HAL
open science

Semiparametric inference of competing risks data with additive hazards and missing cause of failure under MCAR or MAR assumptions

Laurent Bordes, Jean-Yves Dauxois, Pierre Joly

▶ **To cite this version:**

Laurent Bordes, Jean-Yves Dauxois, Pierre Joly. Semiparametric inference of competing risks data with additive hazards and missing cause of failure under MCAR or MAR assumptions. Electronic Journal of Statistics , 2014, 8, pp.41-95. 10.1214/14-EJS876 . hal-00501865v3

HAL Id: hal-00501865

https://hal.science/hal-00501865v3

Submitted on 22 Jan 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Electronic Journal of Statistics
Vol. 0 (2014) 1–56
ISSN: 1935-7524

Semiparametric inference of competing risks data with additive hazards and missing cause of failure under MCAR or MAR assumptions

Laurent Bordes[†]

Université de Pau et des Pays de l'Adour - France
Laboratoire de Mathématiques et de leurs Applications - Pau
UMR CNRS 5142 & FR CNRS 2952
e-mail: laurent.bordes@univ-pau.fr

Jean-Yves Dauxois[†,*]

Université de Toulouse-INSA - France
Institut de Mathématiques de Toulouse
UMR CNRS 5219
e-mail: jean-yves.dauxois@insa-toulouse.fr

Pierre Joly

Université de Bordeaux, ISPED- France
Centre Inserm U897, Bordeaux
e-mail: pierre.joly@isped.u-bordeaux2.fr

Abstract: In this paper, we consider a semiparametric model for lifetime data with competing risks and missing causes of death. We assume that an additive hazards model holds for each cause-specific hazard rate function and that a random right censoring occurs. Our goal is to estimate the regression parameters as well as the functional parameters such as the baseline and cause-specific cumulative hazard rate functions / cumulative incidence functions.

We first introduce preliminary estimators of the unknown (Euclidean and functional) parameters when cause of death indicators are missing completely at random (MCAR). These estimators are obtained using the observations with known cause of failure. The advantage of considering the MCAR model is that the information given by the observed lifetimes with unknown failure cause can be used to improve the preliminary estimates in order to attain an asymptotic optimality criterion. This is the main purpose of our work. However, since it is often more realistic to consider a missing at random (MAR) mechanism, we also derive estimators of the regression and functional parameters under the MAR model. We study the large sample properties of our estimators through martingales and empirical process techniques. We also provide a simulation study to compare the behavior of our three types of estimators under the different mechanisms

*Corresponding Author

†Work supported by the French National Agency of Research (ANR), project AMMSI, number ANR-2011-BS01-021.

of missingness. It is shown that our improved estimators under MCAR assumption are quite robust if only the MAR assumption holds. Finally, three illustrations on real datasets are also given.

MSC 2010 subject classifications: Primary 62G05, 62N01, 62N02; secondary 60F17, 60G15, 60G44, 60G55

.

Keywords and phrases: additive hazards, competing risks, cause-specific cumulative hazard rate function, counting process, cumulative incidence function, large sample behavior, missing indicator, missing at random, missing completely at random, regression parameter, survival analysis, Reliability

.

Contents

1 Introduction . . . 3
2 Framework under MCAR assumption . . . 5
3 Statistical Inference on the Euclidean parameters under MCAR . . . 8
3.1 Estimators . . . 8
3.2 Solving problems (P_i) . . . 10
4 Asymptotic behavior of the regression parameters estimators under MCAR . . . 11
4.1 Additional notations and assumptions, preliminary results . . . 11
4.2 Large sample behavior of $\hat{\alpha}$ . . . 15
4.3 Large sample behavior of $\hat{\beta}$ . . . 17
4.4 The optimal estimator $\tilde{\beta}$ . . . 18
5 Statistical inference on the functional parameters under MCAR . . . 20
5.1 Inference in the nonparameteric model . . . 20
5.1.1 First estimators of the cause-specific cumulative hazard rate functions . . . 21
5.1.2 Optimal estimators of the cause-specific cumulative hazard rate functions . . . 24
5.1.3 Estimation of the cumulative incidence functions and the survival function of T . . . 26
5.2 With explanatory variables . . . 28
5.2.1 First estimators of the cause-specific cumulative hazard rate functions: the WMC estimators . . . 29
5.2.2 Improved estimators of the cause-specific cumulative hazard rate functions: the T-optimal estimators . . . 32
6 Semiparametric inference under the MAR Assumption . . . 35
6.1 MAR Assumption . . . 36
6.2 Parametric model on the MAR mechanism . . . 36
6.3 The double robust estimators of β_j and $\Lambda_{0j}(\cdot)$ . . . 37
6.4 Large sample behavior . . . 39
7 Numerical study . . . 43
7.1 Simulation study . . . 43
7.1.1 Simulation design . . . 43

7.1.2 Regression parameter estimators 44
7.1.3 Estimators of the baseline cause-specific cumulative hazard functions 48
7.2 Applications on real datasets 49
7.2.1 Hodgkin's disease dataset 49
7.2.2 Breast Cancer dataset 50
7.2.3 Hard drive dataset 51
Acknowledgments 53
References 54

1. Introduction

The competing risks models are useful in Survival Analysis and Reliability in order to take into account the different causes of death of a patient or the of failure of a device. The observations usually include the (possibly censored) lifetime and the indicator of cause of death. Sometimes, a vector of (time dependent) covariates is also available. Based on such data, one can carry out a statistical inference assuming a parametric, semi-parametric or nonparametric model. One can find an extensive review of these models and their statistical inference in, for example, Crowder (2001) or Andersen *et al.* (1993).

In practice, it may happen that the cause of death or failure is missing for some individuals. Many papers have considered the problem of missing information in competing risks models. We refer to Miyakawa (1984), Usher and Hodgson (1988), Lin *et al.* (1993), Schabe (1994), Goetghebeur and Ryan (1995), Guttman *et al.* (1995), Reiser *et al.* (1996), Basu *et al.* (1999), Lu and Tsiatis (2001), Flehinger *et al.* (2002), Craiu and Duchesne (2004), Gao and Tsiatis (2005), Craiu and Reiser (2007), Lu and Liang (2008), Bakoyannis *et al.* (2010) and Lee et al. (2011).

Many authors have developed methods for accurate modeling of the missingness mechanism. Some of these works are based on parametric models. When a latent variable represents the missingness mechanism, an EM-type algorithm can be used to estimate the model parameters. In Craiu and Duchesne (2004), such a procedure is proposed when the missingness mechanism depends both on the failure cause and the failure time. Recently, Craiu and Reiser (2007) considered a very complete parametric model including dependence with the failure causes.

The special case of a possibly censored single failure cause differs from the competing risks model by the fact that, in the former case, the censoring is not an event of interest and is (usually) supposed to be independent of the lifetime. However, when the censoring information is missing, we are close to the competing risks situation where failure causes are possibly missing. Some specific inferential methods have been derived for various models with or without covariates and several missingness mechanisms, see *e.g.* Gijbels *et al.* (1993), McKeague and Subramanian (1998), van der Laan and McKeague (1998), Zhou and Sun (2003), Subramanian (2004) and Song *et al.* (2010).

Other authors have developed estimation procedures in a semi or nonparametric framework for two or more failure causes with missing indicators (see *e.g.* Myakawa, 1984, Dinse, 1986, Lo, 1991, Schabe, 1994). In particular, the case of a proportional hazards model for the cause-specific hazard rate functions has been studied by Goetghebeur and Ryan (1995), Lu and Tsiatis (2001) and, more recently, by Lee et al. (2011). On the other hand, Gao and Tsiatis (2005) have considered a linear transformation competing risks model whereas Bakoyannis *et al.* (2010) have focused on the well-known Fine and Gray (2009) model.

As far as we know, Lu and Liang (2008) is the only paper which considers an additive risk model for competing risks data with missing indicator. The missing indicator situation is an interesting alternative to the proportional hazards model. Lu and Liang (2008) have assumed such a model on only one of the cause-specific hazard rate functions, namely the cause of primary interest. Using estimating equations based on inverse probability weighted (IPW) and double robust (DR) techniques, they are able to estimate the regression parameters. In their model, as in Bakoyannis *et al.* (2010), Lee et al. (2011) and Gao and Tsiatis (2005), the mechanism of missingness may depend on the failure time, which corresponds to the missing at random (MAR) assumption.

Our aim in this paper is to consider an additive risk model for competing risks data when indicators of cause of death are missing at random. Our model differs from the one of Lu and Liang (2008) in the sense that we assume an additive risk model for each cause-specific hazard rate function and not just for the cause of interest. We do not assume independence between the causes of death. It has to be noted that, unlike in Goetghebeur and Ryan (1995), these functions are not supposed to be linked in our model. Furthermore, we are interested in estimating not only the regression parameters but also the cumulative incidence functions and, finally, the overall survival function. Note that Lu and Liang (2008) do not consider the problem of estimating the cumulative incidence function and by providing an estimator of the survival functions associated to the cause-specific hazard rate functions of interest implicitly assume independence between the various causes of death. We believe that obtaining estimation of the cumulative incidence functions is also of interest due to the importance of these functions in a competing risks model. As far as we know, the only paper which considers estimation of the cumulative function when indicators are missing is Bakoyannis *et al.* (2010). But, as noted earlier, their work is under a different model, since they assume a Fine and Gray model.

In this paper, we first consider the case of the most restrictive mechanism of missing information, namely the missing completely at random (MCAR) mechanism. We show that the advantage of this type of missingness mechanism lies in the fact that it doesn't affect the additive shape of the cause-specific hazard rate functions. Thus, one can estimate the regression parameters following the idea of estimating equations introduced by Lin and Ying (1994). We also show that the observations with unknown failure cause are usable to estimate the sum of the previous parameters. Then, and we think that this is the main contribution of our approach, we develop a method that allows us to take into

account this information in an optimal way in order to improve our preliminary estimators of the regression parameters. The improved estimators are built in order to attain an asymptotic efficiency criterion. Such an improvement does not appear to be possible with any other missingness mechanism. We are also able to derive improved estimators of the cumulative hazard rate and the cumulative incidence functions.

We also consider the case of the MAR mechanism. We show that the Double Robust technique, introduced by Robins *et al.* (1994) and later used by Lu and Liang (2008) in a context similar to ours, can be used to derive estimates of the Euclidean and functional parameters of our model. Our work in this part extends the results of Lu and Liang (2008) in the same manner as described earlier: incorporating additive risks for each cause-specific hazard rate function, not making any assumption of independence between the causes, estimating the cause-specific cumulative hazard rate functions, etc...

The paper is organized as follows. In Section 2, the model and the MCAR assumptions are introduced and seen as a specific case of a nonhomogeneous Markov process. In Section 3, first estimators under MCAR of the Euclidean parameters are obtained using estimating equations. Then, we introduce our improved estimators, still under MCAR, which have minimum asymptotic variance. In Section 4, we show that our parameter estimators are consistent and asymptotically Gaussian. Section 5 deals with the estimation of the functional parameters under MCAR assumption. Initial and improved estimators are introduced and we prove their consistency and their asymptotic Gaussian behavior. Consistent estimators of the asymptotic variances are also provided in each section. Note that our proofs of the asymptotic results in Sections 4 and 5 are based on the martingale theory and are, thus, very different from the methods commonly used (see, *e.g.* Gao and Tsiatis, 2005, and Lu and Liang, 2008) in the literature on this topic. Section 6 is devoted to the consideration of the MAR model. We introduce estimators of the Euclidean and functional parameters under this model and obtain their large sample behaviors. Finally, a Monte Carlo study is performed in Section 7 in order to assess the behavior of our estimators for finite sample sizes and to compare their properties under the MCAR and MAR models. In addition, three examples using real data sets illustrate our estimation methods.

2. Framework under MCAR assumption

Consider a population in which each individual is liable to die from any of $p \geq 2$ causes. The causes are not necessarily independent but we assume that each death is due to a single cause. Let us denote by T the individual lifetime and $d \in \{1, \ldots, p\}$ its cause of death. Suppose that our interest focuses on the effect of a time-varying covariate vector $Z(\cdot)$ of dimension k. More precisely, let $\mathbb{Z}(t) = \{Z(u); u \leq t\}$ be the history up to time t of this covariate process and assume that an additive hazard model holds on the cause-specific hazard rate

function, that is:

$$\lambda_j(t|\mathbb{Z}) = \lim_{h\to 0^+} \frac{1}{h} P(T \in [t, t+h[, d = j | T \geq t, \mathbb{Z}(t)) = \lambda_{0j}(t) + \beta_j^T Z(t), \quad t \geq 0, \tag{2.1}$$

for $j = 1, \ldots, p$ where $\lambda_{0j}(\cdot)$ is the baseline jth cause-specific hazard rate function and $\beta_j \in \mathbb{R}^k$ is the vector of regression parameters associated to the jth cause.

Suppose also that the lifetime T is right-censored by a random variable (r.v.) C and write

$$\begin{cases} X &= T \wedge C \equiv \min(T, C) \\ \delta &= I(T \leq C) \end{cases},$$

where $I(\cdot)$ is the set indicator function. Let $\lambda_C(\cdot)$ denote the hazard rate function of the r.v. C and assume that conditionally on $\mathbb{Z}$, the r.v. C is independent of (T, d).

Of course, the future cause of death is not known if C is observed instead of T. But, in some situations, it may happen that d is also not known even if T is observed. Let R denote the missingness indicator, i.e. $R = 1$ if the cause is known and $R = 0$ otherwise. Thus, we are in a situation where the available observation for an individual is

$$(X, \delta, D, \mathbb{Z}(X)),$$

where $D = \delta R d$ reveals the failure cause d when the failure time is uncensored ($\delta = 1$) and $R = 1$ and is equal to zero otherwise. In the following, we assume that the missing mechanism is such that:

$$P(R = 1 | X, \mathbb{Z}, d, \delta = 1) = P(R = 1 | \delta = 1) = \alpha \in [0, 1],$$

where α is an unknown parameter, and

$$P(R = 0 | X, \mathbb{Z}, d, \delta = 0) = P(R = 0 | \delta = 0) = 1.$$

This is the Missing Completely At Random (MCAR) assumption.

One can see the observation of the vector (X, δ, D), conditionally on $\mathbb{Z}$, as the realization of a $(p+3)$-states nonhomogeneous Markov process (see Fig. 1) with space set $\{0, 1, \ldots, p, m, c\}$: state 0 is the initial state; state i, for $i = 1, \ldots, p$, corresponds to the observation of the lifetime T with known cause of death i; state m to the observation of the lifetime with missing cause; state c to a censored observation. Except 0, all the states are absorbing states. From the assumptions on M and the independence between (T, d) and C, conditionally on Z, one can easily get that the instantaneous transition rates of this Markov process are, conditionally on $\mathbb{Z}$:

$$\begin{cases} \lambda'_j(t|\mathbb{Z}) &= \alpha \lambda_j(t|\mathbb{Z}), \text{ for } j \in \{1, \ldots, p\}, \\ \lambda'_m(t|\mathbb{Z}) &= (1-\alpha) \sum_{j=1}^p \lambda_j(t|\mathbb{Z}), \\ \lambda'_c(t|\mathbb{Z}) &= \lambda_C(t|\mathbb{Z}). \end{cases}$$

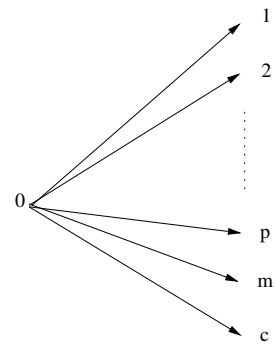


FIG 1. *Markov graph associated to* (X, δ, D)

In order to simplify the notation in the following derivations, let us denote by $p+1$ the index corresponding to a missing cause (previously denoted by m). It is important to note that, up to a multiplicative constant (α or $1-\alpha$), the additive form of the instantaneous rates is preserved (except for the transition $0 \to c$). It is obvious for $\lambda'_j(t|\mathbb{Z})$ with $1 \leq j \leq p$ but also for $\lambda'_{p+1}(\cdot|\mathbb{Z})$ since

$$\lambda'_{p+1}(t|\mathbb{Z}) = (1-\alpha)\left(\lambda_{0p+1}(t) + \beta_{p+1}^T Z(t)\right),$$

where $\lambda_{0p+1}(\cdot) = \sum_{j=1}^p \lambda_{0j}(\cdot)$ and $\beta_{p+1} = \sum_{j=1}^p \beta_j$. This will help us to improve the estimation of the regression parameters $\beta_1, \ldots, \beta_p$.

Now, let us suppose that we observe a sample $(X_i, \delta_i, D_i, \mathbb{Z}_i(X_i))_{1 \leq i \leq n}$ of $(X, \delta, D, \mathbb{Z}(X))$. Let $\tau < +\infty$ be the upper bound of the interval of study which means that individuals are only observed on the time interval $[0, \tau]$. Let $N_{ij}(\cdot)$, for $j \in \{1, \ldots, p+1\}$, be the counting processes defined by:

$$\begin{aligned} N_{ij}(t) &= 1(X_i \leq t, D_i = j) \text{ for } 1 \leq j \leq p, \\ \text{and } N_{ip+1}(t) &= 1(X_i \leq t, \delta_i = 1, D_i = 0). \end{aligned}$$

Finally write $Y_i(t) = 1(X_i \geq t)$ the individual risk process, for $i = 1, \ldots, n$.

From Andersen *et al.* (1993) or Fleming and Harrington (1991), we know that the processes $M_{ij}(\cdot)$, for $j = 1, \ldots, p+1$, defined by

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_i(s)\lambda'_j(s|\mathbb{Z})ds,$$

for $t \geq 0$, are zero mean martingales with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$ defined by

$$\mathcal{F}_t = \sigma\{N_{ij}(s), N_{im}(s), Y_i(s), \mathbb{Z}_i(s); s \leq t; 1 \leq i \leq n, j \in \{1, \ldots, p+1\}\}.$$

3. Statistical Inference on the Euclidean parameters under MCAR

3.1. Estimators

The finite dimensional parameters of our model are: the probability α of knowing the cause of death; and the regression parameters $\beta_1, \ldots, \beta_p$ of each cause-specific hazard rate functions. Recall that $\alpha = P(R = 1|\delta = 1)$. Thus, the maximum likelihood estimator of α is nothing but the proportion of lifetimes with known cause of death among the uncensored lifetimes, that is:

$$\hat{\alpha} = \frac{\sum_{i=1}^n 1(D_i > 0)}{\sum_{i=1}^n 1(\delta_i = 1)} = \frac{\sum_{j=1}^p N_{\cdot j}(\tau)}{N_{\cdot\cdot}(\tau)},$$

where

$$N_{\cdot j}(t) = \sum_{i=1}^n N_{ij}(t) \quad \text{and} \quad N_{\cdot\cdot}(t) = \sum_{j=1}^{p+1} N_{\cdot j}(t), \text{ for all } t.$$

Extending an approach proposed by Lin and Ying (1994) in case of a single cause of death, one can estimate β_j, for $j = 1, \ldots, p$, by the solution $\hat{\beta}_j$ of the estimating equation $\mathcal{U}_j(\beta, \hat{\alpha}, \tau) = 0$ where

$$\mathcal{U}_j(\beta, \hat{\alpha}, \tau) = \sum_{i=1}^n \int_0^\tau \left[Z_i(s) - \bar{Z}(s)\right] \left[dN_{ij}(s) - \hat{\alpha}\beta^T Z_i(s) Y_i(s) ds\right], \tag{3.1}$$

and

$$\bar{Z}(s) = \frac{\sum_{i=1}^n Y_i(s) Z_i(s)}{\sum_{i=1}^n Y_i(s)}.$$

In the sequel we write

$$\hat{\beta}_{WMC} = \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} \tag{3.2}$$

the estimator of the regression parameters without taking into account the subjects with unknown cause of death (WMC abbreviates "Without taking into account the Missing Causes"). Note that it is a vector of dimension kp.

Now, since it has been seen that the cause-specific hazard rate function $\lambda'_{p+1}(\cdot)$ associated with a missing cause has an additive form too, one can also estimate β_{p+1} by the solution $\hat{\beta}_{p+1}$ of the estimating equation $\mathcal{U}_{p+1}(\beta, \hat{\alpha}, \tau) = 0$ where

$$\mathcal{U}_{p+1}(\beta, \hat{\alpha}, \tau) = \sum_{i=1}^n \int_0^\tau \left[Z_i(s) - \bar{Z}(s)\right] \left[dN_{ip+1}(s) - (1 - \hat{\alpha})\beta^T Z_i(s) Y_i(s) ds\right].$$

Closed-form expressions of these estimators are available and given below (see Equation (4.9)).

At this stage, we are in a situation where each parameter β_j, for $j = 1, \ldots, p$, has its own estimator $\hat{\beta}_j$. But we also have an estimator $\hat{\beta}_{p+1}$ of their sum $\beta_{p+1} = \beta_1 + \cdots + \beta_p$. This estimator only uses the information from the transitions $0 \to p+1$ and hence is not equal to $\hat{\beta}_1 + \cdots + \hat{\beta}_p$. It is, of course, of interest to use it in order to improve the estimation of the first parameters β_j, for $j = 1, \ldots, p$. To this end, we propose to find the linear transformation of our estimator $(\hat{\beta}_1^T, \ldots, \hat{\beta}_p^T, \hat{\beta}_{p+1}^T)^T$ which will lead to an estimator of $(\beta_1^T, \ldots, \beta_p^T)^T$ with minimum asymptotic variance. More precisely, let $\mathcal{H}$ be the family of all the block matrices

$$H = \begin{pmatrix} H_{11} & H_{12} & \cdots & H_{1p+1} \\ H_{21} & H_{22} & \cdots & H_{2p+1} \\ \vdots & \vdots & & \vdots \\ H_{p1} & H_{p2} & \cdots & H_{pp+1} \end{pmatrix},$$

where the H_{ij}, for $i = 1, \ldots, p$ and $j = 1, \ldots, p+1$, are $k \times k$ real valued matrices, such that

$$H \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \\ \beta_{p+1} \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \tag{3.3}$$

for all vectors $\beta_1, \ldots, \beta_p$ in $\mathbb{R}^k$ and $\beta_{p+1} = \beta_1 + \cdots + \beta_p$. Write

$$\hat{q}(H) = \mathrm{trace}(H \hat{\Sigma}_{\hat{\beta},\infty} H^T),$$

where $\hat{\Sigma}_{\hat{\beta},\infty}$ is an estimator of the asymptotic variance-covariance matrix of $(\hat{\beta}_1^T, \ldots, \hat{\beta}_p^T, \hat{\beta}_{p+1}^T)^T$. With

$$\hat{H} = \underset{H \in \mathcal{H}}{\mathrm{argmin}}\ \hat{q}(H),$$

an estimator of $(\beta_1^T, \ldots, \beta_p^T)$ with minimal asymptotic variance is given by:

$$\begin{pmatrix} \tilde{\beta}_1 \\ \vdots \\ \tilde{\beta}_p \end{pmatrix} = \hat{H} \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \\ \hat{\beta}_{p+1} \end{pmatrix}.$$

Thus, the optimal estimators of the regression parameters are

$$\tilde{\beta}_i = \sum_{j=1}^{p+1} \hat{H}_{ij} \hat{\beta}_j, \text{ for } i = 1, \ldots, p, \tag{3.4}$$

and are called T–optimal estimators in the sequel. Let us denote by

$$\tilde{\beta}_{Topt} = \begin{pmatrix} \tilde{\beta}_1 \\ \vdots \\ \tilde{\beta}_p \end{pmatrix} \tag{3.5}$$

this vector of dimension kp of T–optimal estimators. Such a way to improve estimators has been considered in Balakrishnan *et al.* (2010).

Note that the constraints on H, given by (3.3), are linear and that they act separately on the rows of H. Indeed, denoting by I_k and 0_k the identity matrix of order k and the null matrix of order k respectively, these constraints may be written

$$\begin{cases} H_{ii} + H_{ip+1} &= I_k \\ H_{ij} + H_{ip+1} &= 0_k, \end{cases}$$

for $1 \leq i \leq p$ and $j \in \{1, \ldots, p\} \backslash \{i\}$.

On the other hand, we have

$$\hat{q}(H) = \sum_{i=1}^{p} \text{trace}(H_{i\bullet} \hat{\Sigma}_{\hat{\beta},\infty} H_{i\bullet}^T) = \sum_{i=1}^{p} \hat{q}_i(H)$$

where $\hat{q}_i(H) = \text{trace}(H_{i\bullet} \hat{\Sigma}_{\hat{\beta},\infty} H_{i\bullet}^T)$ and $H_{i\bullet}$ is the ith row block of H.

Thus, it is sufficient to solve separately the following problems (P_i), for $i = 1, \ldots, p$:

$$(P_i) \quad \begin{cases} \text{Find } H_{i1}, \ldots, H_{ip+1} \text{ which minimize } \text{trace}(H_{i\bullet} \hat{\Sigma}_{\hat{\beta},\infty} H_{i\bullet}^T) \\ \text{such that: } H_{ii} + H_{ip+1} = I_k, \\ \qquad\qquad H_{ij} + H_{ip+1} = 0, \text{ for } j \neq i. \end{cases}$$

3.2. Solving problems (P_i)

Since problems $(P_1), \ldots, (P_p)$ are identical, we only give the method to solve (P_1). Let us introduce some temporary notations. For $j = 1, \ldots, p+1$, write $H^{(j)} = H_{1j}$. Let us introduce the bloc diagonal matrix

$$\hat{Q} = \text{diag} \underbrace{\left(\hat{\Sigma}_{\hat{\beta},\infty}, \ldots, \hat{\Sigma}_{\hat{\beta},\infty} \right)}_{p \text{ times}}$$

and

$$L = \left(H_{1\bullet}^{(1)}, \ldots, H_{1\bullet}^{(p+1)}, H_{2\bullet}^{(1)}, \ldots, H_{2\bullet}^{(p+1)}, \ldots, H_{p\bullet}^{(1)}, \ldots, H_{p\bullet}^{(p+1)} \right)^T,$$

where $H_{l\bullet}^{(j)}$ is the lth row of $H^{(j)}$.

One can rewrite the function $\hat{q}_1(H)$ to be minimized like:

$$\hat{q}_1(H) = \text{trace}\left(\left(H^{(1)}\dots H^{(p+1)}\right)\hat{\Sigma}_{\hat{\beta},\infty}\left(H^{(1)}\dots H^{(p+1)}\right)^T\right) = L\hat{Q}L^T.$$

Now, the constraints on H are

$$\left\{\begin{array}{lcl} H^{(1)} + H^{(p+1)} & = & I_k, \\ H^{(2)} + H^{(p+1)} & = & 0_k, \\ & \vdots & \\ H^{(p)} + H^{(p+1)} & = & 0_k. \end{array}\right.$$

These constraints can be rewritten $\mathcal{C}L = d$ where $\mathcal{C}$ is a $(pk^2) \times (k^2(p+1))$ matrix and d a pk^2-column vector.

Thus, the Lagrange function for this optimization problem with linear constraints is given by

$$\ell(L,\lambda) = \frac{1}{2}L^T\hat{Q}L + (\mathcal{C}L - d)^T\lambda,$$

where λ is a Lagrange multiplier vector. The optimal parameters $\hat{L}$ and $\hat{\lambda}$ necessarily satisfy the first-order conditions

$$\left\{\begin{array}{lclcl} \dfrac{\partial \ell}{\partial L}(L,\lambda) & = & 0 & = & \hat{Q}L + \mathcal{C}^T\lambda \\ \mathcal{C}L & = & d & & \end{array}\right. ,$$

leading to

$$\hat{\lambda} = -[\mathcal{C}\hat{Q}^{-1}\mathcal{C}^T]^{-1}d,$$

and

$$\hat{L} = \hat{Q}^{-1}\mathcal{C}^T[C\hat{Q}^{-1}\mathcal{C}^T]^{-1}d.$$

It has to be noted that $\hat{Q}$ is invertible whenever $\hat{\Sigma}_{\hat{\beta},\infty}$ is. Finally, having $\hat{L}$ we obtain $\hat{H}^{(j)}$ for $j = 1,\dots,p+1$.

4. Asymptotic behavior of the regression parameters estimators under MCAR

4.1. Additional notations and assumptions, preliminary results

Let

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \\ \beta_{p+1} \end{pmatrix} \text{ and } \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \\ \hat{\beta}_{p+1} \end{pmatrix}.$$

If z is a column vector in $\mathbb{R}^k$, let us write

$$x^{\otimes l} = \begin{cases} 1 & \text{when} \quad l = 0, \\ z & \text{when} \quad l = 1, \\ zz^T & \text{when} \quad l = 2. \end{cases}$$

Finally, let us denote by $S_l(\cdot)$, for $l = 0, 1, 2$, the processes defined by

$$S_l(s) = \frac{1}{n} \sum_{i=1}^{n} Y_i(s) Z_i^{\otimes l}(s),$$

for $0 \leq s \leq \tau$ and for b in $\mathbb{R}^k$ the process $S_3(\cdot; b)$ defined by

$$S_3(s; b) = \frac{1}{n} \sum_{i=1}^{n} Y_i(s) Z_i^{\otimes 2}(s) b^T Z_i(s).$$

From now on we make the following assumptions.

A1. The probability α to get a known cause of death is strictly positive.

A2. The upper bound τ of the time interval of study is such that

$$0 < \int_0^\tau \lambda_{0j}(s) ds < +\infty, \text{ for}, j = 1, \ldots, p,$$

and the covariate processes $Z_i(\cdot)$ are $(\mathcal{F}_t)_{t \geq 0}$-predictable and uniformly bounded with respect to $i \geq 1$.

A3. For $0 \leq l \leq 2$, there exists functions $s_k(\cdot)$ defined on $[0, \tau]$ such that

$$\max_{0 \leq l \leq 2} \sup_{s \in [0,\tau]} \|S_l(s) - s_l(s)\| \xrightarrow{P} 0, \text{ when } n \to +\infty.$$

Moreover, the function $s_0(\cdot)$ is bounded below by a positive real number.

A4. With the notations $a(u) = s_2(u) - s_1^{\otimes 2}(u)/s_0(u)$,

$$A(t) = \int_0^t a(u) du \text{ and } \theta(t) = \int_0^t \left[s_0(u) \lambda_{0p+1}(u) + \beta_{p+1}^T s_1(u) \right] du, \quad (4.1)$$

the matrix $A(\tau)$ is positive definite and the real number $\theta(\tau)$ is strictly positive.

The matrix

$$\int_0^\tau \left[S_2(s) - S_1^{\otimes 2}(s)/S_0(s) \right] ds$$

is also positive definite. Note that, from previous assumptions, it was already true asymptotically.

A5. For all $b \in \mathbb{R}^k$, let $S_4(\cdot; b)$ be the process defined by

$$S_4(s; b) = \frac{1}{n} \sum_{i=1}^{n} (b^T Z_i(s)) Y_i(s) (s_1(s) Z_i^T(s)/s_0(s)).$$

There exist functions $s_3(\cdot;b)$ and $s_4(\cdot;b)$ such that, for all $b \in \mathbb{R}^k$,

$$\max_{3\leq l\leq 4} \sup_{s\in[0,\tau]} \|S_l(s;b) - s_l(s;b)\| \xrightarrow{P} 0, \text{ when } n \to +\infty.$$

A6. The following functions are integrable on $[0,\tau]$:

$$a(\cdot), a(\cdot)\lambda_{0j}(\cdot), s_0(\cdot)\lambda_{0j}(\cdot), s_1(\cdot), \frac{s_1^{\otimes 2}(\cdot)}{s_0^2(\cdot)}, s_3(\cdot;\beta_j) \text{ and } s_4(\cdot;\beta_j), \text{ for } j = 1,\ldots,p.$$

Now, let us introduce, for all $b \in \mathbb{R}^k$, the processes $\mathcal{U}_j(b,\alpha_j,\cdot)$ defined, for $j = 1,\ldots,p+1$ and $t \in [0,\tau]$, by:

$$\mathcal{U}_j(b,\alpha_j,t) = \sum_{i=1}^n \int_0^t \left(Z_i(s) - \bar{Z}(s)\right)\left(dN_{ij}(s) - \alpha_j b^T Z_i(s) Y_i(s) ds\right),$$

where

$$\alpha_j = \begin{cases} \alpha & \text{if} \quad 1 \leq j \leq p, \\ 1-\alpha & \text{if} \quad j = p+1. \end{cases} \tag{4.2}$$

Note that values of these processes at $s = \tau$ appear in the estimating equations of Section 3.1.

It is easily seen that these processes can be rewritten as

$$\mathcal{U}_j(b,\alpha_j,\cdot) = \sum_{i=1}^n \int_0^{\cdot} \left(Z_i(s) - \bar{Z}(s)\right) dM_{ij}(s),$$

for $j = 1,\ldots,p+1$. Hence, they are local square integrable $(\mathcal{F}_t)_{t\geq 0}$-martingales as sum of stochastic integrals of predictable and bounded processes with respect to local square integrable martingales.

Now, let us introduce two technical results useful in the following section. Let us note that all the functional convergence results of this paper are considered in the Skorohod space of cadlag functions $D[0,\tau]$.

Proposition 4.1. *Under Assumptions A1-A6, the multivariate process $\mathbb{U}_n(\cdot)$ defined, for all $t \in [0,\tau]$, by*

$$\mathbb{U}_n(t) \quad = \quad \frac{1}{\sqrt{n}} \begin{pmatrix} \mathcal{U}_1(\beta_1,\alpha_1,t) \\ \vdots \\ \mathcal{U}_{p+1}(\beta_{p+1},\alpha_{p+1},t) \\ \displaystyle\sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta(\tau)} M_{\cdot j}(t) \end{pmatrix},$$

where $\alpha_j^* = 1-\alpha$ *for* $1 \leq j \leq p$ *and* $\alpha_{p+1}^* = -\alpha$, *converges weakly in* $D^{k(p+1)+1}[0,\tau]$, *as* n *tends to infinity, to a zero mean multivariate Gaussian martingale* $\mathbb{U}_\infty(\cdot)$ *with covariance matrix defined for all* $t \in [0,\tau]$ *by*

$$\Sigma_{\mathbb{U}_\infty}(t) = \begin{pmatrix} \Theta_1(t) & 0 & 0 & \xi_1(t) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \Theta_{p+1}(t) & \xi_{p+1}(t) \\ \xi_1^T(t) & \cdots & \xi_{p+1}^T(t) & \xi_\alpha(t) \end{pmatrix},$$

where, for $j = 1,\ldots,p+1$,

$$\begin{aligned}
\Theta_j(t) &= \alpha_j \int_0^t \left[a(s)\lambda_{0j}(s) + s_3(s;\beta_j) - s_4(s;\beta_j) - s_4^T(s;\beta_j) + (\beta_j^T s_1(s)) s_1^{\otimes 2}(s)/s_0^2(s)\right] ds, \\
\xi_j(t) &= \kappa_j \frac{\alpha(1-\alpha)}{\theta(\tau)} A(t)\beta_j, \text{ with } \kappa_j = 1, \text{ for } 1 \leq j \leq p \text{ and } \kappa_{p+1} = -1, \\
\xi_\alpha(t) &= \frac{\alpha(1-\alpha)}{\theta^2(\tau)}\theta(t)
\end{aligned}$$

and $A(\cdot)$ *as well as* $\theta(\cdot)$ *are defined in (4.1).*

Proof. As in Andersen and Gill (1982), the main idea is to apply Rebolledo's Theorem (see Rebolledo, 1980, or Andersen *et al.*, 1993, p. 83–84). We only derive here the limit of the predictable variation process associated to $\mathbb{U}_n(\cdot)$. This will give us the asymptotic variance-covariance matrix function $\Sigma_{\mathbb{U}_\infty}(\cdot)$ of $\mathbb{U}_n(\cdot)$.

On one hand, straightforward calculations show that we have, for $j = 1,\ldots,p+1$,

$$\begin{aligned}
&\left\langle \frac{1}{\sqrt{n}}\mathcal{U}_j(\beta_j,\alpha,\cdot)\right\rangle(t) = \frac{1}{n}\sum_{i=1}^n \int_0^t \left(Z_i(s)-\bar{Z}(s)\right)\left(Z_i(s)-\bar{Z}(s)\right)^T Y_i(s)\lambda_j'(s|\mathbb{Z})ds \\
= \;& \alpha_j \int_0^t \left[S_2(s) - S_1^{\otimes 2}(s)/S_0(s)\right]\lambda_{0j}(s)ds \\
+ \;& \alpha_j \int_0^t \left[S_3(s;\beta_j) - S_4(s;\beta_j) - S_4^T(s;\beta_j) + (\beta_j^T S_1(s)) S_1^{\otimes 2}(s)/S_0^2(s)\right] ds + o_P(1),
\end{aligned}$$

which, by Assumptions A2-A6, converges in probability, as n tends to infinity, to $\Theta_j(t)$ given in the Theorem.

On the other hand, since the martingales M_{ij} and $M_{i'j'}$ are orthogonal for all $1 \leq i,i' \leq n$ and $1 \leq j \neq j' \leq p+1$, we have:

$$\langle \mathcal{U}_j(\beta_j,\alpha_j,\cdot), \mathcal{U}_{j'}(\beta_{j'},\alpha_{j'},\cdot)\rangle(t) = 0,$$

whenever j is different from j' and for all $t \in [0,\tau]$. This justifies the null terms in the asymptotic covariance matrix $\Sigma_{\mathbb{U}_\infty}(t)$.

Now, thanks again to the orthogonality between martingales with different indices, it is easy to show that, for $1 \leq l \leq p+1$, we have

$$\frac{1}{n}\left\langle \mathcal{U}_l(\cdot), \sum_{j=1}^{p+1} \frac{\alpha_j^* M_{\cdot j}(\cdot)}{\theta(\tau)} \right\rangle (t) = \frac{\alpha_l^*}{n\theta(\tau)} \sum_{i=1}^{n} \int_0^t \left(Z_i(s) - \bar{Z}(s)\right) Y_i(s) \lambda_l'(s|\mathbb{Z}) ds$$

which converges in probability to

$$\xi_l(t) = \kappa_l \frac{\alpha(1-\alpha)}{\theta(\tau)} A(t)\beta_l,$$

when n tends to infinity. Finally, with the same kind of arguments we have

$$\left\langle \frac{1}{\sqrt{n}} \sum_{j=1}^{p+1} \frac{\alpha_j^* M_{\cdot j}(\cdot)}{\theta(\tau)} \right\rangle (t) = \frac{1}{n} \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta^2(\tau)} \sum_{i=1}^{n} \int_0^t Y_i(s) \lambda_j'(s|\mathbb{Z}) ds \xrightarrow{P} \xi_\alpha(t) = \frac{\alpha(1-\alpha)}{\theta^2(\tau)} \theta(t),$$

when n tends to infinity. □

4.2. Large sample behavior of $\hat{\alpha}$

Lemma 4.2. *Under Assumptions A1-A4, we have*

$$\sqrt{n}(\hat{\alpha} - \alpha) = \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta(\tau)} \frac{1}{\sqrt{n}} M_{\cdot j}(\tau) + o_P(1). \tag{4.3}$$

Moreover $\sqrt{n}(\hat{\alpha}-\alpha)$ converges weakly to a $N(0, \alpha(1-\alpha)/\theta(\tau))$ distribution when $n \to +\infty$.

Remark. The second result of this lemma is straightforward and doesn't require the first step. It arises from an easy application of the central limit theorem. However the first result will be useful in the next section.

Proof. From the definition of $\hat{\alpha}$ we can write:

$$\begin{aligned}
\hat{\alpha} &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\int_0^\tau dN_{ij}(s)}{\sum_{i=1}^{n}\sum_{j=1}^{p+1}\int_0^\tau dN_{ij}(s)} \\
&= \frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\int_0^\tau dM_{ij}(s) + \sum_{i=1}^{n}\sum_{j=1}^{p}\int_0^\tau Y_i(s)\lambda'_j(s|\mathbb{Z})ds}{\sum_{i=1}^{n}\sum_{j=1}^{p+1}\int_0^\tau dM_{ij}(s) + \sum_{i=1}^{n}\sum_{j=1}^{p+1}\int_0^\tau Y_i(s)\lambda'_j(s|\mathbb{Z})ds} \\
&= \frac{\sum_{j=1}^{p} M_{\cdot j}(\tau) + \alpha\sum_{i=1}^{n}\sum_{j=1}^{p}\int_0^\tau Y_i(s)\lambda_j(s|\mathbb{Z})ds}{M_{\cdot\cdot}(\tau) + \sum_{i=1}^{n}\sum_{j=1}^{p}\int_0^\tau Y_i(s)\lambda_j(s|\mathbb{Z})ds} \\
&= \frac{\sum_{j=1}^{p} M_{\cdot j}(\tau) + \alpha n\int_0^\tau \left[S_0(u)\lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)\right]du}{M_{\cdot\cdot}(\tau) + n\int_0^\tau \left[S_0(u)\lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)\right]du},
\end{aligned}$$

where $M_{\cdot\cdot}(\cdot)$ denotes the process $\sum_{i=1}^{n}\sum_{j=1}^{p+1} M_{ij}(\cdot)$. It follows that

$$\begin{aligned}
&\sqrt{n}(\hat{\alpha}-\alpha) \\
&= \frac{\frac{1-\alpha}{\sqrt{n}}\sum_{j=1}^{p} M_{\cdot j}(\tau) - \frac{\alpha}{\sqrt{n}}M_{\cdot p+1}(\tau)}{\frac{1}{n}M_{\cdot\cdot}(\tau) + \int_0^\tau \left[S_0(u)\lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)\right]du}1(N_{\cdot\cdot}(\tau)>0) - \sqrt{n}\alpha 1(N_{\cdot\cdot}(\tau)=0) \\
&= \frac{\frac{1-\alpha}{\sqrt{n}}\sum_{j=1}^{p} M_{\cdot j}(\tau) - \frac{\alpha}{\sqrt{n}}M_{\cdot p+1}(\tau)}{\frac{1}{n}M_{\cdot\cdot}(\tau) + \int_0^\tau \left[S_0(u)\lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)\right]du} + o_P(1).
\end{aligned}$$

Moreover, using Lenglart's inequality (see *e.g.* Andersen *et al.*, 1993) and Assumption A3, it is easily seen that

$$\frac{1}{n}M_{\cdot\cdot}(\tau) \xrightarrow{P} 0, \text{ as } n \to +\infty. \tag{4.4}$$

These two last results with Assumptions A3 and A4 complete the proof of (4.3).

As said in the previous remark, the asymptotic normality of $\hat{\alpha}$ is straightforward. □

4.3. Large sample behavior of $\hat{\beta}$

Theorem 4.3. *Under Assumptions A1-A6, the random vector $\sqrt{n}(\hat{\beta}-\beta)$ is asymptotically Gaussian, with zero mean and positive definite covariance matrix*

$$\Sigma_{\hat{\beta},\infty} = \Sigma_1(\tau)\Sigma_{\mathbb{U}_\infty}(\tau)\Sigma_1^T(\tau),$$

where

$$\Sigma_1(\tau) \quad = \quad \begin{pmatrix} \frac{A^{-1}(\tau)}{\alpha} & 0 & \cdots & 0 & -\frac{\beta_1}{\alpha} \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \frac{A^{-1}(\tau)}{\alpha} & 0 & -\frac{\beta_p}{\alpha} \\ 0 & \cdots & 0 & \frac{A^{-1}(\tau)}{1-\alpha} & \frac{\beta_{p+1}}{1-\alpha} \end{pmatrix}.$$

Proof. Write

$$\hat{\alpha}_j = \begin{cases} \hat{\alpha} & \text{if} \quad 1 \le j \le p \\ 1-\hat{\alpha} & \text{if} \quad j = p+1 \end{cases}.$$

From their definition, the estimators $\hat{\beta}_j$, for $j = 1, \ldots, p$, are such that $\mathcal{U}_j(\hat{\beta}_j, \hat{\alpha}_j, \tau) = 0$. Thus, one can write :

$$\frac{1}{\sqrt{n}}\mathcal{U}_j(\beta_j, \alpha_j, \tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \left(Z_i(s) - \bar{Z}(s)\right)\left(\hat{\alpha}_j\hat{\beta}_j^T Z_i(s)Y_i(s)ds - \alpha_j\beta_j^T Z_i(s)Y_i(s)ds\right)$$
$$= \left[\int_0^\tau \frac{1}{n}\sum_{i=1}^n \left(Z_i(s)Z_i^T(s) - \bar{Z}(s)Z_i^T(s)\right)Y_i(s)ds\right] \times \left[\sqrt{n}(\hat{\alpha}_j - \alpha_j)\hat{\beta}_j + \sqrt{n}\alpha_j(\hat{\beta}_j - \beta_j)\right].$$

With the notation

$$\hat{A}(\tau) = \int_0^\tau \left[S_2(s) - S_1^{\otimes 2}(s)/S_0(s)\right] ds,$$

we obtain

$$\frac{1}{\sqrt{n}}\mathcal{U}_j(\beta_j, \alpha_j, \tau) = \hat{A}(\tau)\left[\sqrt{n}(\hat{\alpha}_j - \alpha_j)\hat{\beta}_j + \alpha_j\sqrt{n}(\hat{\beta}_j - \beta_j)\right].$$

It follows that

$$\hat{\beta}_j - \beta_j = \frac{1}{\hat{\alpha}_j}\hat{A}^{-1}(\tau)\frac{1}{n}\mathcal{U}_j(\beta_j, \alpha_j, \tau) + \left(\frac{\alpha_j}{\hat{\alpha}_j} - 1\right)\beta_j, \tag{4.5}$$

thanks to Assumption A4 which insures that $\hat{A}(\tau)$ is invertible. Now, using Assumptions A3 and A4, Proposition 4.1 and Lemma 4.2, it is easily seen that the right-hand side of (4.5) converges to zero when n tends to infinity. Thus $\hat{\beta}$ is consistent.

Furthermore, after some straightforward calculations on equation (4.5), one can write, for $j = 1, \ldots, p+1$:

$$\begin{aligned}\sqrt{n}(\hat{\beta}_j - \beta_j) &= \frac{1}{\alpha_j}\left(A^{-1}(\tau)\frac{\mathcal{U}_j(\beta_j, \alpha_j, \tau)}{\sqrt{n}} - \beta_j\sqrt{n}(\hat{\alpha}_j - \alpha_j)\right) \\ &\quad + \frac{1}{\alpha_j}\left(\left(\hat{A}^{-1}(\tau) - A^{-1}(\tau)\right)\frac{\mathcal{U}_j(\beta_j, \alpha_j, \tau)}{\sqrt{n}} - \sqrt{n}(\hat{\alpha}_j - \alpha_j)(\hat{\beta}_j - \beta_j)\right).\end{aligned}$$

Using again Assumptions A3 and A4 (which ensure that $\hat{A}^{-1}(\tau)$ converges in probability to $A^{-1}(\tau)$), Proposition 4.1 and Lemma 4.2 as well as the consistency of $\hat{\beta}$, one can prove that the second term of the right-hand side of this last equality is an $o_P(1)$. Thus we can write:

$$\sqrt{n}(\hat{\beta}_j - \beta_j) = \frac{1}{\alpha}\left(A^{-1}(\tau)\frac{\mathcal{U}_j(\beta_j, \alpha_j, \tau)}{\sqrt{n}} - \beta_j\sqrt{n}(\hat{\alpha} - \alpha)\right) + o_P(1), \tag{4.6}$$

for $j = 1 \ldots, p$, and

$$\begin{aligned}&\sqrt{n}(\hat{\beta}_{p+1} - \beta_{p+1}) \\ &= \frac{1}{1-\alpha}\left(A^{-1}(\tau)\frac{\mathcal{U}_{p+1}(\beta_{p+1}, \alpha_{p+1}, \tau)}{\sqrt{n}} + \beta_{p+1}\sqrt{n}(\hat{\alpha} - \alpha)\right) + o_P(1).\end{aligned} \tag{4.7}$$

Hence, from (4.3), (4.6) and (4.7) we finally obtain

$$\sqrt{n}(\hat{\beta} - \beta) = \Sigma_1(\tau)\mathbb{U}_n(\tau) + o_P(1), \tag{4.8}$$

where $\Sigma_1(\tau)$ is given in Theorem 4.3. This and Proposition 4.1 complete the proof. Note that because $\Sigma_{\mathbb{U}_\infty}(\tau)$ is positive definite, the matrix $\Sigma_{\hat{\beta},\infty}$ is positive definite too. □

4.4. The optimal estimator $\tilde{\beta}$

Recall that our T–optimal estimator is $\tilde{\beta} = \hat{H}\hat{\beta}$, where

$$\hat{H} = \text{argmin}_{H\in\mathcal{H}}\hat{q}(H) = \text{argmin}_{H\in\mathcal{H}}\text{trace}(H\hat{\Sigma}_{\hat{\beta},\infty}H^T),$$

and $\hat{\Sigma}_{\hat{\beta},\infty}$ is required to be an estimator of the asymptotic covariance matrix $\Sigma_{\hat{\beta},\infty}$ given in Theorem 4.3. Thus, we first have to find such an estimator of $\Sigma_{\hat{\beta},\infty}$.

It is easy to check from their definition that the explicit expressions of the estimators $\hat{\beta}_j$, for $j = 1, \ldots, p+1$, are

$$\hat{\beta}_j = \frac{1}{\hat{\alpha}_j}\hat{A}^{-1}(\tau)\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left(Z_i(s) - \bar{Z}(s)\right) dN_{ij}(s). \tag{4.9}$$

Furthermore, let

$$\hat{\Theta}_j(\tau) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left(Z_i(s) - \bar{Z}(s)\right)^{\otimes 2} dN_{ij}(s),$$

and

$$\hat{\xi}_j(\tau) = \kappa_j \frac{\hat{\alpha}(1-\hat{\alpha})}{\hat{\theta}(\tau)} \hat{A}(\tau)\hat{\beta}_j,$$

where $\hat{\theta}(\tau) = N_{\cdot\cdot}(\tau)/n$, and finally

$$\hat{\xi}_\alpha(\tau) = \frac{\hat{\alpha}(1-\hat{\alpha})}{\hat{\theta}(\tau)}.$$

Therefore, one can use respectively

$$\hat{\Sigma}_1(\tau) \quad = \quad \begin{pmatrix} \frac{\hat{A}^{-1}(\tau)}{\hat{\alpha}} & 0 & \cdots & 0 & -\frac{\hat{\beta}_1}{\hat{\alpha}} \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \frac{\hat{A}^{-1}(\tau)}{\hat{\alpha}} & 0 & -\frac{\hat{\beta}_p}{\hat{\alpha}} \\ 0 & \cdots & 0 & \frac{\hat{A}^{-1}(\tau)}{1-\hat{\alpha}} & \frac{\hat{\beta}_{p+1}}{1-\hat{\alpha}} \end{pmatrix}$$

and

$$\hat{\Sigma}_{\mathbb{U}_\infty}(\tau) \quad = \quad \begin{pmatrix} \hat{\Theta}_1(\tau) & 0 & 0 & \hat{\xi}_1(\tau) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \hat{\Theta}_{p+1}(\tau) & \hat{\xi}_{p+1}(\tau) \\ \hat{\xi}_1^T(\tau) & \cdots & \hat{\xi}_{p+1}^T(\tau) & \hat{\xi}_\alpha(\tau) \end{pmatrix}$$

as estimators of $\Sigma_1(\tau)$ and $\Sigma_{\mathbb{U}_\infty}(\tau)$. Finally, let $\beta^* = (\beta_1^T, \ldots, \beta_p^T)^T$.

Theorem 4.4. *Under Assumptions A1-A6, the estimator*

$$\hat{\Sigma}_{\hat{\beta},\infty} = \hat{\Sigma}_1(\tau)\hat{\Sigma}_{\mathbb{U}_\infty}(\tau)\hat{\Sigma}_1^T(\tau)$$

converges in probability to the matrix $\Sigma_{\hat{\beta},\infty}$.

Moreover, $\sqrt{n}(\tilde{\beta}_{Topt} - \beta^*)$ *is asymptotically Gaussian distributed with mean zero and covariance matrix whose trace minimizes* $q(H) = trace(H\Sigma_{\hat{\beta},\infty}H^T)$ *over* $H \in \mathcal{H}$.

Proof. We know from Lemma 4.2 that $\hat{\alpha}$ converges in probability to α. We have also seen in the proof of Theorem 4.3 that $\hat{A}^{-1}(\tau)$ and $\hat{\beta}_j$, for $j = 1, \ldots, p+1$, converge in probability respectively to $A^{-1}(\tau)$ and β_j. Thus, we get the convergence in probability of $\hat{\Sigma}_1(\tau)$ to $\Sigma_1(\tau)$.

Moreover, we have seen in the proof of Lemma 4.2 that

$$N_{\cdot\cdot}(\tau) = M_{\cdot\cdot}(\tau) + n\int_0^\tau \left[S_0(u)\lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)\right] du.$$

Thanks to Assumptions A3 and A4, the integral in the right-hand side of the last equation converges in probability to $\theta(\tau)$. This and (4.4) give the convergence in probability of $\hat{\theta}(\tau)$ to $\theta(\tau)$.

From the above convergences we have the convergence in probability of $\hat{\xi}_j(\tau)$ to $\xi_j(\tau)$, for $j = 1, \ldots, p+1$, as well as the one of $\hat{\xi}_\alpha$ to ξ_α.

On the other hand, one can write, for $j = 1, \ldots, p$:

$$\begin{aligned}\hat{\Theta}_j(\tau) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left(Z_i(s) - \bar{Z}(s)\right)^{\otimes 2} dM_{ij}(s) \\ &+ \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left(Z_i(s) - \bar{Z}(s)\right)^{\otimes 2} Y_i(s)\lambda'_j(s|\mathbb{Z}_i)ds. \qquad (4.10)\end{aligned}$$

With the notation $A^{\otimes 2} = AA^T$ when A is also a matrix, it is easily seen that

$$\begin{aligned}&\left\langle \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\int_0^{\cdot} t\left(Z_i(s) - \bar{Z}(s)\right)^{\otimes 2} dM_{ij}(s)\right\rangle(\tau) \\ &= \tfrac{1}{n^2}\textstyle\sum_{i=1}^{n}\int_0^\tau \left(\left(Z_i(s) - \bar{Z}(s)\right)^{\otimes 2}\right)^{\otimes 2} Y_i(s)\lambda'_j(s|\mathbb{Z}_i)ds,\end{aligned}$$

which converges to zero in probability when n tends to infinity, by Assumptions A2 and A3. This and Lenglart inequality yield the convergence in probability to zero of the first term of (4.10). Moreover, we have seen in the proof of Proposition 4.1 that the second term of (4.10) is equal to $\langle \mathcal{U}_j(\beta_j, \alpha, \tau)/\sqrt{n}\rangle$ and converges in probability to $\Theta_j(\tau)$, when n tends to infinity. These two last convergences prove that $\hat{\Theta}_j(\tau)$ converges to $\Theta_j(\tau)$, for $j = 1, \ldots, p+1$. Thus we get the consistency of the estimator $\hat{\Sigma}_{\mathbb{U}_\infty}(\tau)$ and finally the one of $\hat{\Sigma}_{\hat{\beta},\infty}$.

From the above and the continuous dependency of $\hat{H}$ on $\hat{\Sigma}_{\hat{\beta},\infty}$ we deduce that $\hat{H}$ converges in probability to the matrix H^{opt} in $\mathcal{H}$ which minimizes trace$(H\Sigma_{\hat{\beta},\infty}H^T)$. We recall that the existence of such an optimal matrix is ensured because $\Sigma_{\hat{\beta},\infty}$ is positive definite. Since $\hat{H}\beta = \beta^*$, we get from the above and Theorem 4.3 that $\sqrt{n}(\tilde{\beta}_{Topt} - \beta^*) = \hat{H}\sqrt{n}(\hat{\beta} - \beta)$ converges to a zero mean Gaussian distributed random variable with covariance matrix $H^{opt}\Sigma_{\hat{\beta},\infty}(H^{opt})^T$ which is optimal in the sense defined earlier. □

5. Statistical inference on the functional parameters under MCAR

Even though it is not the model considered at the beginning of this paper, we will first consider the fully nonparametric case; that is, a model without covariates where only functional parameters have to be estimated. Of course, this model is also of interest for applications. Then we will come back to our semiparametric model and will see how to estimate its functional parameters.

5.1. Inference in the nonparameteric model

In this case, no parametric form is assumed on the cause-specific hazard rate functions $\lambda_j(\cdot)$, for $j = 1, \ldots, p$ and we also do not take into account any covariate. Let $\Lambda_j(\cdot)$, for $j = 1, \ldots, p$, denote the cause-specific cumulative hazard

rate functions defined by:

$$\Lambda_j(t) = \int_0^t \lambda_j(u)du,$$

for t in $[0, \tau]$ and

$$\Lambda_{p+1}(t) = \sum_{j=1}^{p} \Lambda_j(t).$$

With the assumption on M and the hypothesis of independence between (T, d) and C, the instantaneous transition rates of the Markov process with graph given in Figure 1 are:

$$\begin{cases} \lambda'_j(t) & = & \alpha\lambda_j(t), \text{ for } j = 1, \ldots, p, \\ \lambda'_{p+1}(t) & = & (1-\alpha)\sum_{j=1}^{p} \lambda_j(t) = (1-\alpha)\lambda_{p+1}(t), \\ \lambda'_C(t) & = & \lambda_C(t). \end{cases}$$

Contrary to the model of Goetghebeur and Ryan (1995), when there is no covariate in the data, the model is still of interest because it allows different failure rates for failure causes.

5.1.1. First estimators of the cause-specific cumulative hazard rate functions

By standard arguments, one can easily estimate $\Lambda_j(t)$, for $j = 1, \ldots, p+1$ and $t \in [0, \tau]$, by

$$\hat{\Lambda}_j(t) = \frac{1}{\hat{\alpha}_j}\hat{\Lambda}'_j(t), \tag{5.1}$$

where $\hat{\alpha}_j$ is defined at the beginning of the proof of Theorem 4.3, $Y(s) = \sum_{i=1}^{n} Y_i(s)$ is the number at risk at time s and

$$\hat{\Lambda}'_j(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_{ij}(s)}{Y(s)}$$

is the well-known Nelson-Aalen estimator of

$$\Lambda'_j(t) = \int_0^t \lambda'_j(u)du.$$

Write:

$$\Lambda(\cdot) = \begin{pmatrix} \Lambda_1(\cdot) \\ \vdots \\ \Lambda_p(\cdot) \\ \Lambda_{p+1}(\cdot) \end{pmatrix} \quad \text{and} \quad \hat{\Lambda}(\cdot) = \begin{pmatrix} \hat{\Lambda}_1(\cdot) \\ \vdots \\ \hat{\Lambda}_p(\cdot) \\ \hat{\Lambda}_{p+1}(\cdot) \end{pmatrix}.$$

Note that, as for the regression parameters case, the estimator $\hat{\Lambda}_{p+1}(\cdot)$ differs from $\hat{\Lambda}_1(\cdot)+\cdots+\hat{\Lambda}_p(\cdot)$ since it is based on information coming from the transition $0 \to p+1$.

Now, let us introduce two classical assumptions that allow to obtain the following asymptotic results.

B1. τ satisfies $0 < \Lambda_j(\tau) < +\infty$, for $j = 1, \ldots, p$.

B2. There exists a function $s_0(\cdot)$, defined on $[0, \tau]$, and bounded away from 0, such that

$$\sup_{s\in[0,\tau]} \left| \frac{Y(s)}{n} - s_0(s) \right| \xrightarrow{P} 0, \text{ when } n \to +\infty.$$

These assumptions are nothing but Assumptions A2 and A3 of Section 4 adapted to this new model. It is well known that Assumptions B1 and B2 are fulfilled whenever τ is such that $S(\tau)\bar{G}(\tau) > 0$, where $S(\cdot)$ and $\bar{G}(\cdot)$ are respectively the survival functions of T and C.

Note also that the functions $s_0(\cdot)\lambda_j(\cdot)$ and $\lambda_j(\cdot)/s_0(\cdot)$ are integrable on $[0, \tau]$.

Finally, let us define:

$$\begin{aligned}
\eta_j(\cdot) &= \alpha_j \int_0^{\cdot} \lambda_j(u)/s_0(u)du, \text{ for } j = 1, \ldots, p, \\
\eta_{p+1}(\cdot) &= \sum_{j=1}^{p} \eta_j(\cdot), \\
\theta_0(\cdot) &= \int_0^{\cdot} s_0(u)\lambda_{p+1}(u)du \\
\rho_j(\cdot) &= \kappa_j \alpha(1-\alpha)\Lambda_j(\cdot)/\theta_0(\tau), \text{ for } 1 \le j \le p+1, \\
\text{and } \rho_\alpha(\tau) &= \alpha(1-\alpha)/\theta_0(\tau).
\end{aligned}$$

Theorem 5.1. *Under Assumptions A1, B1 and B2, the process* $\sqrt{n}(\hat{\Lambda}(\cdot) - \Lambda(\cdot))$ *converges weakly in* $D^{p+1}[0, \tau]$ *to a zero mean Gaussian process* $\mathbb{L}(\cdot)$ *with covariance matrix function*

$$\Sigma_{\mathbb{L}}(t) = \Sigma_2(t)\Sigma_{\mathbb{V}_\infty}(t)\Sigma_2^T(t),$$

where

$$\Sigma_2(t) = \begin{pmatrix} \frac{1}{\alpha} & 0 & \cdots & 0 & -\frac{\Lambda_1(t)}{\alpha} \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \frac{1}{\alpha} & 0 & -\frac{\Lambda_p(t)}{\alpha} \\ 0 & \cdots & 0 & \frac{1}{1-\alpha} & \frac{\Lambda_{p+1}(t)}{1-\alpha} \end{pmatrix}$$

and

$$\Sigma_{\mathbb{V}_\infty}(t) = \begin{pmatrix} \eta_1(t) & 0 & 0 & \rho_1(t) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \eta_{p+1}(t) & \rho_{p+1}(t) \\ \rho_1(t) & \cdots & \rho_{p+1}(t) & \rho_\alpha(\tau) \end{pmatrix}.$$

Proof. Our proof starts with the observation that:

$$\hat{\Lambda}_j(t) - \Lambda_j(t) = \frac{1}{\hat{\alpha}_j}\left(\hat{\Lambda}'_j(t) - \Lambda'_j(t)\right) - \frac{\Lambda'_j(t)}{\sqrt{n}\alpha_j\alpha'_j}\sqrt{n}\left(\hat{\alpha}_j - \alpha\right). \tag{5.2}$$

A straightforward application of the central limit theorem gives the $\sqrt{n}$-asymptotical normality of $\hat{\alpha}$ and thus that $\sqrt{n}\,(\hat{\alpha}-\alpha)$ is an $O_p(1)$. Moreover, under Assumptions B1 and B2, the Nelson-Aalen estimators $\hat{\Lambda}'_j(\cdot)$, for $j = 1,\ldots,p+1$, are well-known to be uniformly consistent (see *e.g.* Andersen *et al.*, 1993). Hence, under the assumptions of Theorem 5.1, one can get easily from (5.2) the uniform convergence in $D[0,\tau]$ of $\hat{\Lambda}_j(\cdot)$ to $\Lambda_j(\cdot)$, for $j = 1,\ldots,p+1$. These convergences and (5.2) yield

$$\sqrt{n}(\hat{\Lambda}_j(\cdot)-\Lambda_j(\cdot)) = \frac{\sqrt{n}}{\alpha_j}\int_0^{\cdot}\frac{dM_{\cdot j}(s)}{Y(s)} - \frac{\Lambda_j(\cdot)}{\alpha_j}\sqrt{n}(\hat{\alpha}_j-\alpha_j)+o_P(1), \tag{5.3}$$

in $D[0,\tau]$ and for $j = 1,\ldots,p+1$.

But, following the lines of Lemma 4.2, we show that

$$\sqrt{n}(\hat{\alpha}-\alpha) = \sum_{j=1}^{p+1}\frac{\alpha_j^*}{\theta_0(\tau)}\frac{1}{\sqrt{n}}M_{\cdot j}(\tau)+o_P(1),$$

where $\theta_0(\tau) > 0$ by B1 and B2. From this and (5.3), we deduce that

$$\sqrt{n}(\hat{\Lambda}(t)-\Lambda(t)) = \Sigma_2(t)\mathbb{V}_n(t)+o_P(1), \tag{5.4}$$

in $D^{p+1}[0,\tau]$, where $\Sigma_2(t)$ is defined in Theorem 5.1 and

$$\mathbb{V}_n(t) = \sqrt{n}\begin{pmatrix} \int_0^t \frac{dM_{\cdot 1}(s)}{Y(s)} \\ \vdots \\ \int_0^t \frac{dM_{\cdot p+1}(s)}{Y(s)} \\ \frac{1}{n}\sum_{j=1}^{p+1}\frac{\alpha_j^*}{\theta_0(\tau)}M_{\cdot j}(\tau) \end{pmatrix}.$$

Using the Rebolledo Theorem or an approach given in Dauxois (2000) one can easily prove that, under Assumptions B1 and B2, the process $\mathbb{V}_n(\cdot)$ converges weakly in $D^{p+2}[0,\tau]$ to a zero mean Gaussian process $\mathbb{V}_\infty(\cdot)$. It remains to specify its covariance function $\Sigma_{\mathbb{V}_\infty}(\cdot)$. First, note that we can write $\mathbb{V}_n(t) = \mathbb{V}_{1,n}(t)+\mathbb{V}_{2,n}(t)$ where

$$\mathbb{V}_{1,n}(t) = \sqrt{n}\begin{pmatrix} \int_0^t \frac{dM_{\cdot 1}(s)}{Y(s)} \\ \vdots \\ \int_0^t \frac{dM_{\cdot p+1}(s)}{Y(s)} \\ \frac{1}{n}\sum_{j=1}^{p+1}\frac{\alpha_j^*}{\theta_0(\tau)}M_{\cdot j}(t) \end{pmatrix}$$

and

$$\mathbb{V}_{2,n}(t) = \frac{1}{\sqrt{n}}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \sum_{j=1}^{p+1}\frac{\alpha_j^*}{\theta_0(\tau)}(M_{\cdot j}(\tau)-M_{\cdot j}(t)) \end{pmatrix}.$$

Since the martingales $M_{\cdot j}$ are orthogonal, for $j = 1, \ldots, p+1$, we have

$$\mathbb{E}\left[\mathbb{V}_n^{\otimes 2}(t)\right] = \mathbb{E}\left[\mathbb{V}_{1,n}^{\otimes 2}(t)\right] + \mathbb{E}\left[\mathbb{V}_{2,n}^{\otimes 2}(t)\right]$$

and thus

$$\Sigma_{\mathbb{V}_\infty}(t) = \lim_{n\to\infty} \mathbb{E}\left[\mathbb{V}_{1,n}^{\otimes 2}(t)\right] + \lim_{n\to\infty} \mathbb{E}\left[\mathbb{V}_{2,n}^{\otimes 2}(t)\right].$$

Straightforward calculus lead to the expression of $\Sigma_{\mathbb{V}_\infty}(\cdot)$ given in the theorem. From this and (5.4) we get the desired result. □

One can get a consistent estimator of the covariance function $\Sigma_{\mathbb{L}}(\cdot)$ of the limit process $\mathbb{L}(\cdot)$. To this end, let us define, for $j = 1, \ldots, p+1$:

$$\hat{\eta}_j(t) = n \int_0^t \frac{dN_{\cdot j}(s)}{Y^2(s)}.$$

Define also:

$$\hat{\rho}_j(t) = \kappa_j \frac{\hat{\alpha}(1-\hat{\alpha})\hat{\Lambda}_j(t)}{\hat{\theta}_0(\tau)}, \text{ for } j = 1, \ldots, p+1, \text{ and } \hat{\rho}_\alpha = \frac{\hat{\alpha}(1-\hat{\alpha})}{\hat{\theta}_0(\tau)},$$

where $\hat{\theta}_0(\tau) = N_{\cdot\cdot}(\tau)/n$.

Then, plug-in estimators of $\Sigma_2(t)$ and $\Sigma_{\mathbb{V}_\infty}(t)$, for $t \in [0, \tau]$, are respectively

$$\hat{\Sigma}_2(t) = \begin{pmatrix} \frac{1}{\hat{\alpha}} & 0 & \cdots & 0 & -\frac{\hat{\Lambda}_1(t)}{\hat{\alpha}} \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \frac{1}{\hat{\alpha}} & 0 & -\frac{\hat{\Lambda}_p(t)}{\hat{\alpha}} \\ 0 & \cdots & 0 & \frac{1}{1-\hat{\alpha}} & \frac{\hat{\Lambda}_{p+1}(t)}{1-\hat{\alpha}} \end{pmatrix}$$

and

$$\hat{\Sigma}_{\mathbb{V}_\infty}(t) = \begin{pmatrix} \hat{\eta}_1(t) & 0 & 0 & \hat{\rho}_1(t) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \hat{\eta}_{p+1}(t) & \hat{\rho}_{p+1}(t) \\ \hat{\rho}_1(t) & \cdots & \hat{\rho}_{p+1}(t) & \hat{\rho}_\alpha \end{pmatrix}.$$

Finally, an empirical estimator of $\Sigma_{\mathbb{L}}(\cdot)$ is given by:

$$\hat{\Sigma}_{\mathbb{L}}(\cdot) = \hat{\Sigma}_2(\cdot)\hat{\Sigma}_{\mathbb{V}_\infty}(\cdot)\hat{\Sigma}_2^T(\cdot).$$

5.1.2. Optimal estimators of the cause-specific cumulative hazard rate functions

Except for the fact that estimators in this section are functions, the situation is the same as the one of Section 3.1. Indeed, our multivariate functional estimator $\hat{\Lambda}(\cdot)$ gives estimators $\hat{\Lambda}_1(\cdot), \ldots, \hat{\Lambda}_p(\cdot)$ of $\Lambda_1(\cdot), \ldots, \Lambda_p(\cdot)$ respectively and

an estimator $\hat{\Lambda}_{p+1}(\cdot)$ of their sum $\Lambda_{p+1}(\cdot) = \sum_{i=1}^{p} \Lambda_i(\cdot)$. Here also we will look for a linear transformation of $\hat{\Lambda}(\cdot)$ which will give us an optimal estimator of $\Lambda^*(\cdot) = (\Lambda_1(\cdot), \ldots, \Lambda_p(\cdot))^T$.

To this end, let us define $\mathcal{H}'$ as the set of $p \times (p+1)$ real valued matrices such that $Ha = a^*$ for all $a^* = (a_1, \ldots, a_p)^T \in \mathbb{R}^p$ and $a = (a^{*T}, \sum_{j=1}^{p} a_j)^T \in \mathbb{R}^{p+1}$. We define

$$\hat{H}(t) = \underset{H \in \mathcal{H}'}{\operatorname{argmin}} \operatorname{trace}(H \hat{\Sigma}_{\mathbb{L}}(t) H^T) \tag{5.5}$$

and set

$$\tilde{\Lambda}(\cdot) = \hat{H}(\cdot)\hat{\Lambda}(\cdot) \tag{5.6}$$

as a new estimator of $\Lambda^*(\cdot)$. The next theorem proves that the later estimator is asymptotically normal and T–optimal.

Theorem 5.2. *Under Assumptions A1, B1 and B2, the matrix $\hat{\Sigma}_{\mathbb{L}}(t)$ converges in probability to the matrix $\Sigma_{\mathbb{L}}(t)$, for all $t \in [0, \tau]$.*

Moreover, let us assume that the matrix $\Sigma_{\mathbb{L}}(t)$ is invertible, for all $t \in]0, \tau]$. If, for all $t \in [0, \tau]$, the matrix $\hat{H}(t)$ is the unique solution of (5.5), then the process $\sqrt{n}(\tilde{\Lambda}(\cdot) - \Lambda^*(\cdot))$ converges weakly in $D^p[0, \tau]$ to a centered Gaussian process $\mathbb{L}'(\cdot) = H_{\Sigma_{\mathbb{L}}}(\cdot)\mathbb{L}(\cdot)$ with covariance function $H_{\Sigma_{\mathbb{L}}}(\cdot)\Sigma_{\mathbb{L}}(\cdot)H^T_{\Sigma_{\mathbb{L}}}(\cdot)$ where

$$H_{\Sigma_{\mathbb{L}}}(t) = \underset{H \in \mathcal{H}'}{\operatorname{argmin}} \operatorname{trace}(H \Sigma_{\mathbb{L}}(t) H^T).$$

Proof. The proof is omitted since it follows the lines of the proof of Theorem 4.4. □

Let us denote by L the column vector in $\mathbb{R}^{p(p+1)}$ defined by $L = (H_1, \ldots, H_p)$ where H_i is the ith row of $H \in \mathcal{H}'$. The link between $L = (l_i)_{1 \le i \le p(p+1)}$ and $H = (h_{i,j})_{1 \le i \le p; 1 \le j \le p+1}$ is therefore $h_{i,j} = l_{(i-1)(p+1)+j}$. One can see that the linear constraints on H may be written on L as $\mathcal{C}L = d$ where $\mathcal{C}$ and d are known. Indeed

$$\begin{aligned}
& \begin{cases} h_{i,i} + h_{i,p+1} &= 1, \text{ for } 1 \le i \le p, \\ h_{i,j} + h_{i,p+1} &= 0, \text{ for } 1 \le i \le p,\ 1 \le j \le p+1 \text{ and } i \ne j, \end{cases} \\
\Leftrightarrow\ & \begin{cases} l_{(i-1)(p+1)+i} + l_{i(p+1)} &= 1, \text{ for } 1 \le i \le p, \\ l_{(i-1)(p+1)+j} + l_{(i+1)p} &= 0, \text{ for } 1 \le i \le p,\ 1 \le j \le p+1 \text{ and } i \ne j, \end{cases} \\
\Leftrightarrow\ & \mathcal{C}L = d,
\end{aligned}$$

where the matrix $\mathcal{C}$ and the vector d are obvious. Moreover, let $\hat{Q}(t)$ be the $p \times p$ block diagonal matrix defined by

$$\hat{Q}(t) = \operatorname{diag} \underbrace{\left(\hat{\Sigma}_{\mathbb{L}}(t), \ldots, \hat{\Sigma}_{\mathbb{L}}(t)\right)}_{p \text{ times}},$$

and note that

$$\mathrm{trace}(H\hat{\Sigma}_{\mathbb{L}}(t)H^T) = L^T\hat{Q}(t)L.$$

Thus, in order to find our optimal estimator, we have to solve the following optimization problems

$$(P_t) \quad \begin{cases} \text{Find } L(t) \text{ which minimizes } L^T(t)\hat{Q}(t)L(t) \\ \text{such that : } \mathcal{C}L(t) = d \end{cases},$$

for $t \in [0, \tau]$.

Following the method of Section 3.2, the solution of (P_t) is:

$$\hat{L}(t) = \hat{Q}^{-1}(t)\mathcal{C}^T(\mathcal{C}\hat{Q}^{-1}(t)\mathcal{C}^T)^{-1}d$$

and $\hat{H}(t)$ is therefore defined by $\hat{h}_{ij}(t) = \hat{l}_{(i-1)(p+1)+j}(t)$ for $1 \leq i \leq p$ and $1 \leq j \leq p+1$.

Remark. Since $\hat{\Sigma}_{\mathbb{L}}(t)$ is piecewise constant, it is sufficient to calculate the matrix $\hat{H}(t)$ at points $t \in [0, \tau]$ where the counting process $N_{\cdot\cdot}$ jumps, that is at points $X_i \in [0, \tau]$ such that $\delta_i = 1$.

5.1.3. Estimation of the cumulative incidence functions and the survival function of T

Our aim in this section is to introduce estimators of the survival function $S(\cdot)$ of the lifetime T as well as estimators of the cumulative incidence functions $F_j(\cdot)$ defined, for all t and $j = 1, \ldots, p$, by

$$F_j(t) = P(T \leq t, d = j).$$

Let us recall that $\Lambda_{\cdot}(\cdot) = \sum_{j=1}^{p} \Lambda_j(\cdot)$ is the cumulative hazard rate function of the survival time T. It is well known that one can write the survival function in terms of the cumulative hazard rate function:

$$S(t) = \prod_{u \in]0,t]} (1 - d\Lambda_{\cdot}(u)),$$

where $\prod$ denotes the product integral (see Gill & Johansen, 1990). Using a plug-in method on this last equation, one can get the Kaplan-Meier estimator of $S(t)$:

$$\hat{S}(t) = \prod_{u \in]0,t]} \left(1 - d\tilde{\Lambda}_{\cdot}(u)\right),$$

where $\tilde{\Lambda}(\cdot)$ is the optimal estimator of the previous paragraph and

$$\tilde{\Lambda}_{\cdot}(\cdot) = \sum_{j=1}^{p} \tilde{\Lambda}_j(\cdot).$$

On the other hand, it is also well-known that one can write, for $t \in [0, \tau]$ and $j = 1, \ldots, p$

$$F_j(t) = \int_0^t S(u^-) d\Lambda_j(u),$$

and that an estimator of this cumulative incidence function is given by the Aalen-Johansen estimator (see Andersen *et al.*, 1993)

$$\hat{F}_j(t) = \int_0^t \hat{S}(u^-) d\tilde{\Lambda}_j(u).$$

The asymptotic behavior of these estimators is given in the following theorem.

Theorem 5.3. *Under Assumptions A1, B1 and B2, we have the following weak convergence in $D^{p+1}[0, \tau]$, when n tends to infinity*

$$\sqrt{n}\left[\begin{pmatrix} \hat{F}_1(\cdot) \\ \vdots \\ \hat{F}_p(\cdot) \\ \hat{S}(\cdot) \end{pmatrix} - \begin{pmatrix} F_1(\cdot) \\ \vdots \\ F_p(\cdot) \\ S(\cdot) \end{pmatrix}\right] \longrightarrow \begin{pmatrix} \int_0^{\cdot} S(u) d\mathbb{L}_1'(u) - \int_0^{\cdot} S(u) \mathbb{L}_{\cdot}'(u) d\Lambda_1(u) \\ \vdots \\ \int_0^{\cdot} S(u) d\mathbb{L}_p'(u) - \int_0^{\cdot} S(u) \mathbb{L}_{\cdot}'(u) d\Lambda_p(u) \\ -S(\cdot)\mathbb{L}_{\cdot}'(\cdot) \end{pmatrix},$$

where $\mathbb{L}_j'(\cdot)$ is the jth component of the limit process $\mathbb{L}'(\cdot)$ of Theorem 5.2 and $\mathbb{L}_{\cdot}'(\cdot) = \mathbb{L}_1'(\cdot) + \cdots + \mathbb{L}_p'(\cdot)$.

Proof. This result is easily obtained from Theorem 5.2 and the functional δ-method (see *e.g.* van der Vaart & Wellner, 1996, for details on this method). Indeed, from the above, one can write

$$\begin{pmatrix} F_1(\cdot) \\ \vdots \\ F_p(\cdot) \\ S(\cdot) \end{pmatrix} = \psi \begin{pmatrix} \Lambda_1(\cdot) \\ \vdots \\ \Lambda_p(\cdot) \end{pmatrix},$$

where $\psi(\cdot)$ is a function from $D^p[0, \tau]$ to $D^{p+1}[0, \tau]$ with coordinate functions defined, for $i = 1, \ldots, p$, by:

$$F_i(\cdot) = \int_0^{\cdot} S(u^-) d\Lambda_i(u) = \psi_i \begin{pmatrix} \Lambda_1(\cdot) \\ \vdots \\ \Lambda_p(\cdot) \end{pmatrix},$$

and

$$S(\cdot) = \psi_{p+1} \begin{pmatrix} \Lambda_1(\cdot) \\ \vdots \\ \Lambda_p(\cdot) \end{pmatrix} = \prod_{u \in]0, \cdot]} (1 - d\Lambda_{\cdot}(u)).$$

We know (see again Van der Vaart and Wellner, 1996) that the product-integral function

$$\begin{array}{rccl} \phi(\cdot): & D[0,\tau] & \to & D[0,\tau] \\ & A & \mapsto & \phi(A) = \prod\limits_{u\in]0,\cdot]} (1 - dA(u)) \end{array}$$

is Hadamard differentiable on $BV_K[0,\tau]$, the subset of functions in $D[0,\tau]$ with total variation bounded by K. Its derivative is

$$\phi'_A(\alpha)(\cdot) = -\int_0^{\cdot} \phi(A)(u^-)\frac{\phi(A)(\cdot)}{\phi(A)(u)}d\alpha(u)$$

and is equal to $-\phi(A)(\cdot)\alpha(\cdot)$ when A is continuous.

We also know that the function

$$\begin{array}{rccl} \varphi(\cdot,\cdot): & D[0,\tau]\times BV_K[0,\tau] & \to & D[0,\tau] \\ & (A,B) & \mapsto & \varphi(A,B) = \int_0^{\cdot} AdB \end{array}$$

is Hadamard differentiable on (A,B) where A is of bounded variation, with derivative:

$$\varphi'_{A,B}(\alpha,\beta)(\cdot) = \int_0^{\cdot} A(u)d\beta(u) + \int_0^{\cdot} \alpha(u)dB(u).$$

Using for instance the chain rule (see Lemma 3.9.3 of Van der Vaart and Wellner, 1996) and the composition

$$(\Lambda_1(\cdot),\ldots,\Lambda_p(\cdot))^T \mapsto (\Lambda_i(\cdot),S(\cdot))^T \mapsto \int_0^{\cdot} S(u^-)d\Lambda_i(u),$$

one can prove that the coordinate functions $\psi_1(\cdot),\ldots,\psi_p(\cdot)$ are differentiable and get their derivatives. The functional δ-method allows us to obtain from Theorem 5.2 the following weak convergence in $D^{p+1}[0,\tau]$:

$$\sqrt{n}\left(\psi(\tilde{\Lambda}(\cdot)) - \psi(\Lambda^*(\cdot))\right) \to \psi'_{\Lambda^*}(\mathbb{L}'(\cdot)),$$

when n tends to infinity. Straightforward differential calculus and the continuity of $S(\cdot)$ give the expression of $\psi'_{\Lambda^*}(\mathbb{L}'(\cdot))$ detailed in the Theorem. □

5.2. With explanatory variables

Now, let us come back to the semiparametric model of equation (2.1) with explanatory variable $Z(\cdot)$. As we have done before, we will first introduce estimators of the cause-specific hazard rate functions which only use the lifetimes Without Missing Causes (WMC). Then, as previously, we will introduce the T-optimal estimators of these functions.

5.2.1. First estimators of the cause-specific cumulative hazard rate functions: the WMC estimators

First note that, even if we will only use the quantities with index $j = 1, \ldots, p$ in this paragraph, we will also introduce here the terms with index $p+1$. These latter will be used in the following paragraph where we will get the T-optimal estimators which also use the lifetimes with unknown cause of failure. In the sequel, $\Lambda_{0p+1}(\cdot)$ and $\Lambda_{p+1}(\cdot|\mathbb{Z})$ denote respectively $\sum_{j=1}^{p} \Lambda_{0j}(\cdot)$ and $\sum_{j=1}^{p} \Lambda_j(\cdot|\mathbb{Z})$.

Recall that we have, for $j = 1, \ldots, p+1$:

$$dN_{\cdot j}(s) = dM_{\cdot j}(s) + \alpha_j Y(s) d\Lambda_{0j}(s) + \alpha_j n S_1^T(s)\beta_j ds$$

where the α_j are defined in (4.2). Thus, an estimator of $\Lambda_{0j}(t)$ is given, for all $t \in [0, \tau]$, by

$$\hat{\Lambda}_{0j}(t) = \frac{1}{\hat{\alpha}_j} \int_0^t \frac{dN_{\cdot j}(s)}{Y(s)} - \hat{\beta}_j^T \int_0^t \frac{S_1(s)}{S_0(s)} ds, \tag{5.7}$$

where $\hat{\beta}_j$ is the estimator introduced in Section 3.1 and is such that $\mathcal{U}_j(\hat{\beta}_j, \alpha_j, \tau) = 0$.

Then we propose to estimate $\Lambda_j(\cdot|\mathbb{Z})$, for $j = 1, \ldots, p+1$, by $\hat{\Lambda}_j(\cdot|\mathbb{Z})$ defined, for $t \geq 0$, by

$$\begin{aligned} \hat{\Lambda}_j(t|\mathbb{Z}) &= \hat{\Lambda}_{0j}(t) + \hat{\beta}_j^T \int_0^t Z(s) ds \\ &= \frac{1}{\hat{\alpha}_j} \int_0^t \frac{dN_{\cdot j}(s)}{Y(s)} + \hat{\beta}_j^T \left(\int_0^t Z(s) ds - \int_0^t \frac{S_1(s)}{S_0(s)} ds \right), \end{aligned} \tag{5.8}$$

where by abuse of notation $\int_0^t Z(s)ds$ denotes the vector of integrals of each coordinate of $Z(\cdot)$. Let us write:

$$\hat{\Lambda}_{WMC}(\cdot|\mathbb{Z}) = \begin{pmatrix} \hat{\Lambda}_1(\cdot|\mathbb{Z}) \\ \vdots \\ \hat{\Lambda}_p(\cdot|\mathbb{Z}) \end{pmatrix}. \tag{5.9}$$

Before giving the large sample behavior of this estimator in the following theorem, let us introduce other notations. For $j = 1, \ldots, p+1$ let us define:

$$\begin{aligned} \mu_{1j}(\cdot) &= \alpha_j \int_0^{\cdot} \frac{s_0(s)\lambda_{0j}(s) + \beta_j^T s_1(s)}{s_0^2(s)} ds, \\ \mu_{2j}(\cdot) &= \frac{\alpha_j \alpha_j^*}{\theta(\tau)} \int_0^{\cdot} \frac{s_0(s)\lambda_{0j}(s) + \beta_j^T s_1(s)}{s_0(s)} ds, \\ \mu_{3j}(\cdot) &= \alpha_j \int_0^{\cdot} \frac{a(s)}{s_0(s)} ds \beta_j, \end{aligned}$$

where the α_j^*, for $j = 1, \ldots, p+1$, are defined in Proposition 4.1. Finally, let us define the block matrix

$$\Sigma_5(t) = \frac{1}{\alpha} \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & K(t) & 0 & 0 & \ldots & 0 & -\Lambda_1(t|\mathbb{Z}) \\ 0 & 1 & 0 & \ldots & 0 & 0 & K(t) & 0 & \ldots & 0 & -\Lambda_2(t|\mathbb{Z}) \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & 1 & 0 & 0 & \ldots & 0 & K(t) & 0 & -\Lambda_{p-1}(t|\mathbb{Z}) \\ 0 & \ldots & 0 & 0 & 1 & 0 & \ldots & 0 & 0 & K(t) & -\Lambda_p(t|\mathbb{Z}) \end{pmatrix}$$

where $K(t)$ is a $1 \times k$ vector defined by

$$K(t) = \left(\int_0^t Z(s)ds - \int_0^t \frac{s_1(s)}{s_0(s)} ds \right)^T A^{-1}(\tau), \text{ for all } t \in [0, \tau].$$

Theorem 5.4. *Under Assumptions A1-A6, the process*

$$\sqrt{n} \left(\hat{\Lambda}_{WMC}(\cdot|\mathbb{Z}) - (\Lambda_1(\cdot|\mathbb{Z}), \ldots, \Lambda_p(\cdot|\mathbb{Z}))^T \right)$$

converges weakly in $D^p[0, \tau]$, when $n \to +\infty$, to a zero mean Gaussian process with covariance matrix $\Sigma_5(t)\Sigma_{\mathbb{W}'_{p,\infty}}(t)\Sigma_5^T(t)$, where

$$\Sigma_{\mathbb{W}'_{p,\infty}}(t) = \frac{1}{\alpha} \begin{pmatrix} \Sigma^{(11)}_{\mathbb{W}'_{p,\infty}}(t) & \Sigma^{(12)}_{\mathbb{W}'_{p,\infty}}(t) \\ \Sigma^{(12)}_{\mathbb{W}'_{p,\infty}}(t)^T & \Sigma^{(22)}_{\mathbb{W}'_{p,\infty}}(t) \end{pmatrix},$$

with

$$\Sigma^{(11)}_{\mathbb{W}'_{p,\infty}}(t) = \begin{pmatrix} \mu_{11}(t) & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mu_{1p}(t) \end{pmatrix},$$

$$\Sigma^{(12)}_{\mathbb{W}'_{p,\infty}}(t) = \begin{pmatrix} \mu_{31}^T(t) & 0 & 0 & \ldots & 0 & \mu_{21}(t) \\ 0 & \mu_{32}^T(t) & 0 & \ldots & 0 & \mu_{22}(t) \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & \mu_{3(p-1)}^T(t) & 0 & \mu_{2(p-1)}(t) \\ 0 & \ldots & 0 & 0 & \mu_{3p}^T(t) & \mu_{2p}(t) \end{pmatrix},$$

and

$$\Sigma^{(22)}_{\mathbb{W}'_{p,\infty}}(t) = \begin{pmatrix} \Theta_1(\tau) & 0 & 0 & \ldots & 0 & \xi_1(\tau) \\ 0 & \Theta_2(\tau) & 0 & \ldots & 0 & \xi_2(\tau) \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & \Theta_{p-1}(\tau) & 0 & \xi_{p-1}(\tau) \\ 0 & \ldots & 0 & 0 & \Theta_p(\tau) & \xi_p(\tau) \\ \xi_1^T(\tau) & \ldots & \ldots & \ldots & \xi_p^T(\tau) & \xi_\alpha(\tau) \end{pmatrix}.$$

Proof. By the definition of $\Lambda_j(\cdot|\mathbb{Z})$ and $\hat{\Lambda}_j(\cdot|\mathbb{Z})$, one can write, for all $t > 0$ and $j = 1, \ldots, p$

$$\begin{aligned} & \sqrt{n}\left(\hat{\Lambda}_j(\cdot|\mathbb{Z}) - \Lambda_j(\cdot|\mathbb{Z})\right) \\ = \quad & \sqrt{n}\left[\left(\frac{1}{\hat{\alpha}}\int_0^t \frac{dN_{\cdot j}(s)}{Y(s)} - \hat{\beta}_j^T \int_0^t \frac{S_1(s)}{S_0(s)}ds - \Lambda_{0j}(t)\right) + (\hat{\beta}_j - \beta_j)^T \int_0^t Z(s)ds\right]. \end{aligned}$$

But it is easy to check that our assumptions and preliminary results yield

$$\begin{aligned} & \sqrt{n}\left(\frac{1}{\hat{\alpha}}\int_0^{\cdot} \frac{dN_{\cdot j}(s)}{Y(s)} - \hat{\beta}_j^T \int_0^{\cdot} \frac{S_1(s)}{S_0(s)}ds - \Lambda_{0j}(\cdot)\right) = \frac{\sqrt{n}}{\hat{\alpha}}\int_0^{\cdot} \frac{dM_{\cdot j}(s)}{Y(s)} \\ - \quad & \frac{\sqrt{n}(\hat{\alpha}-\alpha)}{\alpha}\left(\beta_j^T \int_0^{\cdot} \frac{s_1(s)}{s_0(s)}ds + \Lambda_{0j}(\cdot)\right) - \sqrt{n}(\hat{\beta}_j - \beta_j)^T \int_0^{\cdot} \frac{s_1(s)}{s_0(s)}ds + o_P(1). \end{aligned}$$

These two last equations and (4.6) enable us to write:

$$\begin{aligned} & \sqrt{n}\left(\hat{\Lambda}_j(\cdot|\mathbb{Z}) - \Lambda_j(\cdot|\mathbb{Z})\right) = \frac{\sqrt{n}}{\hat{\alpha}}\int_0^{\cdot} \frac{dM_{\cdot j}(s)}{Y(s)} - \frac{\sqrt{n}(\hat{\alpha}-\alpha)}{\alpha}\Lambda_j(\cdot|\mathbb{Z}) \\ + \quad & \frac{1}{\alpha}\left(A^{-1}(\tau)\frac{\mathcal{U}_j(\beta_j,\alpha_j,\tau)}{\sqrt{n}}\right)^T \left(\int_0^{\cdot} Z(s)ds - \int_0^{\cdot} \frac{s_1(s)}{s_0(s)}ds\right) + o_P(1). \end{aligned}$$

Hence we have

$$\sqrt{n}\begin{pmatrix} \hat{\Lambda}_1(\cdot|\mathbb{Z}) - \Lambda_1(\cdot|\mathbb{Z}) \\ \vdots \\ \hat{\Lambda}_p(\cdot|\mathbb{Z}) - \Lambda_p(\cdot|\mathbb{Z}) \end{pmatrix} = \Sigma_5(\cdot)\mathbb{W}'_{p,n}(\cdot) + o_P(1),$$

where

$$\begin{aligned} \mathbb{W}'_{p,n}(\cdot) \quad = \quad & \left(\sqrt{n}\int_0^{\cdot} \frac{dM_{\cdot 1}(s)}{Y(s)}, \ldots, \sqrt{n}\int_0^{\cdot} \frac{dM_{\cdot p}(s)}{Y(s)}, \right. \\ & \left. \frac{\mathcal{U}_1^T(\beta_1,\alpha_1,\tau)}{\sqrt{n}}, \ldots, \frac{\mathcal{U}_p^T(\beta_p,\alpha_p,\tau)}{\sqrt{n}}, \frac{1}{\sqrt{n}}\sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta(\tau)} M_{\cdot j}(\tau)\right)^T. \end{aligned}$$

But, a straightforward application of Rebolledo's Theorem proves that, under Assumptions A1-A6, the process

$$\begin{aligned} \mathbb{W}_{p,n}(\cdot) \quad = \quad & \left(\sqrt{n}\int_0^{\cdot} \frac{dM_{\cdot 1}(s)}{Y(s)}, \ldots, \sqrt{n}\int_0^{\cdot} \frac{dM_{\cdot p}(s)}{Y(s)}, \right. \\ & \left. \frac{\mathcal{U}_1^T(\beta_1,\alpha_1,\cdot)}{\sqrt{n}}, \ldots, \frac{\mathcal{U}_p^T(\beta_p,\alpha_p,\cdot)}{\sqrt{n}}, \frac{1}{\sqrt{n}}\sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta(\tau)} M_{\cdot j}(\cdot)\right)^T \end{aligned}$$

converges weakly in $D^{p(k+1)+1}[0,\tau]$ to a zero mean multivariate Gaussian martingale $\mathbb{W}_{p,\infty}(\cdot)$. It is easy to check that the covariance matrix $\Sigma_{\mathbb{W}_{p,\infty}}(\cdot)$ of this

limit process is similar to $\Sigma_{\mathbb{W}'_{p,\infty}}(\cdot)$ where τ is everywhere replaced by t. This ends the proof. □

Estimators $\hat{\Theta}_1(\tau), \ldots, \hat{\Theta}_p(\tau), \hat{\xi}_1(\tau), \ldots, \hat{\xi}_p(\tau)$ and $\hat{\xi}_\alpha(\tau)$ of respectively $\Theta_1(\tau)$, $\Theta_2(\tau), \ldots, \Theta_p(\tau)$, $\xi_1(\tau), \ldots, \xi_p(\tau)$ and $\xi_\alpha(\tau)$ have been introduced in Section 4.4. Moreover, one can estimate $\mu_{1j}(t), \mu_{2j}(t), \mu_{3j}(t)$, for $j = 1, \ldots, p+1$ and $t \in [0, \tau]$, with

$$\begin{aligned}
\hat{\mu}_{1j}(t) &= n \int_0^t \frac{dN_{\cdot j}}{Y^2(s)}, \\
\hat{\mu}_{2j}(t) &= \frac{\hat{\alpha}_j^*}{\hat{\theta}(\tau)} \int_0^t \frac{dN_{\cdot j}(s)}{Y(s)}, \\
\hat{\mu}_{3j}(t) &= \hat{\alpha} \int_0^t \frac{S_2(s) - S_1^{\otimes 2}(s)/S_0(s)}{S_0(s)} \hat{\beta}_j
\end{aligned}$$

and an estimator of $K(t)$ is given by:

$$\hat{K}(t) = \left(\int_0^t Z(s)ds - \int_0^t \frac{S_1(s)}{S_0(s)} ds \right)^T \hat{A}^{-1}(\tau).$$

Using these estimators one can get consistent estimators of the matrices $\Sigma_5(t)$ and $\Sigma_{\mathbb{W}'_{p,\infty}}(t)$.

5.2.2. Improved estimators of the cause-specific cumulative hazard rate functions: the T-optimal estimators

The estimators of the cause-specific (baseline) hazard rate functions introduced in the previous paragraph can be improved thanks to the following two arguments:

- replace in equation (5.7) and (5.8) the estimator $\hat{\beta}_{WMC}$ by the improved estimator $\tilde{\beta}_{Topt}$ of the regression parameters;
- apply the idea of Section 5.1.2, i.e. use the estimation of the sum of the cause-specific (baseline) hazard rate functions in order to obtain optimal estimators of each quantity.

Thus, let us first consider the new estimators of $\Lambda_{0j}(t)$ and $\Lambda_j(t|\mathbb{Z})$, for $j = 1, \ldots, p+1$, respectively given by:

$$\tilde{\Lambda}_{0j}(t) = \frac{1}{\hat{\alpha}_j} \int_0^t \frac{dN_{\cdot j}(s)}{Y(s)} - \tilde{\beta}_j^T \int_0^t \frac{S_1(s)}{S_0(s)} ds \tag{5.10}$$

and

$$\tilde{\Lambda}_j(t|\mathbb{Z}) = \tilde{\Lambda}_{0j}(t) + \tilde{\beta}_j^T \int_0^t Z(s)ds,$$

where $\tilde{\beta}_{p+1} = \sum_{j=1}^p \tilde{\beta}_j$.

Some supplementary notations are needed before describing the large sample behavior of these new estimators. Let

- $H^{opt}_{j\bullet} = (H^{opt}_{j1}, \ldots, H^{opt}_{jp}, H^{opt}_{jp+1})$, for $j = 1, \ldots, p$, be the jth row of the optimal block matrix H^{opt} defined in the proof of Theorem 4.4 and thus a matrix of dimension $k \times (p+1)k$;
- $H^{opt}_{+\bullet} = \sum_{j=1}^{p} H^{opt}_{j\bullet}$, a matrix of dimension $k \times (p+1)k$;
- $\bar{\Lambda}_j(t) = \Lambda_{0j}(t) + \beta_j^T \int_0^t \frac{s_1(s)}{s_0(s)} ds$, for $j = 1, \ldots, p+1$;
- $\Gamma_j(t) = K(t)A(\tau)H^{opt}_{j\bullet}\Sigma_1(\tau) - (0, \ldots, 0, \bar{\Lambda}_j(t)/\alpha_j)$, for $j = 1, \ldots, p$, and $\Gamma_{p+1}(t) = K(t)A(\tau)H^{opt}_{+\bullet}\Sigma_1(\tau) + (0, \ldots, 0, \bar{\Lambda}_{p+1}(t)/\alpha_{p+1})$, all these matrices being of dimension $1 \times [k(p+1)+1]$.

Finally, let us define the block matrix $\Sigma_6(t)$, for all $t \in [0, \tau]$, by:

$$\Sigma_6(t) = \begin{pmatrix} 1/\alpha_1 & 0 & \cdots & 0 & \Gamma_1(t) \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & 1/\alpha_p & 0 & \Gamma_p(t) \\ 0 & \cdots & 0 & 1/\alpha_{p+1} & \Gamma_{p+1}(t) \end{pmatrix}.$$

The large sample behavior of the above estimators of the cause-specific cumulative hazard rate functions is given in next Theorem.

Theorem 5.5. *Under Assumptions A1-A6, the process*

$$\sqrt{n} \begin{pmatrix} \tilde{\Lambda}_1(\cdot|\mathbb{Z}) - \Lambda_1(\cdot|\mathbb{Z}) \\ \vdots \\ \tilde{\Lambda}_p(\cdot|\mathbb{Z}) - \Lambda_p(\cdot|\mathbb{Z}) \\ \tilde{\Lambda}_{p+1}(\cdot|\mathbb{Z}) - \Lambda_{p+1}(\cdot|\mathbb{Z}) \end{pmatrix}$$

converges weakly in $D^{p+1}[0, \tau]$, when $n \to +\infty$, to a zero mean Gaussian process with covariance matrix $\Sigma_6(t)\Sigma_{\mathbb{W}'_{p+1,\infty}}(t)\Sigma_6^T(t)$, where

$$\Sigma_{\mathbb{W}'_{p+1,\infty}}(t) = \begin{pmatrix} \Sigma^{(11)}_{\mathbb{W}'_{p+1,\infty}}(t) & \Sigma^{(12)}_{\mathbb{W}'_{p+1,\infty}}(t) \\ \Sigma^{(12)}_{\mathbb{W}'_{p+1,\infty}}(t)^T & \Sigma^{(22)}_{\mathbb{W}'_{p+1,\infty}}(t) \end{pmatrix},$$

with

$$\Sigma^{(11)}_{\mathbb{W}'_{p+1,\infty}}(t) = \begin{pmatrix} \mu_{11}(t) & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \mu_{1p}(t) & 0 \\ 0 & \cdots & 0 & \mu_{1p+1}(t) \end{pmatrix}$$

$$\Sigma^{(12)}_{\mathbb{W}'_{p+1,\infty}}(t) = \begin{pmatrix} \mu_{31}^T(t) & 0 & 0 & \ldots & 0 & \mu_{21}(t) \\ 0 & \mu_{32}^T(t) & 0 & \ldots & 0 & \mu_{22}(t) \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & \mu_{3p}^T(t) & 0 & \mu_{2p}(t) \\ 0 & \ldots & 0 & 0 & \mu_{3(p+1)}^T(t) & \mu_{2(p+1)}(t) \end{pmatrix},$$

and

$$\Sigma^{(22)}_{\mathbb{W}'_{p+1,\infty}}(t) = \Sigma_{\mathbb{U}_\infty}(t) = \begin{pmatrix} \Theta_1(t) & 0 & 0 & \xi_1(t) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \Theta_{p+1}(t) & \xi_{p+1}(t) \\ \xi_1^T(t) & \cdots & \xi_{p+1}^T(t) & \xi_\alpha(t) \end{pmatrix}.$$

Proof. One can prove that, for all $j = 1, \ldots, p$ and $t \in [0, \tau]$, we have:

$$\begin{aligned} \sqrt{n}\left(\tilde{\Lambda}_j(t|\mathbb{Z}) - \Lambda_j(t|\mathbb{Z})\right) &= \frac{\sqrt{n}}{\alpha_j}\int_0^t \frac{dM_{\cdot j}(s)}{Y(s)} - \frac{\bar{\Lambda}_j(t)}{\alpha_j}\sum_{l=1}^{p+1}\frac{\alpha_l^*}{\theta(\tau)}\frac{M_{\cdot l}(\tau)}{\sqrt{n}} \\ &+ K(t)A(\tau)H^{opt}_{j\bullet}\Sigma_1(\tau)\mathbb{U}_n(\tau) + o_P(1), \end{aligned}$$

where $o_P(1)$, here and in the following, is uniform for $t \in [0, \tau]$. Similarly, one obtains:

$$\begin{aligned} \sqrt{n}\left(\tilde{\Lambda}_{p+1}(t|\mathbb{Z}) - \Lambda_{p+1}(t|\mathbb{Z})\right) &= \frac{\sqrt{n}}{\alpha_{p+1}}\int_0^t \frac{dM_{\cdot p+1}(s)}{Y(s)} + \frac{\bar{\Lambda}_{p+1}(t)}{\alpha_{p+1}}\sum_{l=1}^{p+1}\frac{\alpha_l^*}{\theta(\tau)}\frac{M_{\cdot l}(\tau)}{\sqrt{n}} \\ &+ K(t)A(\tau)H^{opt}_{+\bullet}\Sigma_1(\tau)\mathbb{U}_n(\tau) + o_P(1). \end{aligned}$$

Hence

$$\sqrt{n}\begin{pmatrix} \tilde{\Lambda}_1(\cdot|\mathbb{Z}) - \Lambda_1(\cdot|\mathbb{Z}) \\ \vdots \\ \tilde{\Lambda}_p(\cdot|\mathbb{Z}) - \Lambda_p(\cdot|\mathbb{Z}) \\ \tilde{\Lambda}_{p+1}(\cdot|\mathbb{Z}) - \Lambda_{p+1}(\cdot|\mathbb{Z}) \end{pmatrix} = \Sigma_6(\cdot)W'_{p+1,n}(\cdot) + o_P(1),$$

where

$$\begin{aligned} \mathbb{W}'_{p+1,n}(\cdot) = \Big(&\sqrt{n}\textstyle\int_0^\cdot \frac{dM_{\cdot 1}(s)}{Y(s)}, \ldots, \sqrt{n}\int_0^\cdot \frac{dM_{\cdot p}(s)}{Y(s)}, \sqrt{n}\int_0^\cdot \frac{dM_{\cdot p+1}(s)}{Y(s)}, \\ &\frac{\mathcal{U}_1^T(\beta_1,\alpha_1,\tau)}{\sqrt{n}}, \ldots, \frac{\mathcal{U}_p^T(\beta_p,\alpha_p,\tau)}{\sqrt{n}}, \frac{\mathcal{U}_{p+1}^T(\beta_{p+1},\alpha_{p+1},\tau)}{\sqrt{n}}, \frac{1}{\sqrt{n}}\textstyle\sum_{j=1}^{p+1}\frac{\alpha_j^*}{\theta(\tau)}M_{\cdot j}(\tau)\Big)^T. \end{aligned}$$

The weak convergence of $W'_{p+1,n}(\cdot)$ to $W'_{p+1,\infty}(\cdot)$ in $D^{(k+1)(p+1)+1}[0, \tau]$ is easily obtained thanks to arguments similar than those used in previous sections (see in particular the proofs of Proposition 4.1 and Theorem 5.1). This ends the proof. □

Recall that estimators of the quantities which appears in the limiting covariance matrix have been introduced in previous sections (see in particular Section 4.4 and Section 5.2.1).

Now, T-optimal estimators of the (baseline) cause-specific cumulative hazard rate functions can be obtained following the line of Section 5.1.2. For example,

from the estimator $\tilde{\Lambda}_{0j}(\cdot)$, for $j = 1, \ldots, p+1$, given in equation (5.10), one can obtain T-optimal estimators of the baseline cause-specific cumulative hazard rate functions. Indeed, write

$$\tilde{\Lambda}_0(\cdot) = \begin{pmatrix} \tilde{\Lambda}_{01}(\cdot) \\ \vdots \\ \tilde{\Lambda}_{0p}(\cdot) \\ \tilde{\Lambda}_{0p+1}(\cdot) \end{pmatrix}$$

and

$$\hat{H}(t) = \underset{H \in \mathcal{H}'}{\operatorname{argmin}} \operatorname{trace}(H \hat{\Sigma}_{\tilde{\Lambda}_0}(t) H^T),$$

where $\hat{\Sigma}_{\tilde{\Lambda}_0}(t)$ is the estimator of the asymptotic covariance matrix of the estimator $\tilde{\Lambda}_0(t)$. Then the T-optimal estimator of $\Lambda_0(\cdot) = (\Lambda_{01}(\cdot), \cdots, \Lambda_{0p}(\cdot))^T$ is given by:

$$\tilde{\Lambda}_0^{Topt}(\cdot) = \hat{H}(\cdot)\tilde{\Lambda}_0(\cdot). \tag{5.11}$$

Finally, it has to be noted that, as in Section 5.1.3, it is also possible to get estimators of the survival function and the cumulative incidence functions as well as to get their large sample behavior (like in Theorem 5.3). But this is omitted here since it is very similar to what has been done in Section 5.1.3. Let us just note that, from a practical point of view, if for some $t \in [0, \tau]$ the determinant of the matrix $\hat{\Sigma}_{\tilde{\Lambda}_0}(t)$ is too small, then we set $\tilde{\Lambda}_0^{Topt}(t) = (\tilde{\Lambda}_{01}(t), \ldots, \tilde{\Lambda}_{0p}(t))^T$.

6. Semiparametric inference under the MAR Assumption

Following the request of the Associate Editor and the referees, we have considered the possibility of weakening the assumption on the missingness mechanism. Indeed, one knows that the MCAR assumption can be too strong in some situations and it could be of interest to consider the case where the probability that the cause is missing depends on the time X as in the MAR case. The aim of this section is to consider such a situation. In this Section, we show that under the MAR assumption, one can obtain estimators of the Euclidean parameters (the regression parameters) as well as the functional parameters (the baseline cause-specific cumulative hazard rate function and, as a consequence, of the Cumulative Incidence Function and the survival function). Our approach is based on the double robust technique already used by Lu and Liang (2008) but it generalizes their work in the following ways. We do not assume independence between the causes of death; we assume an additive form on each cause-specific hazard rate function and, finally, we obtain estimators of the cumulative incidence functions as well as the survival function of the lifetime T (all causes confounded). As mentioned earlier in Section 1, there is, however, a disadvantage in considering such a more general situation since it won't be possible to improve the estimators as done in the previous section.

6.1. MAR Assumption

In this section we still assume that an additive hazard model holds on the cause-specific hazard rate function, that is:

$$\lambda_j(t|\mathbb{Z}) = \lim_{h\to 0^+} \frac{1}{h} P(T \in [t, t+h[, d = j | T \geq t, \mathbb{Z}(t)) = \lambda_{0j}(t) + \beta_j^T Z(t), \quad t \geq 0, \tag{6.1}$$

for $j = 1, \ldots, p$, where $\lambda_{0j}(\cdot)$ is the baseline jth cause-specific hazard rate function, $\mathbb{Z}(t) = \{Z(u); u \leq t\}$ is the history up to time t of this covariate process and $\beta_j \in \mathbb{R}^k$ is the vector of regression parameters associated to the jth cause.

But we only assume a MAR assumption (Rubin, 1976) on the mechanism of missing information on the cause of death (and not a MCAR assumption as in previous sections). More precisely, we suppose that:

$$P(R = 1|d, \delta = 1, X, \mathbb{Z}, G) = P(R = 1|\delta = 1, X, \mathbb{Z}, G) = \pi(X, \mathbb{Z}, G) = \pi(\mathbb{O}),$$

where G is an auxiliary covariate collected for each individual and where $\mathbb{O}$ denotes $(X, \mathbb{Z}, G)$ to shorten notation. Thus, given $\delta = 1$ and $\mathbb{O}$, the probability that the cause is missing depends only on $\mathbb{O}$ (which is observable) and not on the cause of death (not always observed). It is also assumed that

$$P(R = 0|d, \delta = 0, \mathbb{O}) = 1.$$

The above assumption of independence between R and d, given $\delta = 1$ and $\mathbb{O}$, allows us to write, for $j = 1, \ldots, p$:

$$P(d = j|\delta = 1, R = 0, \mathbb{O}) = P(d = j|\delta = 1, R = 1, \mathbb{O}) = P(d = j|\delta = 1, \mathbb{O}).$$

Assuming that these probabilities don't depend on the auxiliary covariate G, let us denote them $h_j(\mathbb{Q})$, for $j = 1, \ldots, p$, where $\mathbb{Q} = (X, \mathbb{Z})$.

In the following, we will assume that we observe a sample

$$(X_i, \delta_i, D_i, \mathbb{Z}_i(X_i), G_i)_{1\leq i\leq n}$$

of $(X, \delta, D, \mathbb{Z}(X), G)$, where we still write $D = \delta R d$.

6.2. Parametric model on the MAR mechanism

Following Lu and Liang (2008), one can introduce parametric models on the unknown probabilities $\pi(\mathbb{O})$ and $h_j(\mathbb{Q})$, for $j = 1, \ldots, p$. To this end, let us denote by $\pi(\mathbb{O}, \gamma)$ the parametric model posited on the probability $\pi(\mathbb{O})$, where γ is a regression parameter of dimension k_γ. A logistic regression model could be used, among others. Let us denote by $h_j(\mathbb{Q}, \zeta)$, where ζ is a regression parameter of dimension k_ζ, the parametric model used for the $h_j(\mathbb{Q})$, for $j = 1, \ldots, p$; for example, a multinomial logistic regression model. Due to the MAR assumption,

one can get estimators $\hat{\gamma}$ and $\hat{\zeta}$ of γ and ζ respectively by maximization of the likelihoods

$$\prod_{i=1}^{n} (\pi(\mathbb{O}_i, \gamma))^{R_i \delta_i} (1 - \pi(\mathbb{O}_i, \gamma))^{(1-R_i)\delta_i},$$

and

$$\prod_{i=1}^{n} \left[\prod_{j=1}^{p-1} (h_j(\mathbb{Q}_j, \zeta))^{I\{D_i=j\}} \right] \times (1 - h_1(\mathbb{Q}_i, \zeta) - \cdots - h_{p-1}(\mathbb{Q}_i, \zeta))^{I\{D_i=p\}}.$$

If their corresponding parametric models are correctly specified $\hat{\gamma}$ and $\hat{\zeta}$ are well-known to be consistent estimators of the true values of the parameters γ and ζ.

6.3. The double robust estimators of β_j and $\Lambda_{0j}(\cdot)$

Once estimators of γ and ζ are obtained, the question is how to estimate the regression parameters β_j and the cause-specific cumulative hazard rate functions $\Lambda_{0j}(\cdot)$, for $j = 1 \ldots, p$. One possible approach is to construct estimating equations based on the double robust technique introduced by Robins *et al.* (1994) and later used by Lu and Liang (2008) in a context similar to the present.

To this end, let us introduce two new families of counting processes:

$$\tilde{N}_{ij}(t) = I\{X_i \leq t, d_i = j\} \text{ and } N_i^*(t) = I\{X_i \leq t, \delta_i = 1\},$$

for $i = 1, \ldots, n$, $j = 1, \ldots, p$ and all $t > 0$. Note that, due to missingness of cause of death, the counting process $\tilde{N}_{ij}(\cdot)$ are not always observable, but will always appear in the sequel multiplied by R_i (itself given by the observation of δ_i and D_i), which ensures that the product is observable.

Estimators of the β_j and the $\Lambda_{0j}(\cdot)$, for $j = 1, \ldots, p$, will be obtained as solutions of the following estimating equations:

$$\begin{aligned} \sum_{i=1}^{n} \quad & \left\{ \frac{R_i}{\pi(\mathbb{O}_i, \hat{\gamma})} d\tilde{N}_{ij}(t) - \frac{R_i - \pi(\mathbb{O}_i, \hat{\gamma})}{\pi(\mathbb{O}_i, \hat{\gamma})} h_j(\mathbb{Q}_i, \hat{\zeta}) dN_i^*(t) \right. \\ & \left. - Y_i(t)\beta_j^T Z_i(t)dt - Y_i(t)d\Lambda_{0j}(t) \right\} = 0, \text{ for all } t > 0 \end{aligned} \tag{6.2}$$

$$\begin{aligned} \sum_{i=1}^{n} \quad & \int_0^{\tau} Z_i(t) \left\{ \frac{R_i}{\pi(\mathbb{O}_i, \hat{\gamma})} d\tilde{N}_{ij}(t) - \frac{R_i - \pi(\mathbb{O}_i, \hat{\gamma})}{\pi(\mathbb{O}_i, \hat{\gamma})} h_j(\mathbb{Q}_i, \hat{\zeta}) dN_i^*(t) \right. \\ & \left. - Y_i(t)\beta_j^T Z_i(t)dt - Y_i(t)d\Lambda_{0j}(t) \right\} = 0. \end{aligned} \tag{6.3}$$

Let γ_∞ and ζ_∞ be respectively the limit of the maximum likelihood estimators $\hat{\gamma}$ and $\hat{\zeta}$ introduced in previous subsection. As long as at least one of the two

models used on $\pi(\mathbb{O})$ and on the $h_j(\mathbb{Q})$, for $j = 1, \ldots, p$ is well specified, one can show that we have, for all $t > 0$:

$$\mathbb{E}\left[\frac{R_i}{\pi(\mathbb{O}_i, \hat{\gamma})}\tilde{N}_{ij}(t) - \frac{R_i - \pi(\mathbb{O}_i, \hat{\gamma})}{\pi(\mathbb{O}_i, \hat{\gamma})}h_j(\mathbb{Q}_i, \hat{\zeta})N_i^*(t) - \int_0^t Y_i(t)\left(\beta_j^T Z_i(t)dt - d\Lambda_{0j}(t)\right)\right] = 0.$$

This result is obtained thanks to consideration of the conditional expectation: given $\{\mathbb{O}_i, d_i, \delta_i = 1\}$, in the case where the model on $\pi(\mathbb{O})$ is correct; given $\{\mathbb{O}_i, R_i, \delta_i = 1\}$, in the case where it is the model on the $h_j(\mathbb{Q})$, for $j = 1, \ldots, p$ which is well specified.

From equations (6.2) and (6.3), one can get easily the expression of the Double Robust (DR) estimators:

$$\hat{\beta}_j = \hat{A}^{-1}(\tau)\frac{1}{n}\sum_{i=1}^n \int_0^\tau \left(Z_i(t) - \bar{Z}(t)\right)\left\{\frac{R_i}{\pi(\mathbb{O}_i, \hat{\gamma})}d\tilde{N}_{ij}(t) - \frac{R_i - \pi(\mathbb{O}_i, \hat{\gamma})}{\pi(\mathbb{O}_i, \hat{\gamma})}h_j(\mathbb{Q}_i, \hat{\zeta})dN_i^*(t)\right\}, \tag{6.4}$$

where $\hat{A}(\tau)$ and $\bar{Z}(t)$ are defined respectively in sections 4.3 and 3.1 and

$$\hat{\Lambda}_{0j}(\cdot) = \int_0^\cdot \frac{\sum_{i=1}^n \left\{\frac{R_i}{\pi(\mathbb{O}_i,\hat{\gamma})}d\tilde{N}_{ij}(t) - \frac{R_i - \pi(\mathbb{O}_i,\hat{\gamma})}{\pi(\mathbb{O}_i,\hat{\gamma})}h_j(\mathbb{Q}_i, \hat{\zeta})dN_i^*(t) - Y_i(t)\hat{\beta}_j^T Z_i(t)dt\right\}}{\sum_{i=1}^n Y_i(t)}, \tag{6.5}$$

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

Remark. In the case where there is no covariate, one can fix the estimators $\hat{\beta}_j$, for $j = 1, \ldots, p$, to zero in (6.5) to obtain nonparametric estimators $\hat{\Lambda}_j$ of the p cause-specific cumulative hazard functions.

Let us denote by

$$\hat{\beta}_{DR} = \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} \tag{6.6}$$

the vector of dimension kp concatenating the Double Robust estimators of the regression parameters and

$$\hat{\Lambda}_0^{DR}(\cdot) = \begin{pmatrix} \hat{\Lambda}_{01}(\cdot) \\ \vdots \\ \hat{\Lambda}_{0p}(\cdot) \end{pmatrix} \tag{6.7}$$

the Double Robust estimator of the baseline cause-specific hazard rate functions.

6.4. Large sample behavior

Let us introduce the following notations:

$$
\begin{aligned}
&M_{ij}^*(t,\beta_j,\gamma,\zeta,\Lambda_{0j}(\cdot)) = \frac{R_i}{\pi(\mathbb{O}_i,\gamma)}\tilde{N}_{ij}(t) - \frac{R_i-\pi(\mathbb{O}_i,\gamma)}{\pi(\mathbb{O}_i,\gamma)}h_j(\mathbb{Q}_i,\zeta)N_i^*(t)\\
&-\int_0^t Y_i(t)\left(\beta_j^T Z_i(t)dt - d\Lambda_{0j}(t)\right),\\
&B_{\gamma,j} = \mathbb{E}\left[\int_0^\tau \left(Z_1(t)-\frac{s_1(t)}{s_0(t)}\right)\frac{R_1\nabla_\gamma\pi(\mathbb{O}_1,\gamma_\infty)}{\pi^2(\mathbb{O}_1,\gamma_\infty)}\left(dN_{1j}(t)-h_j(\mathbb{Q}_1,\zeta_\infty)dN_1^*(t)\right)\right],\\
&B_{\zeta,j} = \mathbb{E}\left[\int_0^\tau \left(Z_1(t)-\frac{s_1(t)}{s_0(t)}\right)\frac{R_1-\pi(\mathbb{O}_1,\gamma_\infty)}{\pi(\mathbb{O}_1,\gamma_\infty)}\nabla_\zeta h_j(\mathbb{Q}_1,\zeta_\infty)dN_1^*(t)\right],\\
&S_{\gamma,i} = \frac{\delta_i\left(R_i-\pi(\mathbb{O}_i,\gamma_\infty)\right)}{\pi(\mathbb{O}_i,\gamma_\infty)\left(1-\pi(\mathbb{O}_i,\gamma_\infty)\right)}\nabla_\gamma\pi(\mathbb{O}_i,\gamma_\infty),\\
&S_{\zeta,i} = I\{D_i=1\}\frac{\nabla_\zeta h_1(\mathbb{Q}_i,\zeta_\infty)}{h_1(\mathbb{Q}_i,\zeta_\infty)} + \cdots + I\{D_i=p-1\}\frac{\nabla_\zeta h_{p-1}(\mathbb{Q}_i,\zeta_\infty)}{h_{p-1}(\mathbb{Q}_i,\zeta_\infty)}\\
&-I\{D_i=p\}\frac{\nabla_\zeta h_1(\mathbb{Q}_i,\zeta_\infty)+\cdots+\nabla_\zeta h_{p-1}(\mathbb{Q}_i,\zeta_\infty)}{1-h_1(\mathbb{Q}_i,\zeta_\infty)-\cdots-h_{p-1}(\mathbb{Q}_i,\zeta_\infty)},\\
&I_\gamma = \mathbb{E}\left(S_{\gamma,1}S_{\gamma,1}^T\right),\\
&I_\zeta = \mathbb{E}\left(S_{\zeta,1}S_{\zeta,1}^T\right),
\end{aligned}
$$

where $\nabla_u f(\cdot)$ denotes the gradient function of a function $f(\cdot)$ computed at u, and finally

$$
\phi_{i,j} = \int_0^\tau \left(Z_i(t)-\frac{s_1(t)}{s_0(t)}\right)dM_{ij}^*(t,\beta_j^0,\gamma_\infty,\zeta_\infty,\Lambda_{0j}^0(\cdot)) - B_{\gamma,j}^T I_\gamma^{-1}S_{\gamma,i} - B_{\zeta,j}^T I_\zeta^{-1}S_{\zeta,i},
$$

where β_j^0 and $\Lambda_{0j}^0(\cdot)$ are the true values of β_j and $\Lambda_{0j}(\cdot)$. Note that $S_{\gamma,i}$ and $S_{\zeta,i}$ are the score functions for individual i associated to the models on $\pi(\mathbb{O})$ and the $h_j(\mathbb{Q})$, for $j=1,\ldots,p$, respectively. The matrices I_γ and I_ζ are nothing but the corresponding Fisher Information matrices.

Theorem 6.1. *Under Assumptions A2-A4, A6 and if at least one of the two parametric models assumed on the MAR mechanism is well specified, then the r.v.*

$$
\sqrt{n}\left(\hat{\beta}_1-\beta_1,\ldots,\hat{\beta}_p-\beta_p\right)^T
$$

converges in distribution in $\mathbb{R}^{pk}$ to a mean zero Gaussian r.v. with covariance block matrix:

$$
\Sigma_{\hat{\beta},\infty} = \left(A^{-1}(\tau)\mathbb{E}(\phi_{1j}\phi_{1j'}^T)A^{-1}(\tau)\right)_{1\le j,j'\le p}.
$$

Proof. Since it is an adaptation of the proof of Theorem 1 of Lu and Liang (2008), only the skeleton of the proof is given. Fix j between 1 and p. The

solution of (6.2) in terms of $\Lambda_{0j}(\cdot)$, for all β_j in $\mathbb{R}^k$ is:

$$\hat{\Lambda}_{0j}(t,\beta_j) = \int_0^t \frac{\sum_{i=1}^n \left\{ \frac{R_i}{\pi(\mathbb{O}_i,\hat{\gamma})} d\tilde{N}_{ij}(t) - \frac{R_i - \pi(\mathbb{O}_i,\hat{\gamma})}{\pi(\mathbb{O}_i,\hat{\gamma})} h_j(\mathbb{Q}_i,\hat{\zeta}) dN_i^*(t) - Y_i(t)\beta_j^T Z_i(t) dt \right\}}{\sum_{i=1}^n Y_i(t)},$$

for all $t > 0$. Replacing $\Lambda_{0j}(\cdot)$ by $\hat{\Lambda}_{0j}(\cdot,\beta_j)$ in (6.3) one has now to solve the following equation in terms of β_j:

$$\sum_{i=1}^n \int_0^\tau \left(Z_i(t) - \bar{Z}(t)\right) \left\{ \frac{R_i}{\pi(\mathbb{O}_i,\hat{\gamma})} d\tilde{N}_{ij}(t) - \frac{R_i - \pi(\mathbb{O}_i,\hat{\gamma})}{\pi(\mathbb{O}_i,\hat{\gamma})} h_j(\mathbb{Q}_i,\hat{\zeta}) dN_i^*(t) - Y_i(t)\beta_j^T Z_i(t) \right\} = 0,$$

that we will briefly write

$$\mathcal{U}_j(\beta_j,\hat{\gamma},\hat{\zeta}) = 0.$$

A Taylor expansion of $\mathcal{U}_j(\beta_j^0,\hat{\gamma},\hat{\zeta})$ around γ_∞ and ζ_∞ and rather standard arguments in empirical process and maximum likelihood areas enable us to obtain after some cumbersome calculus:

$$\mathcal{U}_j(\beta_j^0,\hat{\gamma},\hat{\zeta}) = \sum_{i=1}^n \phi_{i,j} + o_P(\sqrt{n}).$$

On the other hand, since by definition of $\hat{\beta}_j$ one has $\mathcal{U}_j(\hat{\beta}_j,\hat{\gamma},\hat{\zeta}) = 0$, we can write:

$$\mathcal{U}_j(\beta_j^0,\hat{\gamma},\hat{\zeta}) = n\hat{A}(\tau)(\hat{\beta}_j - \beta_j^0).$$

Finally, we get

$$\sqrt{n}(\hat{\beta}_j - \beta_j^0) = \hat{A}^{-1}(\tau)\frac{1}{\sqrt{n}}\left(\sum_{i=1}^n \phi_{i,j} + o_P(\sqrt{n})\right)$$

and the result of the theorem follows from an application of the Central Limit Theorem. □

Note that the block (j,j'), for $1 \le j,j' \le p$, of the asymptotic covariance matrix given in the last theorem can be consistently estimated by

$$\hat{A}^{-1}(\tau)\left(\frac{1}{n}\sum_{i=1}^n \hat{\phi}_{i,j}\hat{\phi}_{i,j'}^T\right)\hat{A}^{-1}(\tau),$$

where

$$\hat{\phi}_{i,j} = \int_0^\tau \left(Z_i(t) - \bar{Z}(t)\right) dM_{ij}^*(t,\hat{\beta}_j,\hat{\gamma},\hat{\zeta},\hat{\Lambda}_{0j}(\cdot)) - \hat{B}_{\gamma,j}^T \hat{I}_\gamma^{-1}\hat{S}_{\gamma,i} - \hat{B}_{\zeta,j}^T \hat{I}_\zeta^{-1}\hat{S}_{\zeta,i},$$

and $\hat{B}_{\gamma,j}$, $\hat{B}_{\zeta,j}$, $\hat{I}_\gamma$, $\hat{I}_\zeta$, $\hat{S}_{\gamma,i}$ and $\hat{S}_{\zeta,i}$ are obtained from $B_{\gamma,j}$, $B_{\zeta,j}$, I_γ, I_ζ, $S_{\gamma,i}$ and $S_{\zeta,i}$ by replacement of γ_∞ and ζ_∞ by their maximum likelihood estimators, replacement of $s_0(\cdot)$ and $s_1(\cdot)$ by $S_0(\cdot)$ and $S_1(\cdot)$ and replacement of the expectations by their empirical counterparts.

Let us now consider the asymptotic behavior of the estimator of the baseline cause-specific cumulative hazard rate function $\Lambda_{0j}(\cdot)$. To this end, let us write:

$$C_{\gamma,j} = \mathbb{E}\left[\int_0^\tau \frac{R_1 \nabla_\gamma \pi(\mathbb{O}_1,\gamma_\infty)}{\pi^2(\mathbb{O}_1,\gamma_\infty)s_0(t)}\left(dN_{1j}(t) - h_j(\mathbb{Q}_1,\zeta_\infty)dN_1^*(t)\right)\right],$$
$$C_{\zeta,j} = \mathbb{E}\left[\int_0^\tau \frac{R_1 - \pi(\mathbb{O}_1,\gamma_\infty)}{\pi(\mathbb{O}_1,\gamma_\infty)s_0(t)}\nabla_\zeta h_j(\mathbb{Q}_1,\zeta_\infty)dN_1^*(t)\right]$$

and

$$\begin{aligned}\psi_{i,j}(t) &= \int_0^t \frac{dM_{ij}^*(t,\beta_j^0,\gamma_\infty,\zeta_\infty,\Lambda_{0j}^0(\cdot))}{s_0(t)} - \left(\int_0^t \frac{s_1(u)}{s_0(u)}du\right)^T A^{-1}(\tau)\phi_{i,j} \\ &- C_{\gamma,j}^T I_\gamma^{-1} S_{\gamma,i} - C_{\zeta,j}^T I_\zeta^{-1} S_{\zeta,i},\end{aligned}$$

Theorem 6.2. *Under Assumptions A2-A4, A6 and if at least one of the two parametric models assumed on the MAR mechanism is well specified, then the multivariate process*

$$\sqrt{n}\left(\hat{\Lambda}_{01}(\cdot) - \Lambda_{01}^0(\cdot),\ldots,\hat{\Lambda}_{0p}(\cdot) - \Lambda_{0p}^0(\cdot)\right)^T$$

converges weakly in $D^p[0,\tau]$, when $n \to +\infty$, to a zero mean Gaussian process with block covariance matrix functions given by

$$\mathbb{E}\left(\psi_{1,j}(s)\psi_{1,j'}^T(t)\right),\ \textit{for } (s,t) \in [0,\tau]^2 \textit{ and } 1 \le j,j' \le p.$$

As in the case of the previous theorem, the blocks of the above asymptotic covariance matrix function can also be consistently estimated by

$$\frac{1}{n}\sum_{i=1}^n \hat{\psi}_{i,j}(s)\hat{\psi}_{i,j'}^T(t)$$

where $\hat{\psi}_{i,j}(\cdot)$ is obtained from $\psi_{i,j}(\cdot)$ with replacement of the unknown parameters by their estimators and the expectations by their empirical counterparts.

Proof of the Theorem. Again, the proof is on the same lines as in Lu and Liang (2008). We only mention here that using empirical processes arguments and a Taylor expansion around γ_∞ and ζ_∞, one can write, for $j = 1,\ldots,p$ and $t \in [0,\tau]$:

$$\sqrt{n}\left(\hat{\Lambda}_{01}(t) - \Lambda_{01}^0(t)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_{i,j}(t) + o_P(1)$$

and the functional central limit theorem completes the proof. □

We are now in a position to consider the estimation of the cause-specific cumulative hazard rate function, given the covariate $\mathbb{Z}$. Recall that the additive hazard model considered in this paper assumes that we have, for all t in $[0,\tau]$:

$$\Lambda_j(t|\mathbb{Z}) = \Lambda_{0j}(t) + \int_0^t \beta_j^T Z(u)du.$$

Thus, an estimator of the conditional cumulative hazard rate function is given by:

$$\hat{\Lambda}_j(t|\mathbb{Z}) = \hat{\Lambda}_{0j}(t) + \int_0^t \hat{\beta}_j^T Z(u)du,$$

where $\hat{\Lambda}_{0j}(\cdot)$ and $\hat{\beta}_j$ are the double robust estimators introduced in Section 6.3.

From the above, one can easily write, for all t in $[0,\tau]$ and $j = 1,\ldots,p$:

$$\sqrt{n}\left(\hat{\Lambda}_j(t|\mathbb{Z}) - \Lambda_j^0(t|\mathbb{Z})\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Psi_{i,j}(t,\mathbb{Z}) + o_P(1), \tag{6.8}$$

where

$$\Psi_{i,j}(t,\mathbb{Z}) = \psi_{i,j}(t) + \left(\int_0^t Z(u)du\right)^T A^{-1}(\tau)\phi_{ij}.$$

Theorem 6.3. *Under Assumptions A2-A4, A6 and if at least one of the two parametric models assumed on the MAR mechanism is well specified, then the multivariate process*

$$\sqrt{n}\left(\hat{\Lambda}_1(\cdot|\mathbb{Z}) - \Lambda_1^0(\cdot|\mathbb{Z}),\ldots,\hat{\Lambda}_p(\cdot|\mathbb{Z}) - \Lambda_p^0(\cdot|\mathbb{Z})\right)^T$$

converges weakly in $D^p[0,\tau]$, when $n \to +\infty$, to a zero mean Gaussian process with block covariance matrix function given by

$$\mathbb{E}\left(\Psi_{1,j}(s)\Psi_{1,j'}^T(t)\right), \text{ for } (s,t) \in [0,\tau]^2 \text{ and } 1 \le j,j' \le p.$$

The proof is omitted since the result immediately follows from equation (6.8) and the central limit theorem.

Again, the blocks of the above asymptotic covariance matrix function can be consistently estimated by:

$$\frac{1}{n}\sum_{i=1}^n \hat{\Psi}_{i,j}(s)\hat{\Psi}_{i,j'}^T(t),$$

where

$$\hat{\Psi}_{i,j}(t) = \hat{\psi}_{i,j}(t) + \left(\int_0^t Z(u)du\right)^T \hat{A}^{-1}(\tau)\hat{\phi}_{i,j}.$$

We end this section by mentioning that it is also possible to get estimators of the survival function and the cumulative incidence functions as well as to obtain their large sample behavior (as in Theorem 5.3). To keep this paper shorter, we omit the details which require just an application of the approach of Section 5.1.3: one need only use the Gaussian process which appears in the limit of Theorem 6.3 as the process $\mathbb{L}'(\cdot)$ in Theorem 5.3.

7. Numerical study

In this section we first present a simulation study to compare the properties of our estimators for different sample sizes, their robustness to model misspecification and their efficiency. Then we illustrate the use of our estimators through an application on three different real datasets.

The simulation work has been implemented with the R statistical software (see R Development Core Team, 2010).

7.1. Simulation study

7.1.1. Simulation design

We carried out Monte Carlo simulations with only two causes of death and a time-independent covariate Z of dimension 1 with continuous distribution. More precisely, let T_1 (resp. T_2) denote the continuous latent, or potential, survival time associated with risk/cause 1 (resp. 2). Hence, the individual lifetime T introduced in Section 2 is given by $T = \min(T_1, T_2)$ and $d = 1$ (resp. $d = 2$) if $T = T_1$ (resp. $T = T_2$). Note that since the covariate Z is time-independent, the notation $\mathbb{Z}$ is useless here and won't be used in the sequel.

We assumed that the bivariate survival function of (T_1, T_2), conditionally on $Z = z$, is given by

$$S_{T_1,T_2}(t_1, t_2|z) = \exp\left(-(\lambda_1 + \beta_1 z)t_1 - (\lambda_2 + \beta_2 z)t_2 - \lambda_3 t_1 t_2\right),$$

where λ_3 is such that $0 \leq \lambda_3 < \min(\lambda_1, \lambda_2)$. One can see that, conditionally on $Z = z$, the r.v. T_1 and T_2 are dependent if $\lambda_3 > 0$ and that the distribution of T_j is exponential with rate $\lambda_j + \beta_j z$ for $j = 1, 2$. It is also easy to check that the assumption of an additive hazard rate model is fulfilled in this case since, conditionally on $Z = z$, the cause-specific hazard rate function of T_j is

$$\lambda_j(t|z) = \lambda_j + \lambda_3 t + \beta_j z, \qquad \text{for } t \geq 0,$$

for $j = 1, 2$. Finally, we assumed that the covariate Z is uniformly distributed on $(0, \tau_Z)$ and that the censoring time has an exponential distribution with rate λ_C. Our computer program first simulates the covariate Z according to the uniform distribution, then simulates T_1, given Z, with an exponential distribution and finally simulates T_2, given (T_1, Z), by numerical inversion of its conditional cumulative distribution function.

We chose the following set of parameters: $(\lambda_1, \lambda_2, \lambda_3) = (1, 1, 0.5)$, $(\beta_1, \beta_2) = (1, 2)$, $\tau_Z = 2$ and $\lambda_C = 3$. These parameters yield empirically (based on 100 000 replications of the above simulation scenario): 38.2% of censoring, 25.4% of failure from cause 1, and 36.4% of failure from cause 2.

Then we considered separately two mechanisms of causes missingness:

- the MCAR model with $P(R = 1|X, Z, d, \delta = 1) = \alpha = 0.8$ which leads empirically to 12,6% of missing cause of failure, 20.2% of failure from cause 1, and 29.0% of failure from cause 2;
- the MAR model with a logistic regression model for the conditional probability $\pi(X, Z)$, i.e.

$$\pi(x, z) = P(R = 1|X = x, Z = z, d, \delta = 1) = \frac{e^{\gamma_0+\gamma_1 x+\gamma_2 z}}{1 + e^{\gamma_0+\gamma_1 x+\gamma_2 z}},$$

  denoted briefly in the sequel by $\text{logit}(P(R = 1|X, Z, d, \delta = 1)) = \gamma_0 + \gamma_1 X + \gamma_2 Z$. We chose $\gamma_0 = -0.0855$ and $\gamma_1 = \gamma_2 = 1$ which leads empirically to 17,1% of missing cause of failure, 17.4% of failure from cause 1 and 26.1% of failure from cause 2.

Three sample sizes have been considered in our simulation study: $n = 100, n = 400$ and $n = 1000$. This simulation design was replicated 10 000 times.

Note finally that, under the MAR assumption, the estimates are obtained with a logistic regression model assigned to the conditional probabilities $h_j(x, z) = P(d = j|\delta = 1, R = 0, X = x, Z = z)$, for $j = 1, 2$. More precisely:

$$h_1(x, z) = \frac{1}{1 + e^{\zeta_0+\zeta_1 x+\zeta_2 z}} \quad \text{and} \quad h_2(x, z) = 1 - h_1(x, z).$$

7.1.2. Regression parameter estimators

Let us first study the behavior of the three different estimators of the regression parameters introduced in previous sections:

- the WMC estimator $\hat{\beta}_{WMC}$ given in (3.2) which generalizes an approach introduced by Lin and Ying (1994) and doesn't take into account the data with missing cause;
- our T–optimal estimator $\tilde{\beta}_{Topt}$ introduced in (3.5) which uses in an optimal way the information carried by the missing causes in order to get an estimator of asymptotic minimal variance;
- The DR estimator $\hat{\beta}_{DR}$ introduced in (6.6) and obtained through a generalization of the Double Robust estimators of Lu and Liang (2008).

Recall that the two first estimators (WMC and T–optimal) are obtained under the MCAR assumption whereas the DR estimator is based on the MAR assumption.

Comparison of the three estimators under the MCAR and MAR models

From the 10 000 times replication of the simulation design described in Section 7.1.1 we derived, for each estimator, the Monte-Carlo estimates of the following quantities:

- the bias (Bias);
- the variance (Emp. Var.);
- the mean of the estimates of the theoretical variance given in the previous sections (Var. Mean);
- the coverage probability of the 95% confidence intervals (Coverage %).

Table 1 lists the results when the MCAR model was used for simulations whereas Table 2 gives the results for the MAR model.

TABLE 1

Simulation results for the three regression parameter estimators (WMC, T–optimal and DR) under MCAR assumption. Monte Carlo estimates of the bias (Bias), the variance (Emp. Var.), the mean of estimated variance (Var. Mean) and the coverage percentage (Coverage %).

n	Estimator	WMC		T–optimal		DR	
		β_1	β_2	β_1	β_2	β_1	β_2
100	Bias	0.021	0.057	0.014	0.032	0.022	0.046
	Emp. Var.	0.699	1.051	0.638	0.951	0.649	0.937
	Var. Mean	0.701	1.037	0.638	0.908	0.794	0.970
	Coverage %	94.5	94.7	94.6	94.4	98.1	97.3
400	Bias	0.000	0.016	-0.001	0.011	-0.001	0.016
	Emp. Var.	0.161	0.231	0.148	0.207	0.150	0.208
	Var. Mean	0.162	0.239	0.150	0.213	0.164	0.198
	Coverage %	94.8	95.4	94.8	95.0	96.7	95.1
1000	Bias	0.003	-0.004	0.003	-0.004	0.003	-0.002
	Emp. Var.	0.065	0.094	0.060	0.085	0.060	0.085
	Var. Mean	0.064	0.094	0.059	0.084	0.062	0.076
	Coverage %	94.8	95.0	94.8	94.8	95.6	94.0

Table 1 shows that, under the MCAR model, the T–optimal estimator has generally lower standard-errors than the others. Moreover it has the smallest bias in the case of a small sample size. For all the estimators, the larger is the sample size the better behaves the estimator. One can also see that, for all the estimators, the coverage probabilities are almost always close to the nominal value of 95%, even if for $n = 100$ the coverage probabilities are rather large for the DR estimator.

It is interesting to note in Table 2 that the DR estimator outperforms the two other estimators in terms of bias. The WMC estimator has a bad behavior with a significant bias and with a coverage percentage which deteriorates when the sample size increases. Contrary to the WMC estimator, the T–optimal estimator proves to be robust under the MAR assumption; its standard-error and its coverage probability appear to be competitive with respect to the ones of the DR estimator.

Efficiency with respect to the parametric model

Table 2
Simulation results for the three regression parameter estimators (WMC, T–optimal and DR) under MAR assumption. Monte Carlo estimates of the bias (Bias), the variance (Emp. Var.), the mean of estimated variance (Var. Mean) and the coverage percentage (Coverage %).

	Estimator	WMC		T–optimal		DR	
n		β_1	β_2	β_1	β_2	β_1	β_2
100	Bias	0.541	0.805	-0.069	-0.133	0.017	0.059
	Emp. Var.	0.814	1.246	0.613	0.929	0.748	1.049
	Var. Mean	0.799	1.197	0.689	0.950	0.878	1.054
	Coverage %	93.4	91.6	95.0	93.8	98.0	97.3
400	Bias	0.516	0.745	-0.051	-0.118	0.001	0.014
	Emp. Var.	0.185	0.278	0.140	0.202	0.162	0.221
	Var. Mean	0.185	0.277	0.162	0.224	0.174	0.208
	Coverage %	79.8	72.6	96.0	94.7	96.5	95.3
1000	Bias	0.512	0.721	-0.045	-0.121	0.002	-0.001
	Emp. Var.	0.074	0.112	0.056	0.083	0.064	0.089
	Var. Mean	0.073	0.109	0.064	0.089	0.066	0.079
	Coverage %	53.7	41.4	95.7	93.2	95.6	94.1

Since the model used for simulations is fully parametric, one can get maximum likelihood (ML) estimators of the unknown parameters. Indeed, it is easy to see that under the MCAR assumption the likelihood function is given by

$$L_n^{(\mathrm{MCAR})}(\lambda_1,\lambda_2,\lambda_3,\beta_1,\beta_2) = \prod_{i=1}^n \left(\lambda_1(X_i|Z_i)\right)^{I(D_i=1)} \times \left(\lambda_2(X_i|Z_i)\right)^{I(D_i=2)} \times \left(\lambda_1(X_i|Z_i)+\lambda_2(X_i|Z_i)\right)^{I(D_i=0,\delta_i=1)} \times S(X_i|Z_i),$$

where $\lambda_j(X_i|Z_i) = \lambda_j+\beta_j Z_j+\lambda_3 X_i$, for $j=1,2$, and $S(X_i|Z_i) = S_{T_1,T_2}(X_i,X_i|Z_i)$ is the conditional survival function of the lifetime T, given Z_i. On the other hand, under the MAR assumption the likelihood function is given by

$$L_n^{(\mathrm{MAR})}(\lambda_1,\lambda_2,\lambda_3,\beta_1,\beta_2,\gamma_0,\gamma_1,\gamma_2) = \prod_{i=1}^n \left[\left(\lambda_1(X_i|Z_i)I(D_i=1)+\lambda_2(X_i|Z_i)I(D_i=2)\right)\pi(X_i,Z_i) + \left(\lambda_1(X_i|Z_i)+\lambda_2(X_i|Z_i)\right)I(D_i=0,\delta_i=1)(1-\pi(X_i,Z_i))\right] \times S(X_i|Z_i),$$

where $\pi(X_i,Z_i) = \exp(\gamma_0+\gamma_1 X_i+\gamma_2 Z_i)/(1+\exp(\gamma_0+\gamma_1 X_i+\gamma_2 Z_i))$. Hence, we are able to obtain the ML estimator of the regression parameters (β_1,β_2) by maximizing $L_n^{(\mathrm{MCAR})}$ (resp. $L_n^{(\mathrm{MAR})}$) under the MCAR (resp. MAR) assumption.

To study the behavior of the ML estimator, we used the same criteria as for the three other estimators (WMC, T–optimal and DR). Thus, based on the 10 000 replications of the above simulation design, we obtained empirical estimates of its bias, variance, mean of the estimates of the asymptotic variance and coverage probability of the 95% confidence intervals. Note that the asymptotic variances are estimated by inverting the Hessian matrix of the log-likelihood function at its maximum.

We first provide a study of the behavior of the ML estimator under the MCAR and MAR assumptions. The results are summarized in Table 3. Irrespective of

the type of missingness, the ML estimator behaves correctly in case of moderate or large sample size but it has the weakness to provide conservative confidence intervals when the sample size is small.

Table 3
Simulation results for the ML regression parameter estimators under MCAR and MAR assumption. Monte Carlo estimates of the bias (Bias), the variance (Emp. Var.), the mean of estimated variance (Var. Mean) and the coverage percentage (Coverage %).

Data	n	Parameter	Bias	Emp. Var.	Var. Mean	Coverage %
MCAR	100	β_1	0.058	0.465	0.627	99.3
		β_2	0.033	0.726	0.825	96.4
	400	β_1	0.009	0.135	0.136	94.9
		β_2	0.024	0.179	0.182	95.0
	1000	β_1	0.007	0.055	0.054	94.5
		β_2	0.001	0.074	0.072	94.7
MAR	100	β_1	0.054	0.481	0.707	99.5
		β_2	0.028	0.757	0.917	97.0
	400	β_1	0.011	0.146	0.149	94.8
		β_2	0.022	0.194	0.197	95.0
	1000	β_1	0.060	0.059	0.059	94.8
		β_2	0.020	0.079	0.078	94.7

Let us now study the relative efficiency of our estimators (WMC, T–optimal or DR) with respect to the ML estimator. The relative efficiency is defined as the ratio of the empirical Mean Square Error (MSE) of one of our estimators with the empirical MSE of the ML estimator, both based on the same 10 000 simulated samples. These relative efficiencies are listed in Table 4. It is worth noting that the T–optimal and DR estimators have the same performance under MCAR while the WMC is less efficient. In case of MAR mechanism, the T–optimal estimator seems to have the better behavior. Finally, relative efficiencies of the WMC estimator are very large which confirms that this estimator should not be used in this case.

Table 4
Simulation results for the regression parameter estimators under MCAR and MAR assumption. Monte Carlo estimates of the relative efficiency.

Data	n	WMC vs. ML		T–optimal vs. ML		DR vs. ML	
		β_1	β_2	β_1	β_2	β_1	β_2
MCAR	100	1.493	1.450	1.363	1.309	1.387	1.292
	400	1.192	1.288	1.096	1.153	1.110	1.160
	1000	1.181	1.270	1.090	1.149	1.090	1.149
MAR	100	1.287	2.499	1.277	1.249	1.546	1.389
	400	3.088	4.283	0.976	1.110	1.109	1.137
	1000	5.370	7.958	0.927	1.230	1.022	1.121

7.1.3. Estimators of the baseline cause-specific cumulative hazard functions

Let us now study the behavior of the three different estimators of the baseline cause-specific cumulative hazard functions introduced in previous sections:

- the estimators $\hat{\Lambda}_{0j}(\cdot)$, for $j = 1, \ldots, p$, given in (5.7), which use data Without Missing Causes and called WMC estimators in the sequel;
- our T–optimal estimators of the baseline cause-specific cumulative hazard rate functions given in (5.11);
- The DR estimators introduced in (6.7) and obtained through a generalization of the Double Robust estimators of Lu and Liang (2008).

Recall that, here also, the two first estimators (WMC and T–optimal) are obtained under the MCAR assumption whereas the DR estimator is based on the MAR assumption.

We have used the criterion of the Mean Integrated Square Error (MISE) in order to compare the different estimators of the baseline cause-specific cumulative hazard rate functions. More precisely, for cause $j = 1, 2$ and for each type of estimator (WMC, T–optimal and DR), based on N simulated samples we obtained estimates $\hat{\Lambda}_{0j}^{(1)}(\cdot), \ldots, \hat{\Lambda}_{0j}^{(N)}(\cdot)$ of the baseline jth-cause-specific cumulative hazard rate function $\Lambda_j(\cdot)$. An empirical estimate of the MISE of this estimator is then given by:

$$\frac{1}{N} \sum_{k=1}^{N} \int_0^{\tau_k} \left(\hat{\Lambda}_{0j}^{(k)}(s) - \Lambda_{0j}(s) \right)^2 ds,$$

where τ_k is the $\lfloor 0.95n \rfloor$th order statistic of the observed durations in the kth simulated sample. Thus τ_k is nothing but an estimation of the 0.95–quantile of the distribution of X.

Table 5 lists the empirical MISE, based on $N = 1000$ samples, of the three types of estimators (WMC, T–optimal and DR) of the two baseline cause-specific cumulative hazard rate functions $\Lambda_{01}(\cdot)$ and $\Lambda_{02}(\cdot)$, under the MCAR model and also under the MAR model. One can see that the MISE is almost always lower for DR estimator than for the WMC and the T–optimal estimators. As expected, the difference in terms of MISE between the estimators is larger under MAR assumption than under the MCAR assumption where estimates of the MISE are similar. Under MAR assumption, the WMC has a bad behavior, even when the sample size increases. Again the T–optimal estimator outperforms the WMC estimator under the MCAR and MAR assumptions and, as for the regression parameter estimation, it reveals some robustness properties when the MCAR assumption is violated. It is not so for the WMC estimator.

Table 5
Simulation results for the three estimators (WMC, T–optimal and DR) of the baseline cause-specific cumulative hazard rate functions $\Lambda_{01}(\cdot)$ and $\Lambda_{02}(\cdot)$ under MCAR and MAR assumptions. Monte Carlo estimates of MISE ($\times 10^3$).

	Estimator	WMC		T–optimal		DR	
Data	n	$\Lambda_{01}(\cdot)$	$\Lambda_{02}(\cdot)$	$\Lambda_{01}(\cdot)$	$\Lambda_{02}(\cdot)$	$\Lambda_{01}(\cdot)$	$\Lambda_{02}(\cdot)$
MCAR	100	11.69	15.89	11.31	15.19	10.85	14.59
	400	3.05	4.02	2.94	3.84	2.87	3.67
	1000	1.29	1.63	1.23	1.52	1.17	1.37
MAR	100	16.22	23.56	12.29	15.9	13.28	16.43
	400	7.30	12.35	4.17	5.61	3.52	4.18
	1000	5.35	8.69	2.39	2.97	1.39	1.44

7.2. Applications on real datasets

7.2.1. Hodgkin's disease dataset

Let us first consider the real dataset presented in Pintilie (2006) of 616 patients treated with radiation therapy for a Hodgkin's disease at the Princess Margaret Hospital between 1968 and 1986. The event of interest is the incidence of second malignancy, the death without malignancy is seen as a competing risk. The question is whether the younger subjects (30 years old or younger) have the same risk of a new malignancy than the older ones. In this study there are 84 patients with incidence of second malignancy and 195 who died without a second malignancy. There is only one covariate, age, with 308 young patients and 308 older patients. The rate of censoring is about 54%. Let us call this dataset the "Without Missing Cause (WMC) dataset".

With the aim to apply our estimators, we have artificially created two new datasets with some causes of death missing. The first one is obtained from the WMC dataset after the application of a MCAR mechanism with $\alpha = 0.4$ and is called "MCAR dataset". The second one is also obtained from the WMC dataset but after the use of MAR mechanism with probability of missingness given by the logistic regression model $\text{logit}(\pi(X, Z)) = -1 + 0.01X + Z$. Let us call it the "MAR dataset". The different percentages of events for the three datasets are listed in Table 6.

Except for the WMC dataset where only the WMC estimator is used, we have applied our three estimators (WMC, T–optimal and DR) on these datasets. Table 7 lists the estimates of the regression parameters, their standard-errors and the 95% confidence intervals. Our results agree with those of Pintilie (2006). We find that, for all datasets and all estimation methods, an age larger than 30 is a risk factor, whatever the event of interest. The age factor increases more significantly the specific risk of death without second malignancy than the specific risk of incidence of second malignancy. For the MAR dataset, the T–optimal estimates are close to those obtained through the DR method, while the WMC estimates are rather far from results obtained on the initial WMC sample. This last point confirms the fact that the use of the WMC estimator is risky under MAR assumption.

Table 6
WMC, MCAR and MAR datasets on Hodgkin's disease. Percentages of events for incidence of second malignancy (cause 1) and death without malignancy (cause 2).

Data	Censoring	Missing	Cause 1	Cause 2
WMC	54.7	0	13.6	31.7
MCAR	54.7	28.0	5.0	12.3
MAR	54.7	26.1	4.4	14.8

We plotted in Figure 2 the different estimates we obtained, thanks of the WMC, T-optimal and DR methods under MCAR and MAR assumptions, of the two baseline cumulative risk functions associated respectively to the incidence of second malignancy and death without malignancy. The estimates of the cumulative incidence functions are presented in the book of Pintilie (2006). One can note for example that, the probability of a second malignancy for the younger subjects remains very low up to 15 years then it joins the value of the probability of death without second malignancy at about 25 years. This can explain the shape of the estimates of the baseline specific cumulative hazard functions presented in Figure 2. The estimate of the incidence of the second malignancy remains very low for a long period and then quickly increases and catches up or exceeds the estimate for the death without malignancy.

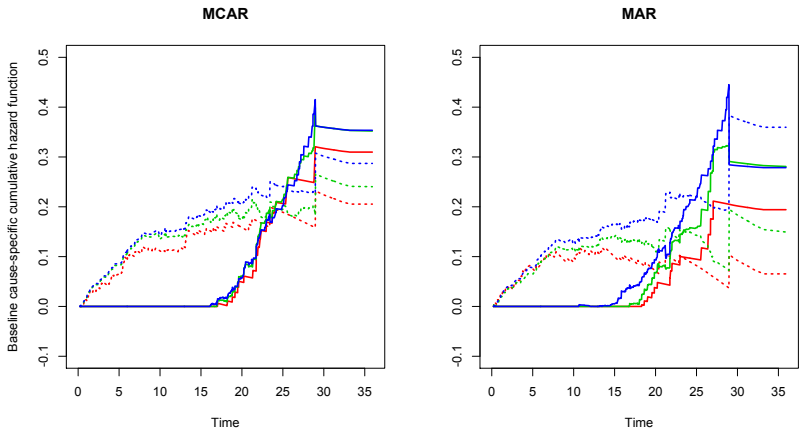
Table 7
WMC, MCAR and MAR datasets on Hodgkin's disease (Pintilie, 2006). WMC, T–optimal and DR estimates of the regression parameters, their estimated standard error (within parenthesis) and a 95% confidence interval for the incidence of second malignancy (β_1) and death without malignancy (β_1).

Data	Method		β_1	β_2
WMC	WMC	Estim. (SE)	0.00558 (0.00198)	0.02435 (0.00305)
		95% CI	[0.00169, 0.00946]	[0.01837, 0.03034]
MCAR	WMC	Estim. (SE)	0.01064 (0.00325)	0.02681 (0.00463)
		95% CI	[0.00426, 0.01701]	[0.01773, 0.03588]
	T–optimal	Estim. (SE)	0.00827 (0.00300)	0.02166 (0.00376)
		95% CI	[0.00239, 0.01142]	[0.01430, 0.02902]
	DR	Estim. (SE)	0.00877 (0.00258)	0.02069 (0.00358)
		95% CI	[0.00370, 0.01383]	[0.01367, 0.02772]
MAR	WMC	Estim. (SE)	0.01083 (0.00287)	0.03882 (0.00427)
		95% CI	[0.00521, 0.01645]	[0.03044, 0.04720]
	T–optimal	Estim. (SE)	0.00576 (0.00272)	0.02601 (0.00360)
		95% CI	[0.00043, 0.01109]	[0.01894, 0.03307]
	DR	Estim. (SE)	0.00567 (0.00278)	0.02399 (0.00368)
		95% CI	[0.00022, 0.01111]	[0.01678, 0.03120]

7.2.2. Breast Cancer dataset

Now, let us consider a dataset from a clinical trial in stage II breast cancer involving a population of elderly women. At the opposite of previous dataset, some

FIG 2. *Hodgkin's disease dataset (Pintilie, 2006). Estimates of the baseline incidence of second malignancy (solid) and death without malignancy (dotted) specific cumulative hazard functions using the WMC method (red), the T–optimal method (green) and the DR method (blue) under MCAR (left) and MAR (right) assumptions.*



patients had unknown cause of death. We use the same data than Cummings et al. (1986), Goetghebeur and Ryan (1995) and Lu and Tsiatis (2001) and, thus, we take into account two binary covariates: the presence of at least four positive nodes and the oestrogen receptor status (ER-negative) of their primary tumor, that were considered as significantly associated with overall survival.

A cause-specific survival analysis study using the additive hazards model is carried out for these data. Of the 169 eligible patients, 79 had died at the time of analysis and cause-of-death information was incomplete. Of these 79, cause of death was available for 61, of whom 44 were classified as dying from their disease and 17 from other causes. The remaining 18 patients had unknown cause of death. The estimates (WMC, T–optimal and DR) for death due to breast cancer are shown in Table 8. It appears that, with the WMC and T–optimal methods, a negative oestrogen receptor status significantly decreases the risk of death due to cancer with very similar estimates. We do no found this result with the DR method although the regression parameter estimate is not too far. The three methods lead to the conclusion that a high number of positive nodes is not associated with death due to breast cancer. The three estimates (thanks to WMC, T–optimal and DR methods) of the baseline cause-specific cumulative hazard rate function associated to death due to cancer are shown in Figure 3. Time is the number of days from randomization.

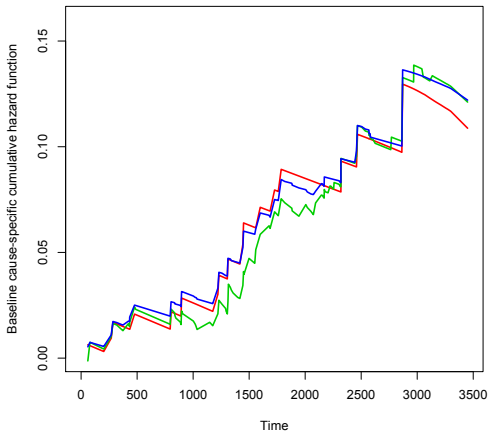
7.2.3. Hard drive dataset

In Section 5.1 we have considered the nonparametric case (i.e. whitout covariate) and have obtained WMC estimators and T-optimal estimators of the functions

TABLE 8
Breast Cancer dataset. WMC, T–optimal and DR estimation of the regression parameters for breast cancer data. For breast cancer cause of death are given WMC, T–optimal and DR estimates of the regression parameters, their estimated standard error (within parenthesis) and a 95% confidence interval.

Method	Covariate	Estim. (SE) ($\times 10^4$)	95% CI ($\times 10^4$)
WMC	nodes ≥ 4	0.533 (0.316)	[-0.086,1.152]
	ER–negative	-0.696 (0.190)	[-1.068,-0.324]
T-optimal	nodes ≥ 4	0.500 (0.303)	[-0.094,1.094]
	ER–negative	-0.696 (0.187)	[-1.063,-0.329]
DR	nodes ≥ 4	0.369 (0.279)	[-0.178,0.916]
	ER–negative	-0.613 (0.439)	[-1.473,0.247]

FIG 3. *Breast Cancer dataset. Estimates of the baseline cause-specific cumulative hazard rate function for the risk of death due to breast cancer using the WMC method (red), the T–optimal method (green) and the DR method (blue). Time is days from randomization.*



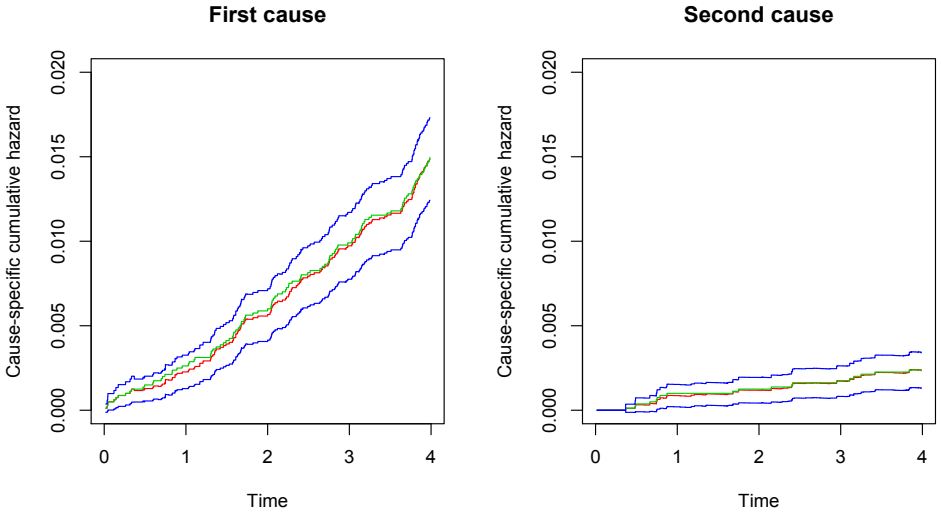
of interest. In order to illustrate this part, let us consider now a dataset in Reliability, without covariates, and introduced by Flehinger *et al.* (2002). These authors consider a scenario in which a company manufacturing computer hard drives tries to analyze the causes of failures of a certain sub-assembly. 10 000 hard drives were manufactured and only 172 items failed during a study period of 4 years. The 9828 other items still functioning at the end of the study are considered as censored observation with time 4 years.

Some of the causes of failure (such as "defective head") are related to components, but others (like "particle contamination") are not. In this dataset, there are three major causes of failures which, without going into details, are denoted as causes 1, 2 and 3. We assume that these causes act in series. Some of the failures were masked and a number of them were analyzed for complete resolu-

tion in a defect isolation laboratory. The observed masked groups were {1, 2, 3} and {1, 3}. Considering causes 1 and 3 as a single cause of failure, we obtain a new dataset with two competing modes of failure: mode 1 (corresponding to causes 1, 3 or masked group {1, 3} in the original data set) and mode 2 (corresponding to the cause 2 in the original data set). The failure cause is missing in this new dataset when none of the three original causes of failure is known (corresponding to the masked group {1, 2, 3} in the original data set). Finally, this new dataset contains 119 failures of type 1, 19 failures of type 2, 34 failures with missing cause and 9828 censored observations.

Assuming that the MCAR assumption holds, the probability α that the cause of failure is missing is empirically estimated by $\hat{\alpha} = 0.802$. We plotted in Figure 4 the WMC and T-optimal estimates of the two cause-specific cumulative hazard rate functions as well as the 95% pointwise confidence intervals of the T–optimal estimators. We can see that our T–optimal estimators are slightly more regular (with smaller size jumps) than the WMC estimators; however the T–optimal estimators may not be fully legitimate (they can be locally decreasing). However, the latter drawback disappears when the sample size increases because of uniform convergence of our estimators.

FIG 4. *Hard drives dataset (Flehinger et al. 2002). Estimates of the Cause-specific cumulative hazard functions for the two failure causes: WMC estimates (green lines); T–optimal estimates (red lines) with their 95% pointwise confidence intervals (blue lines).*



Acknowledgments

The authors are grateful to the editor, an associate editor and two anonymous referees for their constructive comments and suggestions that have led to consid-

erable improvement of an earlier version of the paper. Thanks also to Professor Syed Kirmani who checked the english writing of this paper. The authors are grateful to the Eastern Cooperative Oncology Group (ECOG) for providing the data on the breast cancer.

References

[1] Andersen, P.K., Borgan, O. Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer, New York.

[2] Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study, *Ann. Statist.*, **10**, 1100–1120.

[3] Bakoyannis, G., Siannis, F. and Touloumi, G. (2010). Modelling competing risks data with missing cause of failure. *Stat. Med.* **29** (30), 3172–3185.

[4] Balakrishnan, N., Bordes, L. and Zhao, X. (2010). Minimum-distance parametric estimation under progressive Type-I censoring, *IEEE Tr. Reliab.*, **59**, 413–425.

[5] Basu, S., Basu, A.P. and Mukhopadhyay, C. (1999). Bayesian analysis for masked system failure data using non-identical Weibull models, *J. Statist. Plann. Inference*, **78**, 255–275

[6] Craiu, R.V. and Duchesne, T. (2004). Inference based on the EM algorithm for the competing risks model with masked causes of failure, *Biometrika*, **91**, 543–558.

[7] Craiu, R.V. and Reiser, B. (2007). Inference for the dependent competing risks model with masked causes of failure, *Lifetime Data Analysis*, to appear.

[8] Crowder, M. (2001). *Classical Competing Risks*, Chapman & Hall/CRC Press, London.

[9] Cummings, F.J., Gray, R., Davis, T.E., Tormey, D.C., Harris, J.E., Falkson, G.G. and Arseneau, J. (1986). Tamoxifen versus placebo: double blind adjuvant trial in elderly woman with stage II breast cancer. *Nut. Cancer Inst. Monographs*, **1**, 119–23.

[10] Dauxois, J.Y. (2000). A new method for proving weak convergence results applied to nonparametric estimators in survival analysis, *Stochastic Processes and Their Applications*, **90** (2), 327–334.

[11] Dinse, G.E. (1986). Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death data, *J. Amer. Statist. Assoc.*, **81**, 328–336.

[12] Fine, J.P. and Gray, R. J. (2009). A proportional hazards model for the sub- distribution of a competing risk. *J. Amer. Statist. Assoc.*, **94**, 496–509.

[13] Flehinger, B.J., Reiser, R. and Yashchin, E. (2002). Parametric modeling for survival with competing risks and masked failure causes, *Lifetime Data Analysis*, **8**, 177–203.

[14] Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*, Wiley, New York.

[15] Gijbels, I., Lin, D.Y. and Ying, Z. (2007). Non- and semi-parametric analysis of failure time data with missing failure indicators. Complex

datasets and inverse problems, 203–223, *IMS Lecture Notes Monogr.* Ser., 54, Inst. Math. Statist., Beachwood, OH.

[16] Gao, G. and Tsiatis, A.A. (2005). Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure. *Biometrika* **92** (4), 875–891.

[17] Gill, R.D. and Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18** (4), 1501-1555.

[18] Goetghebeur, E. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing, *Biometrika*, **82**, 821–833.

[19] Guttman, I., Lin, D.K.J., Reiser, B. and Usher, J.S. (1995). Dependent masking and system life data analysis: Bayesian inference for two-component systems, *Lifetime Data Analysis*, **1**, 87–100.

[20] Lee, M., Cronin, K.A., Gail, M.H., Dignam, J.J. and Feuer, E.J. (2011). Multiple imputation methods for inference on cumulative incidence with missing cause of failure. *Biom. J.* **53** (6), 974–993.

[21] Lin, D.K.J., Usher, J.S., Guess, F.M. (1993). Exact maximum likelihood estimation under masked system data, *IEEE Trans. Reliability*, **42**, 631–635.

[22] Lin, D.Y. and Ying, Z. (1994). Semiparametric estimation of the additive risk model, *Biometrika*, **81**, 61–71.

[23] Lo, S.-H. (1991). Estimating a survival function with incomplete cause-of-death data, *J. Multivariate Anal.*, **39**, 217–235.

[24] Lu, W. and Liang, Y. (2008). Analysis of competing risks data with missing cause of failure under additive hazards model. *Statist. Sinica* **18** (1), 219–234.

[25] Lu, K. and Tsiatis, A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics* **57** (4), 1191–1197.

[26] McKeague, I.W. and Subramanian, S. (1998). Product-limit estimators and Cox regression with missing censoring information, *Scandinavian Journal of Statistics*, **25**, 589–601.

[27] Miyakawa, M. (1984). Analysis of incomplete data in competing risks model, *IEEE Trans. Reliability*, **33**, 293–296.

[28] Pintilie, M. (2006). *Competing risks: a practical perspective.* Vol. **58**, Wiley.

[29] R Development Core Team (2010). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[30] Rebolledo, R. (1980). Central limit theorems for local martingales, *Z. Wahrsch. verw. Geb.*, **51**, 269–286.

[31] Reiser, B., Flehinger, B.J. and Conn, A.R. (1996). Estimating component-defect probability from masked system success/failure data, *IEEE Trans. Reliability*, **45**, 238–243.

[32] Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed, *J. Amer.*

Statist. Assoc. , **89**, 846–866.

[33] Rubin, D.B. (1976). Inference and missing data , *Biometrika*, **63**, 581–592.

[34] Schabe, H. (1994). Nonparametric estimation of component lifetime based on masked system life test data, *J. Royal Statist. Soc. Ser. B* **56**, 251–259.

[35] Song, X., Sun, L., Mu, X. and Dinse, G.E. (2010). Additive hazards regression with censoring indicators missing at random. *The Canadian J. Statist.*, **38**, 333–351.

[36] Subramanian, S. (2004). Asymptotically efficient estimation of a survival function in the missing censoring indicator model. *J. Nonparametr. Stat.*, **16**, 797–817.

[37] Usher, J.S., Hodgson, T.J. (1988). Maximum likelihood analysis of component reliability using masked system life-test data, *IEEE Trans. Reliability*, **37**, 550–555.

[38] van der Laan, M. and McKeague, I.W. (1998). Efficient Estimation from Right-Censored Data when Failure Indicators are Missing at Random, *Ann. Statist.*, **26**, 164–182.

[39] van der Vaart, A.W. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New-York.

[40] Zhou, X. and Sun, L. (2003). Additive hazards regression with missing censoring information, *Statistica Sinica*, **13**, 1237–1257.