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.

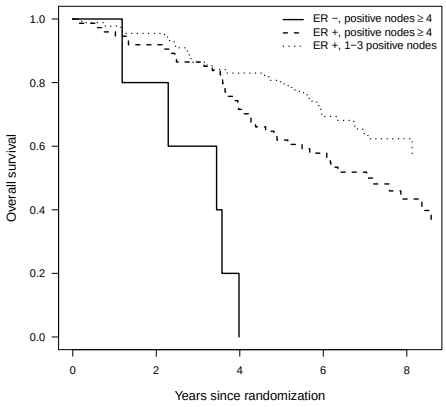


Figure 3.1: Kaplan-Meier survival curves of time to death for each combination of covariates.

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

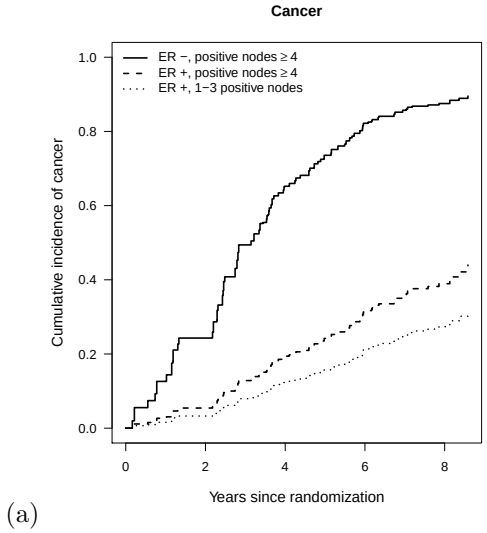
Table 3.3: A proportional hazards model.

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

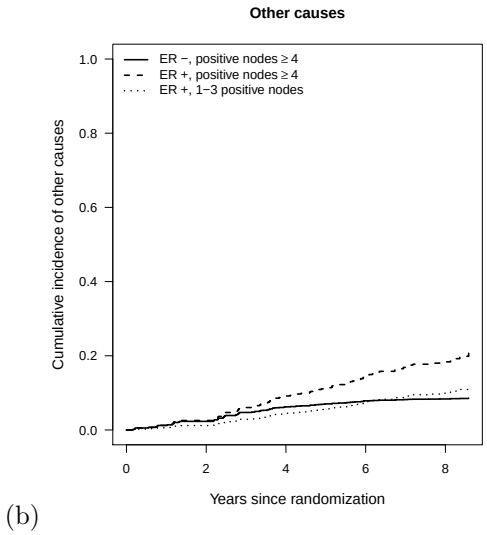
cancer and other causes for each combination of covariate values following from this model, calculated by means of the `mstate` package in `R` (de Wreede et al., 2010). For the combination 1-3 positive nodes and negative estrogen receptor status both cumulative incidence functions are identically zero.

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.

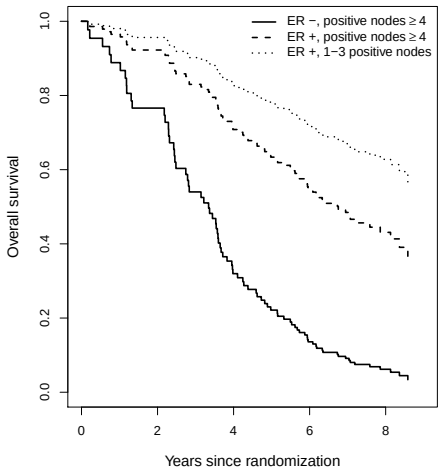


Figure 3.3: Model based survival curves of time to death for each combination of covariates.

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

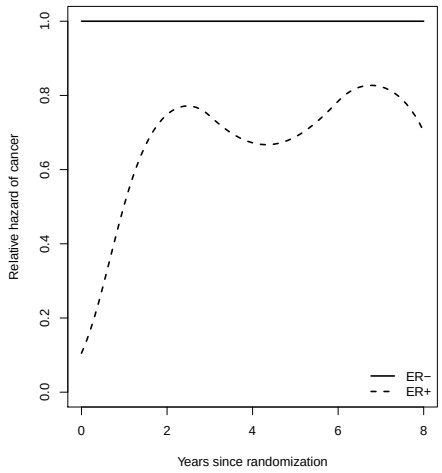


Figure 3.4: The relative hazards of death due to cancer for ER- and ER+.

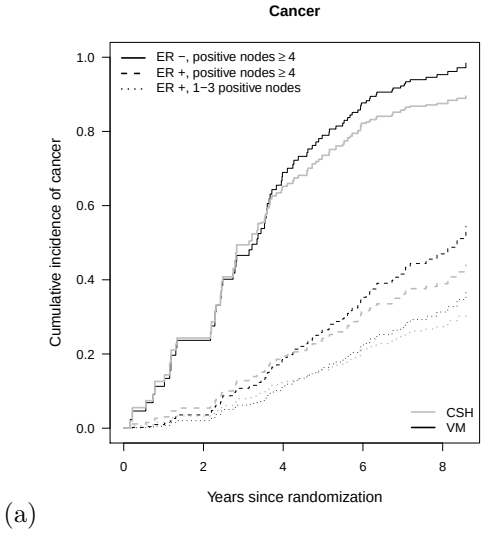
explained by the fact that in the Goetghebeur and Ryan approach occurrence of an event impacts both cause-specific hazards at the same time, via the common baseline hazard $\lambda_{10}(t)$ they share. Vertical modeling acknowledges the fact that no deaths due to other causes were observed for the subgroup 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.

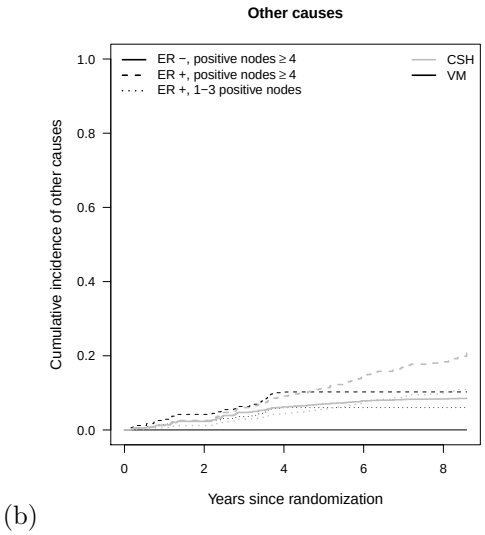
3.5 Vertical modeling in context

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

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

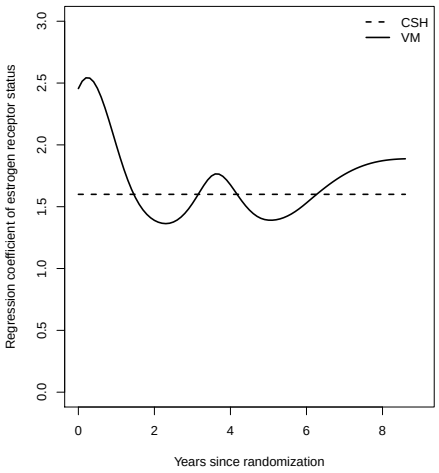


Figure 3.6: Implied time-varying regression coefficient of estrogen receptor status on the cause specific hazard of death due to cancer from Goetghebeur and Ryan's approach (CSH) versus vertical modeling (VM).

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.