8.75 (11.47)	-2.21 (6.56)
4	-4.43 (6.63)	-3.02 (6.25)	-1.55 (8.02)	-0.52 (7.39)	-7.52 (9.51)	-4.34 (7.73)	-12.85 (15.47)	-3.22 (7.73)
5	-3.87 (6.61)	-4.12 (7.90)	-1.63 (9.15)	-0.52 (8.69)	-5.36 (8.51)	-6.51 (10.27)	-15.08 (18.65)	-4.43 (9.81)
	$\widehat{P}^{\text{str}}_{h,\text{CS}}$				$\widehat{P}^{\text{sep}}_{h,\text{CS}}$			
	No intermediate Event		Intermediate Event		No intermediate Event		Intermediate Event	
LM	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4
0	-0.40 (3.84)	0.14 (3.43)			-0.97 (3.53)	-0.58 (3.16)		
1	-0.83 (4.03)	-0.24 (3.61)	-0.04 (7.07)	0.36 (5.75)	-1.09 (4.21)	-0.56 (3.84)	-1.57 (7.93)	-0.73 (6.83)
2	-1.20 (4.86)	-0.54 (4.43)	-1.36 (6.68)	-0.47 (5.56)	-0.65 (4.98)	-0.16 (4.71)	-1.39 (7.53)	-0.78 (6.54)
3	-1.18 (5.74)	-0.50 (5.48)	-1.93 (7.31)	-1.17 (6.13)	-0.20 (5.99)	0.34 (5.94)	-0.85 (7.95)	-0.65 (7.08)
4	-0.33 (6.93)	-0.22 (6.81)	-1.95 (8.95)	-1.54 (7.26)	-0.56 (7.21)	0.47 (7.64)	-0.33 (9.00)	-0.58 (8.02)
5	1.59 (9.15)	-0.72 (8.82)	-2.67 (11.90)	-1.39 (9.24)	-1.61 (8.28)	0.03 (9.86)	1.45 (10.95)	-0.10 (9.65)
	$\widehat{P}^{\text{sup}}_{h,\text{PS}}$				$\widehat{P}^{\text{sep}}_{h,\text{PS}}$			
	No intermediate Event		Intermediate Event		No intermediate Event		Intermediate Event	
LM	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4
0	-0.79 (3.92)	-0.44 (3.50)			-0.80 (3.59)	-0.73 (3.23)		
1	-0.52 (4.06)	-0.84 (3.65)	-0.53 (7.10)	-0.72 (5.91)	-0.68 (4.28)	-0.76 (3.91)	-1.03 (8.05)	-0.93 (6.91)
2	-0.77 (4.84)	-1.02 (4.39)	-0.28 (6.59)	-0.51 (5.65)	-0.65 (5.14)	-0.91 (4.80)	-0.34 (7.52)	-0.46 (6.64)
3	-1.28 (5.55)	-1.19 (5.28)	-0.07 (7.11)	-0.07 (6.28)	-0.79 (6.18)	-1.19 (6.01)	0.25 (8.01)	0.04 (7.24)
4	-1.31 (6.75)	-1.51 (6.94)	0.29 (8.25)	0.34 (7.34)	-1.06 (7.53)	-1.45 (7.78)	0.69 (9.20)	0.53 (8.28)
5	0.32 (8.70)	-2.12 (10.10)	0.96 (11.42)	0.92 (10.14)	-1.37 (8.90)	-1.91 (10.18)	0.83 (11.26)	1.24 (10.09)

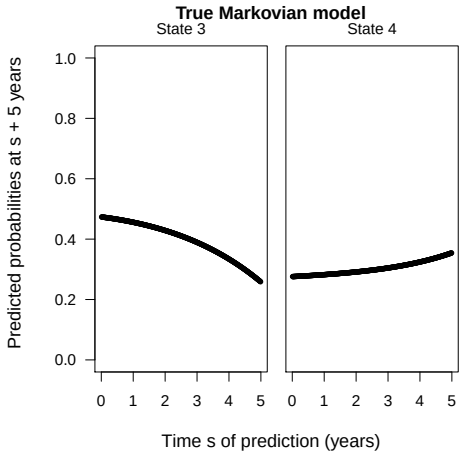


Figure 6.3: The true prediction probabilities $P_{1h}^{(M)}(s, s+5|Z_1 = Z_2 = 0)$, $h = 3, 4$, for varying $s \in [0, 5]$, when the underlying model is the Markovian model used for Table 1 or for Table 2 at the baseline values of covariates $Z_1 = Z_2 = 0$.

6.5.2 True non-Markovian model

Here, we study the case where the true underlying multi-state model is non-Markovian and the baseline covariates exhibit time-fixed effects on the transition intensities. We take $c_1 = c_2 = -0.5$ and $\xi_{23} = -0.5$, $\xi_{24} = 0.5$. Therefore, the distinction between the two competing events becomes apparent at baseline transition intensities and covariate effects levels. Table 6.3 shows the results.

A remarkable fact is that across the landmark models, the estimators remain nearly of the same quality as in the previous scenarios. Instead, the multi-state model of Section 6.4.3 produced larger bias and RMSE; this is especially true for early prediction times and when the intermediate event occurs. Here, for $s_k \in \{0, 1\}$, the multi-state approach is beaten by the last four methods for "no intermediate event"; instead, for "intermediate event", this is true for $s_k \in \{0, 1, 2, 3\}$.

6.6 Discussion

We have compared several modeling approaches to competing risks in order to asses their prediction accuracy when the goal is to do dynamic prediction of the competing events in the presence of time-dependent covariates. Several important

Table 6.3: Estimated bias and RMSE for the non-Markovian scenario for $c_1 = c_2 = -0.5$ and $\xi_{23} = -0.5$, $\xi_{24} = 0.5$.

	$\widehat{P}_{h,\text{MM}}$				$\widehat{P}^{\text{sup}}_{h,\text{CS}}$			
	No intermediate Event		Intermediate Event		No intermediate Event		Intermediate Event	
LM	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4
0	-0.02 (3.28)	0.00 (3.11)			5.15 (6.57)	1.36 (4.13)		
1	2.18 (4.13)	-3.00 (4.61)	-13.68 (14.96)	19.66 (20.62)	1.26 (4.22)	2.32 (4.83)	-1.52 (7.28)	3.55 (8.16)
2	3.17 (5.00)	-4.84 (6.27)	-9.28 (10.87)	15.50 (16.67)	-2.77 (5.13)	1.92 (5.38)	-4.08 (7.33)	2.45 (7.23)
3	2.64 (5.07)	-3.95 (6.12)	-5.06 (7.62)	8.87 (10.92)	-5.55 (7.18)	0.51 (5.85)	-5.69 (8.01)	-0.62 (7.37)
4	0.37 (4.94)	-0.37 (5.65)	-0.70 (6.22)	1.97 (7.19)	-6.91 (8.67)	-1.77 (7.14)	-6.09 (8.20)	-3.74 (9.18)
5	-3.94 (6.89)	5.54 (8.96)	4.62 (8.66)	-5.26 (9.62)	-7.25 (9.66)	-4.28 (9.09)	-5.91 (8.49)	-5.82 (12.13)
	$\widehat{P}^{\text{str}}_{h,\text{CS}}$				$\widehat{P}^{\text{sep}}_{h,\text{CS}}$			
	No intermediate Event		Intermediate Event		No intermediate Event		Intermediate Event	
LM	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4
0	0.18 (3.63)	-0.20 (3.39)			0.13 (3.36)	-0.16 (3.27)		
1	0.37 (3.82)	-0.39 (3.96)	-1.94 (6.99)	1.09 (6.83)	0.03 (3.92)	-0.11 (4.07)	-0.17 (7.56)	0.01 (7.34)
2	0.06 (4.12)	0.12 (4.58)	-0.09 (6.05)	-0.88 (6.44)	0.10 (4.61)	0.18 (4.96)	-0.51 (7.02)	-0.65 (7.27)
3	-0.34 (4.54)	1.03 (5.43)	0.52 (5.71)	-1.94 (6.86)	0.33 (5.54)	0.57 (6.07)	-0.86 (7.04)	-1.16 (7.77)
4	-0.42 (5.30)	1.25 (6.44)	-0.19 (5.57)	-0.62 (6.93)	0.45 (7.01)	0.39 (7.38)	-0.73 (7.37)	-0.63 (8.33)
5	0.29 (6.92)	0.23 (7.88)	-2.77 (6.80)	3.68 (8.93)	0.08 (9.24)	-1.69 (8.89)	-0.22 (8.63)	1.83 (9.76)
	$\widehat{P}^{\text{sup}}_{h,\text{PS}}$				$\widehat{P}^{\text{sep}}_{h,\text{PS}}$			
	No intermediate Event		Intermediate Event		No intermediate Event		Intermediate Event	
LM	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4	Event 3	Event 4
0	-0.52 (3.69)	-0.85 (3.63)			-0.50 (3.45)	-0.97 (3.43)		
1	-0.35 (3.85)	-0.99 (3.95)	-1.24 (7.25)	-0.99 (6.37)	-0.48 (4.03)	-0.94 (4.23)	-0.72 (7.86)	-1.07 (7.35)
2	-0.32 (4.27)	-1.07 (4.74)	-0.51 (6.35)	-0.47 (6.39)	-0.46 (4.68)	-0.88 (5.11)	-0.42 (7.23)	-0.76 (7.35)
3	-0.59 (4.76)	-0.77 (5.49)	0.02 (6.45)	-0.39 (7.03)	-0.46 (5.56)	-0.85 (6.25)	-0.15 (7.24)	-0.29 (7.84)
4	-0.90 (6.15)	-0.68 (6.99)	0.42 (7.02)	0.14 (7.63)	-0.57 (7.03)	-0.89 (7.92)	0.27 (7.68)	0.32 (8.43)
5	-0.68 (8.87)	-1.35 (9.51)	0.74 (8.74)	1.31 (9.51)	-0.69 (9.39)	-1.12 (10.15)	0.55 (9.00)	1.04 (9.81)

distinctions can be made concerning these approaches. From the point of view of functionals involved in the construction of the dynamic prediction probabilities, the multi-state approach uses all the transition-specific intensities, which includes the use of the covariate process; dynamic prediction probabilities are derived as complex functionals of these intensities. In contrast, the landmark approaches are concerned directly with the dynamic prediction probabilities and do need not to specify a model for the covariate process. From the point of view of modeling and estimation, the multi-state approach relies on models for the transition intensities; the estimated regression parameters and dynamic prediction probabilities are obtained by means of one regression analysis on the original data. In contrast, the estimated dynamic prediction probabilities from the landmark approaches are obtained either from one single regression analysis (the (stratified) supermodels) or from several regression analyses (the separate landmark models) based only on the subset of the original data which is necessary for dynamic prediction. From the point of view of the underlying assumptions, the multi-state approach requires adapted modeling to whether the process is Markov or semi-Markov. In contrast, the landmark approaches do not need to account for such assumption, because here modeling is not done over the entire follow-up, but only over the prediction intervals (landmark (super)models on cause-specific hazards) or precisely in the prediction time points (landmark (super)models on dynamic pseudo-observations).

The present competing risks example was meant to assess which is the best model in terms of prediction accuracy when one time-dependent covariate is incorporated. It has shown that the landmark models resulted in less biased estimates when the true underlying model does not fulfil the Markov assumption, compared to the multi-state approach. Retrospectively, this is open to debate because the multi-state model used in estimation stays conveniently within the Markov framework where the Aalen-Johansen formula is available for prediction.

The question arises whether the resulting prediction accuracy from this example can be generalized to other settings; a relevant aspect might be the amount of censoring within the prediction interval $[s, s+w]$. We would expect here the landmark model based on dynamic pseudo-observations to perform best, because censored observations within $[s, s+w]$ provide nonparametric estimators of their event status at $s+w$, while in the multi-state approach and in the landmark approach based on cause-specific hazards censored observations do not contribute anymore once they disappear from the risk set.

The unexpected large bias in $\widehat{P}^{\text{sup}}_{h,\text{CS}}$ needs further investigation. Most probably this is caused by the fact that in our particular setting the prediction intervals $[s, s+5]$, $s \in \{0, 1, \ldots, 5\}$ share overall only the time point $t = 5$. This phenomenon could explain why a model over the interval $[0, 5]$ leads to such different predictions than a model for $[5, 10]$. We would expect considerable improvements

in terms of bias of $\widehat{P}^{\text{sup}}_{h,\text{CS}}$ when the prediction intervals $[s, s+w]$, $s \in [s_1, s_K]$, share a larger amount of common information, that is when the prediction intervals overlap over a continuous line.

Bibliography

Aalen, O. O. (1975). Statistical inference for a family of counting processes. Ph.D. thesis, Univiversity of California, Berkeley.

Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observation. *Scandinavian Journal of Statistics* **5**: 141-150.

Andersen, P. K. and Abildstrøm, S. Z. and Rosthøj, S. (2002). Competing risks as a multi-state model. *Statistical Methods in Medical Research* **11**: 203–215.

Andersen, P. K. and Keiding, N.(2012). Interpretability and importance of functionals in competing risks and multistate models. *Statistics in Medicine* **31**: 1074–1088.

Andersen, P. K. and Klein, J. (2007). Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scandinavian Journal of Statistics* **34**: 3–16.

Andersen, P. K. and Klein, J. P. and Rosthøj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika* **90**: 15–27.

Andersen, P. K. and Perme, M. P. (2010). Pseudo-observations in survival analysis. *Statistical Methods in Medical Research* **19**: 71–99.

Anderson, J. R. and Cain, K. C. and Gelber, R. D. (1983). Analysis of survival by tumor response. *Journal of Clinical Oncology* **11**: 710–719.

Arjas, E. and Eerola, M. (1993). On predictive causality in longitudinal studies. *Journal of Statistical Planning and Inference* **34**: 361–386.

Armitage, P. and Colton, T. (2007). Encyclopedia of Biostatistics. New York: John Wiley and Sons.

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

Beyersmann, J. and Schumacher, M. (2008). Time dependent covariates in the proportional subdistribution hazards model for competing risks. *Biostatistics* **9**: 765–776.

Beyersmann, J. and Wolkewitz, M. and Allignol, A. and Grambauer, N. and Schumacher, M. (2011). Applications of multistate models in hospital epidemiology: Advances and challenges. *Biometrical Journal* **53**: 332–350.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30**: 89–99.

de Bruijne, M. H. and Sijpkens, Y.W. and Paul, L.C. and Westendorp, R.G. and van Houwelingen, H.C. and Zwinderman, A.H. (2003). Predicting kidney graft failure using time-dependent renal function covariates. *Journal of Clinical Epidemiology* **56**: 448-455.

Carrol, R. J. and Rupert, D. and Stefanski, L. A. (2006). Measurement Error in Non-linear Models: A Modern Perspective. 2nd edn, Chapman and Hall/CRC, Boca Raton.

Chang, I. and Hsiung, C. A. and Wen, C. C. and Wu, Y. J. and Yang, C. C. (2007). Non-parametric maximum-likelihood estimation in a semiparametric mixture model for competing-risks data. *Scandinavian Journal of Statistics* **34**: 870–895.

Christensen, E. and Schlichting, P. and Andersen, P. K. and Fauerholdt, L. and Schou, G. and Pedersen, B. V. and Juhl, E. and Poulsen, H. and Tygstrup, N. (1983). Updating prognosis and therapeutic effect evaluation in cirrhosis with Cox multiple-regression model for time-dependent variables. *Scandinavian Journal of Gastroenterology* **21**: 163–174.

Clahsen, P. C. and van der Velde, C. J. H. and Julien, J. P. and Floiras, J. L. and Delozier, T. and Mignolet, F. Y. and Sahmoud, T. M. (1996). Improved local control and disease-free survival after perioperative chemotherapy for early-stage breast cancer. *Journal of Clinical Oncology* **14**: 745-753.

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

Cox, D. (1975). Partial likelihood. *Biometrika* **62**: 269–276.

Cox, D. R. and Oakes, D. (1984). Analysis of Survival Data. London: Chapman and Hall.

Cortese, G. and Andersen, P. K. (2010). Competing risks and time-dependent covariates. *Biometrical Journal* **51**: 138–158.

Cortese, G. and Gerds, T. A. and Andersen, P. K. (2013). Comparison of prediction models for competing risks with time-dependent covariates. *Statistics in Medicine* **32**: 3089-3101.

Craiu, R. V. and Duchesne, T. (2004). Inference based on the EM algorithm for competing risks model with masked causes of failure. *Biometrika* **91**: 543–558.

Cummings, F. J. and Gray, R. and Davis, T. E. and Tormey, D. C. and Harris, J. E. and Falkson, G. G. and Arseneau, J. (1986). Tamoxifen versus placebo: double-blind adjuvant trial in elderly women with stage II breast cancer. *National Cancer Institut Monographs* **1**: 119–123.

Dabrowska, D. M. (1995). Estimation of transition probabilities and bootstrap in a semiparametric Markov renewal model. *Journal of Nonparametric Statistics* **5**: 237–259.

Dabrowska, D. M. (1995). Nonparametric regression with censored observations. *Journal of Multivariate Analysis* **54**: 253–283.

Efron, B. (1977). The efficiency of Cox'likelihood function for censored data. *Journal of the American Statistical Association* **72**: 557–565.

Eilers, P. H. C. and Marx, D.(1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**: 89–121.

Fiocco, M. and Putter, H. and van Houwelingen, J. C. (2005). Reduced rank proportional hazards model for competing risks. *Biostatistics* **6**: 465–478.

Fiocco, M. and Putter, H. and van Houwelingen, J. C. (2008). Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models. *Statistics in Medicine* **27**: 4340–4358.

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

Gerds, T. A. and Scheike, T. H. and Andersen, P. K. (2012). Absolute risk regression for competing risks: interpretation, link functions, and prediction. *Statistics in Medicine* **31**: 3921-3930.

Graw, F. and Gerds, T. A. and Schumacher, M. (2009). On pseudo-values for regression analysis in multi-state models. *Lifetime Data Analysis* **15**: 241–255.

Goetghebeur, E. and Ryan, L. (1995). Analysis of competing risks survival data when some failures types are missing. *Biometrika* **82**: 821–833.

Gran, J. M. and Roysland, K. and Wolbers, M. and Didelez, V. and Sterne, J. A. C. and Ledergerber, B. and Furrer, H. and von Wylf, V. and Aalen, O. O. (2010). A sequential Cox approach for estimating the causal effect of treatment in the presence of time-dependent confounding applied to data from the Swiss HIV Cohort Study. *Statistics in Medicine* **29**: 2757–2768.

Gratwohl, A. and Hermans, J. and Goldman, J. M. et al. (1998). Risk assessment for patients with chronic myeloid leukaemia before allogeneic blood or marrow transplantation. *Lancet* **352**: 1087–1092.

Gratwohl, A. and Brand, R. and Frassoni, F. and Rocha, V. and Niederwieser, D. and Reusser, P. and Einsele, H. and Cordonnier, C. (2005). Cause of death after allogeneic haematopoietic stem cell transplantation (HSCT) in early leukaemias: an EBMT analysis of lethal infectious complications and changes over calendar time. *Bone Marrow Transplantation* **36**: 757–769.

Hachen, D. S. Jr. (1988). The competing risks model: a method for analyzing processes with multiple types of events. *Sociological Methods & Research* **17**: 21–54.

van der Hage, J. A. and van der Velde, C. J. H. and Julien, J. P. and Floiras, J. L. and Delozier, T. and Vandervelden, C. and Duchateau, L. (2001). Improved survival after one course of perioperative chemotherapy in early breat cancer patients: long-term results from the European Organization for Research and Treatment of Cancer (EORTC) trial 10854 . *European Journal of Cancer* **37**: 2184-2193.

Harrell, F. E. and Lee, K. L. and Pollock, B. G. (1988). Regression models in clinical studies: determining relationship between predictors and response. *Journal of National Cancer Institut* **80**: 1198-1202.

Hartgrink, H.H. and van De Velde, C.J.H. and Putter, H. and Bonenkamp, J.J. and Klein Kranenbarg, E. and Songun, I. and Welvaart, K. and Van Krieken, J.H. and Meijer, S. and Plukker., J.T. and Van Elk, P. J. and Obertop, H. and Gouma, D.J. and Van Lanschot, J.J. and Taat, C. W. and De Graaf, P.W. and Von Meyenfeldt, M.F. and Tilanus, H. and Sasako, M. (2004). Extended lymph node dissection for gastric cancer: who may benefit? Final results of the randomized dutch gastric cancer group trial. *Journal of Clinical Oncology* **22**: 2069-2077.

Heagerty, P. J. and Lumley, T. and Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**: 337–344.

Hjort, N. L. (1992). On inference in parametric survival data models. *International Statistical Review* **60**: 355–387.

Horn, R. A. and Johnson, C. R. (1991). Topics in Matrix Analysis. *Cambridge University Press*.

Højsgaard, S. and Halekoh, U. and Yan, J. (2006). The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software* **15**: 1–11.

van Houwelingen, H. C. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics* **34**: 70–85.

van Houwelingen, H. C. and Putter, H. (2008). Dynamic predictic by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. *Lifetime Data Analysis* **14**: 447–463.

van Houwelingen, H. C. and Putter, H. (2012). *Dynamic Prediction in Clinical Survival Analysis*. Chapman & Hall. Boca Raton.

Iosifescu, M. (1980). Finite Markov Processes and Their Applications. New York: John Wiley and Sons.

Jackson, C. H. and Sharples, L. D. and Thompson, S. G. and Duffy, S. W. and Couto, E. (2003). Multistate Markov models for disease progression with classification error. *Journal of the Royal Statistical Society, Series D-The Statistician* **52**: 193–209.

Jeong, J. H. and Fine, J. (2006). Direct parametric inference for the cumulative incidence function. *Journal of the Royal Statistical Society Series C-Applied Statistics* **55**: 187–200.

Johansen, S. (1983). An Extension of Cox's Regression Model. *International Statistical Review* **51**: 258–262.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Klein, J. P. and Andersen, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* **61**: 223–229.

Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association* **43**: 457–481.

Klein, J.P. and Keiding, N. and Copelan, E. A. (1993). Plotting summary predictions in multistate survival models - probabilitites in relapse or remission for bone-marrow transplantation patiens. *Statistics in Medicine* **12**: 2315–2332.

Klein, J. P. and Andersen, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* **61**: 223-229.

Klein, J. P. and Gerster, M. and Andersen, P. K. and Tarima, S. and Perme, M. P. (2008). SAS and R functions to compute pseudo-values for censored data regression. *Computer Methods and Programms in Biomedicine* **89**: 289–300.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.

Koller, T. M. and Raatz, H. and Steyenberg, E. W. and Wolbers, M. (2011). Competing risks and the clinical community: irrelevance or ignorance? *Statistics in Medicine* **68**: 547–548.

Kuk, A. Y. C. (1992). A semiparametric mixture model for the analysis of competing risks data *Australian & New Zealand Journal of Statistics* **34**: 169-180.

Kurland, B. F. and Heagerty, P. J. (2005). Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics* **6**: 241–258.

Larson, M. G. and Dinse, G. E. (1985). A mixture model for the regression-analysis of competing risks data. *Journal of the Royal Statistical Society Series C-Applied Statistics* **2**: 201–211.

Latouche, A. and Boisson, V. and Chevret, S. and Porcher, R. (2007). Misspecified regression model for the subdistribution hazard of a competing risk. *Statistics in Medicine* **26**: 965–974.

Latouche, A. and Porcher, R. and Chevret, S. (2005). A note on including time-dependent covariate in regression model for competing risks data. *Biometrical Journal* **47**: 807–814.

Lau, B. and Cole, S. R. and Moore, R. D. (2008). Evaluating competing adverse and beneficial outcomes using a mixture model. *Statistics in Medicine* **27**: 4313–4327.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Second Edition. John Wiley and Sons. Hoboken, New Jersey.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**: 13-22.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* **84**: 1074-1078.

Little, R. J. A. and Rubin, D. B. (1987). Statistical Analysis with Missing Data. New York: John Wiley & Sons.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society Series B* **44**: 226–233.

Lu, W. and Liang, Y. (2008). Analysis of competing risks data with missing cause of failure under additive hazards model. *Statistica Sinica* **18**: 219–234.

Lu, W. and Peng, L. (2008). Semiparametric analysis of mixture regression models with competing risks data. *Lifetime Data Analysis* **14**: 231–252.

Lu, K. and Tsiatis, A. A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing causes of failure. *Biometrics* **57**: 1191–1197.

Lunn, M. and McNeil, D. (1995). Applying Cox Regression to Competing Risks. *Biometrics* **51**: 524–532.

Madsen, E. B. and Hougaard, P. and Gilpin, E. (1983). Dynamic evaluation of prognosis from time-dependent variables in acute myocardial-infraction. *American Journal of Cardiology* **51**: 1579–1583.

McCarthy, P. L. and Hahn, T. (2011). Self-determinism: alloBMT for AML. *Blood* **117**: 2080–2081.

McCullagh, P. and Nelder, J. A. (1999). *Generalized Linear Models*. Chapman & Hall/CRC.

Nelson, W. (1969). Hazard plotting for incomplete failure data. *Journal of Quality Technology* **1**: 27–52.

Ng, S. K. and McLachlan, G. J. (2003). An EM-based semi-parametric mixture model approach the the regression analysis of competing risks data. *Statistics in Medicine* **22**: 1097–1111.

Nicolaie, M. A. and van Houwelingen, J. C. and Putter, H. (2010). Vertical modeling: a pattern mixture approach for competing risks modeling. *Statistics in Medicine* **29**: 1190-1205.

Nicolaie, M. A. and van Houwelingen, J. C. and Putter, H. (2011). Vertical modeling: analysis of competing risks data with missing causes of failure. *Statistical Methods in Medical Research. In print.*

Nicolaie, M. A. and van Houwelingen, H. C. and de Witte, T. M. and Putter, H. (2013a). Dynamic prediction in competing risks by landmarking. *Statistics in Medicine* **32**: 2031–2047.

Nicolaie, M. A. and van Houwelingen, H. C. and de Witte, T. M. and Putter, H. (2013b). Dynamic pseudo-observations: a robust appraoch to dynamic prediction in competing risks. *Biometrics, DOI:10.1111/biom.12061.*

Nicolaie, M. A. and van Houwelingen, H. C. and Putter, H. (2013c). Comparison of dynamic prediction methods in competing risks. *Submitted manuscript.*

Parast, L. and Cheng, S. C. and Cai, T. (2011). Incorporating short-term outcome information to predict long-term survival with discrete markers. *Biometrical Journal* **53**: 294-307.

Park, C. (2005). Parameter estimation of incomplete data in competing risks using the EM algorithm. *IEEE Transactions on Reliability* **54**: 282–290.

Pepe, M. S. and Heagerty, P. J. and Whitaker, R. (1999). Prediction using partly conditional time-varying coefficients regression models. *Biometrics* **55**: 944–950.

Perperoglou, A. and le Cessie, S. and van Houwelingen, J. C. (2006). Reduced-rank hazard regression for modelling non-proportional hazards. *Statistics in Medicine* **25**: 2831–2845.

Prentice, R. L. and Kalbfleisch, J. D. and Peterson, A. V. and Flournoy, N. and Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**: 541–54.

Proust-Lima, C. and Taylor, J. M. G. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeteated measures of posttreatment PSA: a joint modeling approach. *Biostatistics* **10**: 535–549.

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

Putter, H. and Sasako, M. and Hartgrink, H. H. and van de Velde, C. J. H. and van Houwelingen, J. C. (2005). Long-term survival with non-proportional hazards: results from the Dutch Gastric Cancer Trial. *Statistics in Medicine* **24**: 2807–2821.

Putter, H. and van der Hage, J. and de Bock, G. H. and Elgata, R. and van de Velde, C. J. H. (2006). Estimation and prediction in a multi-state model for breast cancer. *Biometrical Journal* **48**: 366–380.

P. M. Ravdin and L. A. Siminoff and G. J. Davis (2001). Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *Journal of Clinical Oncology* **19**: 980–991.

Rizopoulos, D.(2011) Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time-to-Event Data. *Biometrics* **67**: 819–829.

van Rompaye, B and Goetghebeur, E. and Jaffar, S. (2010). Design and testing for clinical trials faced with misclassified causes of death. *Biostatistics* **11**: 546–558.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**: 581-592.

Sabatier, R. and Eymard, J. C. and Walz, J. and Deville, J. L. and Narbonne, N. and Boher, J. M. and Salem, N. and Marcy, M. and Brunelle, S. and Viens, P. and Bladou, F. and Gravis, G. (2012). Could thyroid dysfunction influence outcome in sunitinib-treated metastatic renal cell carcinoma? *Annals of Oncology* **23**: 714–721.

Schumaker, L. L.(1981). *Spline Functions: Basic Theory*. New York: John Wiley and Sons.

Scheike, T. M. and Zhang, M.-J. and Gerds, T. A. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika* **95**: 205-220.

Smits, J. and van Houwelingen, H. and De Meester, J. and le Cessie, S. and Persijn, G. and Claas, F. H. J. and Frei, U. (2000). Permament detrimental effect of nonimmunological factors on long-term renal graft survival: a parsimonious model of time-dependency. *Transplantation* **70**: 317–323.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73**: 363–369.

Tai, B. C. and White, I. R. and Gebski, V. and Machin, D. (2002). On the issue of 'multiple' first failures in competing risks analysis. *Statistics in Medicine* **21**: 2243–2255.

Xu, R. and O'Quigley, J. (2000). Estimating average regression effect under non-proportional hazards. *Biostatistics* **1**: 423–439.

Zamboni, B. A. and Yothers, G. and Choi, M. and Fuller, C. D. and Dignam, J. J. and Raich, P. C. and Thomas Jr, C. R. and O'Connell, M. J. and Wolmark, N. and Wang, S. J. (2010). Conditional Survival and the Choice of Conditioning Set for Patients With Colon Cancer: An Analysis of NSABP Trials C-03 Through C-07. *Journal of Clinical Oncology* **28**: 2544–2548.

Zheng, Y. and Heagerty, P. J. (2005). Partly conditional survival models for longitudinal data. *Biometrics* **61**: 379–391.

P. W. Wilson and R. B. D'Agostino and D. Levy and A. M. Belanger and H. Silbershatz, W. B. Kannel (1998). Prediction of coronary heart disease using risk factor categories. *Circulation* **97**: 1837-1847.

de Wreede, L. and Fiocco, M. and Putter, H. (2010). The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Computer Methods and Programs in Biomedicine* **99**: 261–274.

Nederlandse samenvatting

Hoofdstuk 1 geeft een algemene introductie tot analysemethoden voor overlevingsduurgegevens met zogenaamde concurrerende risico's (competing risks). De belangrijkste concepten worden geïntroduceerd en enkele reeds bestaande methodes voor de statistische analyse van dergelijke data worden behandeld. Er is een grote hoeveelheid literatuur over overlevingsduurgegevens gewijd aan modelbouw en testen. Het doel van dit hoofdstuk is om te laten zien hoe de standaard methodes voor overlevingsduurgegevens zijn aangepast voor gegevens met concurrerende risico's. De overige hoofdstuken gaan in detail in op een aantal specifieke problemen opgeworpen in de introductie.

Hoofdstuk 2 en 3 zijn gewijd aan de analyse van gegevens met concurrerende risico's, in het bijzonder voor het geval waarin voor sommige individuën de oorzaak van falen ontbreekt. Een nieuwe aanpak voor concurrerende risico's wordt geïntroduceerd, "vertical modeling"genaamd, en de belangrijkste eigenschappen worden beschreven. Hoofdstuk 2 presenteert de wiskundige eigenschappen van de methode; expliciete uitdrukkingen worden gegeven door de variantie van de oorzaak-specifieke cumulatieve incidentie functie, verkregen door middel van de delta-methode. Eigenschappen van de schatters worden bestudeerd in simulatiestudies en vergeleken met niet-parametrische schatters. Een aantrekkelijke eigenschap van onze methode is dat het gaat om natuurlijk observeerbare entiteiten in concurrerende risico's; dit zorgt ervoor dat de parameters makkelijk interpreerbaar zijn. De methode vangt bijvoorbeeld het patroon van faaloorzaken in de tijd. Dit hoofdstuk bevat ook een analyse van echte data, waaruit de praktische toepasbaarheid van de methode blijkt.

Aangezien het in de praktijk lastig kan zijn om concurrerende risico's data te verkrijgen die volledig is, is het belangrijk om in staat te zijn om ook in minder optimale maar meer reële omstandigheden te kunnen opereren. Hoofdstuk 3 laat nog een aantrekkelijke eigenschap zien van vertical modeling, namelijk dat de methode om kan gaan met concurrerende risico's waarin missende faaloorzaken voorkomen. Onder bepaalde redelijke aannames wordt deze situatie op een natuurlijke manier aangepakt door vertical modeling, omdat deze methode alle informatie op het moment van falen gebruikt (ook van de individuen waarvan alleen het moment en niet de oorzaak van falen bekend is), en ook gedeeltelijke

informatie over de precieze oorzaak van falen optimaal gebruikt. De vertical modeling aanpak resulteert in juiste inferentie; de maximum likelihood schatters van de regressie parameters gebaseerd op de volledige likelihood vallen samen met de maximum likelihood parameters verkregen met onze aanpak. Andere voordelen van onze methode ten opzichte van sommige al bestaande methodes worden besproken en geïllustreerd aan de hand van analyses op echte data.

In hoofdstuk 4 wordt een nieuwe aanpak voorgesteld om dynamische voorspellingen te kunnen doen binnen het kader van concurrerende risico's, een aanpak die gezien kan worden als een uitbreiding van de landmark aanpak voor gewone overleving. De aanpak is gebaseerd op het combineren van Cox proportionele hazards modellen van oorzaak-specifieke hazards voor het cohort van overlevers per landmark tijdstip in zogenaamde "supermodellen". Supermodellen, verkregen door het combineren van de oorzaak-specifieke baseline hazards voor een reeks landmark tijdstippen, kunnen omgaan met tijdsvariërende effecten van baseline covariaten of tijdsafhankelijke covariaten, tegelijkertijd rekening houdend met meerdere faaloorzaken. Het schatten van de regressie parameters wordt gedaan door middel van pseudo partiële likelihood. Het voordeel van de methode is dat zij het modelleren vergemakkelijkt, omdat ze de dynamische voorspelkansen direct modelleert; op deze manier wordt alleen de informatie die essentieel is voor het voorspellen gedestilleerd uit het complexe onderliggende proces. De methode wordt empirisch gevalideerd op een echte dataset.

In hoofdstuk 5 wordt een alternatief geboden voor de in hoofdstuk 4 beschreven methode. De nieuwe aanpak is direct gericht op het tijdstip waarop de dynamische voorspelling gewenst is, en gebruikt dus niet het volledige predictie interval raamwerk zoals de methode uit hoofdstuk 4. De voornaamste ingrediënten voor de methode zijn pseudo-observaties die zijn berekend voor het cohort overlevenden per landmark tijdstip, de zogenaamde "dynamische pseudo-observaties". Ze worden samengevoegd in supermodellen voor een reeks landmark tijdstippen door middel van een GLM aanpak, waarin gladgestreken tijdsvariërende effecten van covariaten of tijdsafhankelijke covariaten hun gemiddelde waarden kunnen beïnvloeden. Het schatten wordt gedaan door middel van een GEE methode. We beschrijven de wiskundige eigenschappen van onze methode aangaande het asymptotisch gedrag van de schatters. Het voordeel van het op deze manier modelleren van een enkel tijdstip is de robuustheid tegen model-misspecificaties die snel optreden wanneer men ingewikkelder methoden gebruikt. Onze methode wordt geïllustreerd op een echte data analyse.

Hoofdstuk 6 gaat door middel van simulatie studies in op de eigenschappen van verschillende dynamische voorspelmethodes binnen het concurrerende risico's kader, waaronder de methoden zoals beschreven in hoofdstukken 4 and 5. De focus ligt niet op het passen van het model, maar op de accuraatheid als het gaat om het voorspellen van een toekomstige gebeurtenis. Twee hoofdscenarios wor-

den bekeken; ter eerste het scenario waarin het ware onderliggende stochastische model voldoet aan de Markov eigenschap, ten tweede een scenario waarin niet aan deze eigenschap wordt voldaan.

List of Publications

1. A. Nicolaie. A sufficient condition for the convergence of a finite Markov chain, Math. Rep.(Bucur.), 10(60), 2008, 57-71.
2. A. Nicolaie. On the asymptotic behaviour of finite Markov chains, Carpathian J. Math. 24(2008), 2, 91-100.
3. A. Nicolaie. A generalization of a result concerning the asymptotic behavior of finite Markov chains, Statistics and Probability Letters 18(78), 2008, 3321-3329.
4. M. N. Pascu, A. Nicolaie. On a discret version of the Laugesen-Morpurgo conjecture, Statistics and Probability Letters 6(79), 2009, 797-806.
5. M.A. Nicolaie, Hans van Houwelingen and Hein Putter. Vertical modeling: a pattern mixture approach for competing risks data, Statistics in Medicine, 2010, 29 (11), 1190-1205.
6. M.A. Nicolaie. Competing risks as purged Markov chains, Bulletin of the Transilvania University of Braşov, Vol 3(52) - 2010, Series III: Mathematics, Informatics, Physics, 67-70.
7. M.A. Nicolaie. Hans van Houwelingen and Hein Putter. Vertical modeling: analysis of competing risks data with missing causes of failure. Statistical Methods in Medical Research (2011). DOI: 10.1177/0962280211432067.
8. M.A. Nicolaie. Analysis of missing causes of failure in competing risks in the context of the Larson and Dinse approach. In Mathematical Modelling of Environmental and Life Sciences Problems , 2010, MatrixRom. ISBN: 978-973-755-772-8.
9. M .A. Nicolaie, Hans van Houwelingen, T.M de Witte, Hein Putter. Dynamic prediction in competing risks by landmarking. Statistics in Medicine 32 (12), 2031-2047, 2013.
10. M.A. Nicolaie, Hans van Houwelingen and Hein Putter. Dynamic pseudo-observations: a robust approach to dynamic prediction in competing risks. Biometrics. DOI: 10.1111/biom.12061.

11. M.A. Nicolaie, Hans van Houwelingen and Hein Putter. Comparison of dynamic prediction models. Submitted.

12. M.A. Nicolaie, Catherine Legrand. Vertical modeling: analysis of competing risks data with a cured fraction. Submitted.

13. M.A. Nicolaie, Catherine Legrand. Vertical modeling: analysis of multistate data with a cured fraction. Working paper.

Curriculum vitae

Mioara Alina Nicolaie was born on the 20th of November 1977, în Braşov, Romania. After finishing her secondary education at Liceul de Ştiinţe ale Naturii C. D. Nenitescu, Braşov, she studied Mathematics at the Faculty of Sciences of Transilvania University of Braşov, where she graduated as B.Sc. in 2000 at the top of her class and as M.Sc on an extreme value theory thesis in 2002. Between 2002 and 2007 she was PhD student at the Faculty of Mathematics. Besides, she jointed several research projects in the field of probability theory. She got her first PhD title on December 14, 2012 with a thesis entitled *Convergence and Monotony Theorems for Finite Markov Chains* under the supervision of Prof.dr. Gabriel V. Orman.

In 2008, she started her second PhD studies at the Department of Medical Statistics and Bioinformatics, Leiden University Medical Center. She worked in the project entitled *Prognostic Modeling and Dynamic Prediction for Competing Risks and Multi-State Models* under the supervision of Prof.dr. Hein Putter and Em. Prof.dr. Hans van Houwelingen. Her work was focused on the development of new analysis strategies for competing risks. The results of this research are presented in this thesis.

Since 2012, she works as post-doctoral fellow at the Institute of Statistics, Biostatistics and Actuarial Sciences at the Catholic University of Louvain, Belgium. She develops new statistical methodology for multi-state models with a cured fraction.

Acknowledgments

The research presented in this thesis is the result of my work in the Department of Medical Statistics and Bioinformatics of Leiden University Medical Center. The first thoughts of gratitude I have are for my supervisors Prof. Hans van Houwelingen and Prof. Hein Putter. Prof. Hans van Houwelingen is the father of biostatistics in the Netherlands. He created a new school of thought in medical statistics by translating results from theoretical statistics to applications in medical research. His innovations are characterized by usefulness, deepness, elegance and wideness and they have greatly influenced the international community of (clinical) biostatisticians. Prof. Hein Putter, having given me the opportunity to pursue this PhD, helped me to improve my research skills, to gain a technical writing style and he inspired me on how to tackle a long list of scientific challenges concerning the passage from the theoretical field of research of Markov chains to the more applied one of competing risks. My supervisors will always be for me reference models as researchers, professors and human beings.

Leiden University Medical Center is a great establishment which has acquired genuine researchers from all over the world. Many thanks go to my colleagues from whom I benefited a lot: Marta, Rosa, Roula, Liesbeth, Wendim, Andrea, Bruna, Lies, Henk Jan, Watze, Erik, Jelle, Ron, Ronald, Bart, Saskia, Theo, Anika, Jeanine, Hae Won, Tekla, Stefan, Oksana, Fabrice, Cristian, Leila, Aldo and Livio. Romanians complete the list: Dana, Irina, Cristina, Cristina, Crina, Dana, Andrei, Răzvan and Sandrin.

My best friend Olivia helped me a lot with her unceasing presence when needed, her devoted and kind heart.

My Belgian parents, Father Ioan and Presbytera Christina genuinely inspired me through their living example of how great human values can be preserved in the context of an extremely diverse culture and far away from home.

I am grateful to my family. Mămucă, îţi mulţumesc pentru devotamentul şi dragostea ta. Iulia and Emi, I hope you will share my joy on the defense day. Tătucă şi Ucu, sper că acolo unde v-aţi dus, vă bucuraţi de această zi din viaţa mea.

Alina Nicolaie Genval, Belgium