

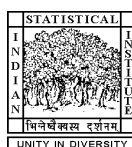
# Regression Analysis of Competing Risks Data with General Missing Pattern in Failure Types

**Technical Report No. ASD/2012/6**

**Dated: 4 May 2012**

Anup Dewanji, Indian Statistical Institute, Kolkata  
P.G. Sankaran, Cochin University of Science and Technology, Cochin  
Debasis Sengupta, Indian Statistical Institute, Kolkata  
and  
Bappa Karmakar, Indian Statistical Institute, Kolkata

Indian Statistical Institute  
Applied Statistics Unit  
Kolkata 700 108



## Regression Analysis of Competing Risks Data with General Missing Pattern in Failure Types

ANUP DEWANJI, P. G. SANKARAN, DEBASIS SENGUPTA AND BAPPA KARMAKAR

*Applied Statistics Unit, Indian Statistical Institute, Kolkata 700108, India*

*Cochin University of Science and Technology, Cochin 682022, India*

*Applied Statistics Unit, Indian Statistical Institute, Kolkata 700108, India*

*Applied Statistics Unit, Indian Statistical Institute, Kolkata 700108, India*

**SUMMARY.** In competing risks data, missing failure types (causes) is a very common phenomenon. In a general missing pattern, if a failure type is not observed, one observes a set of possible types containing the true type along with the failure time. Dewanji and Sengupta (2003) considered nonparametric estimation of the cause-specific hazard rates and suggested a Nelson-Aalen type estimator under such general missing pattern. In this work, we deal with the regression problem, in which the cause-specific hazard rates may depend on some covariates, and consider estimation of the regression coefficients and the cause-specific baseline hazards under the general missing pattern using some semi-parametric models. We consider two different proportional hazards type semi-parametric models for our analysis. Simulation studies from both the models are carried out to investigate the finite sample properties of the estimators. We also consider an example from an animal experiment to illustrate our methodology.

**KEY WORDS:** Competing risks; Cause-specific hazard; Missing failure type; Missing at random; Nelson-Aalen estimator; Semi-parametric model.

## 1 Introduction

In survival studies, the failure (or, death) may be attributed to one of several causes or types, known as competing risks. In such situations, for each individual, we observe a random vector  $(T, J)$ , where  $T$  is possibly censored survival time and  $J$  represents the

cause of death (exactly one of say  $m$  possible causes). However, due to inadequacy in the diagnostic mechanism, often there is uncertainty about the true failure type and so the experimentalists are reluctant to report any specific value of  $J$  for some individuals. This is usually known as the problem of missing failure type in competing risks and has been addressed by many authors. For example, in carcinogenicity studies, besides deaths (failures) without tumor, there are deaths with tumor present due to either the tumor itself or some other causes. Often there is uncertainty in assigning this cause of death even if the presence of tumor can be ascertained (Dinse, 1986; Lagakos and Louis, 1988). In extreme situations, one cannot even ascertain presence or absence of tumor because it is totally cannibalized or autolysed (See Section 6 for details).

Analysis of competing risks data with missing failure types was first considered by Dinse (1982) with the assumption that failure type was either completely available (that is, observed as exactly one of  $m$  possible types) or not available at all (that is, unobserved failure type is any one of the  $m$  possible types). This problem was subsequently studied by different researchers (See Miyakawa, 1984; Racine-poon and Hoel, 1984; Lo, 1991; Mukherjee and Wang, 1993). Goetghebeur and Ryan (1990, 1995) considered the regression problem for two failure types under the assumption that the cause-specific hazards for two failure types are proportional. The method of partial likelihood was employed for estimating the regression parameter. See also Dewanji (1992) and Lu and Tsiatis (2001).

In all the studies mentioned above, missingness meant that no information on failure type was available at all. However, in many contexts, one may be able to narrow down to fewer than  $m$  types to be responsible for failure. In the present work, we consider a general missing pattern in which, for each individual failure, we observe, in addition to survival time, a subset  $g \subseteq \{1, \dots, m\}$  of labels of possible failure types, exactly one of which is the true but unobserved cause of failure (See Section 6 for an example). When  $g$  is a singleton set, then the failure type is exactly observed, and when  $g = \{1, \dots, m\}$ , then the missingness is total. It is usually said that the true failure type is masked in the set  $g$ . Flehinger et al. (1998) considered such general pattern of missing failure types for the purpose of estimating survival due to different types with the strong assumption of proportional hazards due to different types. They also assumed that, for some of the observations with missing failure type, a second stage diagnosis can be performed to pinpoint the type. Craiu and Duchesne (2004) suggested an estimation procedure

using EM algorithm based on piecewise constant cause-specific hazard rates. Under a missing-at-random type assumption and requiring a second stage diagnosis, they developed an EM algorithm to estimate the piecewise constant cause-specific hazard rates and the diagnostic probabilities of the actual cause of failure being  $j$ , given the set  $g$  of observed possible causes. Recently, Dewanji and Sengupta (2003), in addition to suggesting a nonparametric estimator using EM algorithm, developed a Nelson-Aalen type estimator of the cumulative cause-specific hazard rates (and also a smooth estimator of the cause-specific hazard rates), when certain information on the diagnostic probabilities are available from the experimentalists, but the missing pattern could be allowed to be non-ignorable and no second stage diagnosis was required.

In this work, we deal with the regression problem, in which the cause-specific hazard rates may depend on some covariates, and consider estimation of the regression coefficients under some proportional hazards type semi-parametric models, when observation on the failure type exhibits the general missing pattern as discussed before. Recently, Chatterjee et al. (2010) have considered a similar problem in the context of partially observed disease classification data with possibly large number of types. They have suggested a two-stage modeling in which the first stage involves reducing the number of parameters by imposing a natural structure on the underlying disease types and the second stage involves inference through a general extension of the partial likelihood based estimating equation (See Goetghebeur and Ryan, 1995). Apparently, however, they need to make certain assumptions regarding the missing probabilities like most of the work on this issue. Also, Sen et al. (2010) have developed a semiparametric Bayesian approach, where the partial information about the cause of death is incorporated by means of latent variables, and proposed a simulation-based method using Markov Chain Monte Carlo (MCMC) techniques to implement the Bayesian methodology.

We also consider estimation of the cumulative baseline hazards in the spirit of Dewanji and Sengupta (2003). In Section 2, we describe the data and two semi-parametric models to study the effect of covariates. In Section 3, we consider estimation of the regression coefficients, the cumulative baseline cause-specific hazards, the baseline cause-specific hazard rates and the overall survival function for the first model. The same is done for the second model in Section 4. Simulation studies for both the models are carried out in Section 5 to investigate the finite sample properties of the estimators. We illustrate our methodology by means of an example of carcinogenicity study conducted by the British

Industrial Biological Research Association (Peto et al., 1984) in Section 6. Section 7 ends with some concluding remarks.

## 2 Data and Models

For the  $m$  competing causes of failure, the corresponding cause-specific hazard rates, given the covariate value  $Z = z$ , are defined as

$$\lambda_j(t, z) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \Pr [T \in [t, t + \Delta t), J = j | T \geq t, z], \quad (1)$$

for  $j = 1, \dots, m$ , where  $T$  denotes the failure time,  $J$  the failure type and  $Z$  the covariate that may be a vector. The survival function  $S(t, z)$ , for an individual with covariate value  $Z = z$ , can be written in terms of the cause-specific hazard rates as  $S(t, z) = \exp \left[ - \int_0^t \sum_{j=1}^m \lambda_j(u, z) du \right]$ .

Suppose that the data consists of the covariate value  $Z$ , the failure or censoring time  $X = \min(T, C)$ , where  $C$  denotes the censoring time, a censoring indicator  $\delta$  (1 for failure and 0 for censoring), and, if failure occurs, only partial information about the failure type is available in the form of a set  $G \subseteq \{1, \dots, m\}$  representing the possible types of failure. This information on failure type is complete when the observed set  $G = g$  is a singleton set. Let  $(x_i, \delta_i, g_i, z_i)$  be an observation, where the components are the observed values of  $X$ ,  $\delta$ ,  $G$  and  $Z$ , respectively, for the  $i$ th individual, for  $i = 1, \dots, n$ .

Let us write the masking probability  $p_{gj}(t, z)$  as

$$p_{gj}(t, z) = \Pr [G = g | T = t, J = j, \delta = 1, Z = z], \quad (2)$$

which is the conditional probability of observing the set  $g$  as the set of possible failure types, given that there is a failure of type  $j$  at time  $t$  with covariate value  $z$ , for  $j = 1, \dots, m$  and  $g \ni j$ . If  $g$  does not contain  $j$ , this probability is zero so that, for a fixed  $j$ ,  $\sum_g p_{gj}(t, z) = 1$ . Note that the assumption  $p_{gj}(t, z) = p_{g'j}(t, z)$ , for all  $j \neq j' \in g$ , is same as the missing-at-random assumption (Little and Rubin, 1987, p90) in this context. Assuming that the missing mechanism is independent of the censoring mechanism and, for simplicity, the covariate value, the probability  $p_{gj}(t, z)$  equals  $\Pr[G = g | T = t, J = j] = p_{gj}(t)$ , say, which is the conditional probability of observing  $g \ni j$  as the set of possible failure types, given failure of type  $j$  at time  $t$ . Therefore, the missing pattern

is allowed to be non-ignorable. Noting that  $\lambda_j(t, z)dt$  is the conditional probability of instantaneous failure due to type  $j$  at time  $t$ , given survival up to time  $t-$  and  $z$ , it follows that the hazard rate for failure due to cause  $j$  at time  $t$  and with  $g \ni j$  observed as the set of possible causes is  $p_{gj}(t)\lambda_j(t, z)$ . Hence, the hazard rate for failure at time  $t$  with  $g$  observed as the set of possible causes, given covariate value  $z$ , is

$$\lambda_g^*(t, z) = \sum_{j \in g} p_{gj}(t)\lambda_j(t, z) . \quad (3)$$

As expected, the sum of the hazards given by (3), over the set  $\mathcal{G}$  of all non-empty subsets  $g$  of  $\{1, \dots, m\}$ , is the total hazard  $\lambda.(t, z) = \sum_{j=1}^m \lambda_j(t, z)$  since the coefficient of  $\lambda_j(t, z)$ , for a fixed  $j$ , in that sum is  $\sum_{g \ni j} p_{gj}(t) = 1$ . Note that these hazard rates in (3) can also be viewed as another set of cause-specific hazard rates with the different  $g$ 's in  $\mathcal{G}$  representing the different failure types. It is just another decomposition of the total hazard rate  $\lambda.(t, z)$  resulting in a different competing risks problem. However, while information on the original failure types may be masked, the same on these 'modified' failure types is directly available, thereby making estimation of the modified cause-specific hazard rates  $\lambda_g^*(t, z)$ , given by (3), easier.

We consider two different proportional hazards type semi-parametric models to describe the competing risks data. In the first model, referred to as Model 1, we have

$$\lambda_j(t, z) = \lambda_{0j}(t)e^{z\beta} , \quad (4)$$

where  $\lambda_{0j}(t)$  is an unknown and arbitrary function of  $t$  depending on  $j$  representing the baseline cause-specific hazard rate for type  $j$  and  $z\beta$  is a linear combination of the components of  $z$  with  $\beta$  being the vector of coefficient parameters (assumed to be the same for all  $j$ ). The vector of coefficient parameters  $\beta$  measures the effect of the covariate vector (the  $k$ th parameter measuring the effect of the  $k$ th covariate) on the different cause-specific hazard rates and this effect is assumed to be the same for all the failure types. This may be a strong assumption; but we relax this to some extent in the second model.

Alternatively, in the second model, referred to as Model 2, we have

$$\lambda_j(t, z) = \lambda_0(t)e^{\gamma_j + z\beta_j} , \quad (5)$$

with  $\gamma_1 = 0$  for identifiability and where  $\lambda_0(t)$  is an unknown and arbitrary function of  $t$ . The vector of coefficient parameters  $\beta_j$  depends on  $j$ , thereby relaxing the assumption

of same effect of the covariate vector on the different cause-specific hazard rates, as in Model 1. However, the baseline cause-specific hazard rates,  $\lambda_{0j}(t)$ 's, are now assumed to be proportional to each other under this model with  $e^{\gamma_j}$ 's being the proportionality constants. Formally,  $\lambda_{0j}(t) = \lambda_0(t)e^{\gamma_j}$ , for  $j = 2, \dots, m$ , and  $\lambda_{01}(t) = \lambda_0(t)$ . The constants  $\gamma_2, \dots, \gamma_m$  are unknown and treated as parameters along with the  $\beta_j$ 's. Let us write  $\tilde{\gamma} = (\gamma_2, \dots, \gamma_m)^T$ ,  $\tilde{\beta} = (\beta_1, \dots, \beta_m)^T$  and  $\tilde{\theta} = (\tilde{\gamma}, \tilde{\beta})$ .

Note that although Model 2, given by (5), has more number of regression parameters than Model 1, given by (4), Model 1 has  $m$  arbitrary functions (the baseline cause-specific hazard rates) to be estimated while Model 2 has only one. Therefore, clearly, neither model is a special case of the other; both the models are of some interest depending on the situation. Both the models can be independently tested from standard competing risks data without any missingness. However, under the general missing pattern as discussed, none of these can be tested. It may not be possible to analyze the more general model, in which  $\lambda_j(t, z) = \lambda_{0j}(t)e^{z\beta_j}$ , at least by the technique we use in this work.

### 3 Estimation under Model 1

Under Model 1, using (3), the hazard rate for failure at time  $t$  with  $g$  observed as the set of possible failure types, given covariate value  $z$ , is given by

$$\lambda_g^*(t, z) = \lambda_{0g}^*(t)e^{z\beta}, \quad (6)$$

where  $\lambda_{0g}^*(t) = \sum_{j \in g} p_{gj}(t)\lambda_{0j}(t)$ , for all  $g \in \mathcal{G}$  are arbitrary and unknown functions since  $\lambda_{0j}(t)$ 's are. Also, these cause-specific hazard rates for the 'modified' competing risks problem are of the same semi-parametric form as those for the original cause-specific hazard rates in (4). Hence, following partial likelihood is the most appropriate to estimate the regression parameters  $\beta$ , in the absence of any knowledge on the baseline cause-specific hazards  $\lambda_{0g}^*(t)$ , based on 'modified' competing risks data which is available without any missing failure type (Kalbfleisch and Prentice, 1980, Sec. 7.2.3).

### 3.1 Estimation of regression parameters

Let  $t_{(g_1)} < \dots < t_{(g_{k_g})}$  be the  $k_g$  ordered observed failure times (assuming no tie) with missing pattern  $g$  (that is, with  $g$  being the set of possible failure types) and let  $z_{(g_1)}, \dots, z_{(g_{k_g})}$  denote the corresponding covariate values. Then, at each of these  $g$ -type failure times, say  $t_{(gi)}$ , we consider the conditional probability that the individual ( $gi$ ) with covariate value  $z_{(gi)}$  fails at time  $t_{(gi)}$ , given the history up to time  $t_{(gi)}-$  and that one failure with missing pattern  $g$  occurs at time  $t_{(gi)}$ . The method of partial likelihood then gives

$$\begin{aligned} L_{P1}(\beta) &= \prod_{g \in \mathcal{G}} \prod_{i=1}^{k_g} \left[ \frac{\lambda_g^*(t_{(gi)}, z_{(gi)})}{\sum_{l \in R(t_{(gi)})} \lambda_g^*(t_{(gi)}, z_l)} \right] \\ &= \prod_{g \in \mathcal{G}} \prod_{i=1}^{k_g} \left[ \frac{e^{z_{(gi)}\beta}}{\sum_{l \in R(t_{(gi)})} e^{z_l\beta}} \right], \end{aligned} \quad (7)$$

where  $R(t_{(gi)})$  is the risk set at time  $t_{(gi)}-$  and  $z_l$  denotes the covariate value for the  $l$ th individual. Clearly, this partial likelihood (7) can accommodate tied failure times with different missing pattern, but an approximation may be needed to deal with tied failure times with the same missing pattern. This partial likelihood is maximized to get an estimate of  $\beta$ . Let us denote it by  $\hat{\beta}$ . Note that the standard asymptotic likelihood techniques can be applied to this partial likelihood (7) and to the estimate  $\hat{\beta}$  to make inference on  $\beta$ . See Andersen and Gill (1982), Andersen and Borgan (1985) and Andersen et al. (1993, Ch. VII.2). In particular, under some regularity conditions, the asymptotic distribution of  $(\hat{\beta} - \beta)$  is approximately a multivariate normal with mean zero and covariance matrix that may be consistently estimated by  $I^{-1}(\hat{\beta})$ , where  $-I(\beta)$  is the matrix of second order partial derivatives of  $\log L_{P1}(\beta)$ .

### 3.2 Estimation of baseline cumulative cause-specific hazards

The baseline cumulative cause-specific hazards  $\Lambda_{0g}^*(t) = \int_0^t \lambda_{0g}^*(u) du$  for the ‘modified’ competing risks problem can also be estimated as follows. Let us consider the  $n(2^m - 1)$ -dimensional counting process  $\{N_{gi}(t), g \in \mathcal{G}, i = 1, \dots, n\}$ , where  $N_{gi}(t)$  counts the number of failures with missing pattern  $g$  up to time  $t$  for individual  $i$ . Consider the multiplicative intensity model with covariates, in which the corresponding intensity



process is given by

$$\alpha_{gi}(t) = \lambda_g^*(t, z_i)Y_i(t) = \lambda_{0g}^*(t)e^{z_i\hat{\beta}}Y_i(t),$$

where  $Y_i(t) = 1$  if the  $i$ th individual is at risk at time  $t-$  and  $\lambda_{0g}^*(t)$  is as in (6). We have, for each non-empty subset  $g$  of  $\{1, \dots, m\}$  and for each  $i$ ,

$$dN_{gi}(t) = \alpha_{gi}(t)dt + dM_{gi}(t),$$

where  $M_{gi}(t)$ 's are local square integrable martingales. Then, following Andersen and Borgan (1985), the generalized Nelson-Aalen estimate for  $\Lambda_{0g}^*(t) = \int_0^t \lambda_{0g}^*(u)du$ , for  $g \in \mathcal{G}$ , is given by

$$\hat{\Lambda}_{0g}^*(t) = \int_0^t \frac{J(u)}{nS^{(0)}(\hat{\beta}, u)} dN_g(u), \quad (8)$$

where  $N_g(t) = \sum_{i=1}^n N_{gi}(t)$ ,  $Y(t) = \sum_{i=1}^n Y_i(t)$ ,  $J(u) = I\{Y(u) > 0\}$  and  $S^{(0)}(\hat{\beta}, u) = n^{-1} \sum_{i=1}^n Y_i(u)e^{z_i\hat{\beta}}$ . This integral formula for the estimate reduces to a finite sum (Lawless, 2003, p449)

$$\hat{\Lambda}_{0g}^*(t) = \sum_{i:t_{(gi)} \leq t} \frac{1}{\sum_{l=1}^n Y_l(t_{(gi)})e^{z_l\hat{\beta}}}.$$

Under the same set of regularity conditions, as required for the asymptotic normality of  $\hat{\beta}$ , the process  $(\hat{\Lambda}_{0g}^*(t) - \Lambda_{0g}^*(t), g \in \mathcal{G})$  converges weakly to a  $(2^m - 1)$ -variate mean zero Gaussian process (Andersen et al., 1993, Ch VII.2.2). In particular, for fixed  $g, g' \in \mathcal{G}$  and time  $t$ , the covariance of  $(\hat{\Lambda}_{0g}^*(t) - \Lambda_{0g}^*(t), \hat{\Lambda}_{0g'}^*(t) - \Lambda_{0g'}^*(t))$  can be consistently estimated by  $\hat{\sigma}_{gg'}(t) =$

$$\delta_{gg'} \int_0^t S^{(0)}(\hat{\beta}, u)^{-2} dN_g(u) + \left( \int_0^t X(\hat{\beta}, u) dN_g(u) \right)^T I^{-1}(\hat{\beta}) \left( \int_0^t X(\hat{\beta}, u) dN_{g'}(u) \right), \quad (9)$$

where  $\delta_{gg'} = 1$ , if  $g$  and  $g'$  are the same set, and 0 otherwise, and  $X(\hat{\beta}, u) = n^{-1} S^{(1)}(\hat{\beta}, u) \times S^{(0)}(\hat{\beta}, u)^{-2}$  with  $S^{(1)}(\hat{\beta}, u) = n^{-1} \sum_{i=1}^n z_i Y_i(u) e^{z_i\hat{\beta}}$  being a vector of the same size as that of the covariate  $Z$ . Note that this expression (9) for the estimated covariance also reduces to a finite sum.

Note that the masking probabilities  $p_{gj}(t)$ 's are usually not known and need to be estimated. There have been some works concerning estimation of these masking probabilities usually requiring either additional modeling assumptions or secondary data. See Basu (2009) for a review on this, which also finds that the performance of model-based estimates is less than desirable. Nevertheless, in order to be able to estimate these probabilities in practice, as in Dewanji and Sengupta (2003), we make the simplifying

assumption that  $p_{gj}(t)$ 's are independent of time  $t$ , though it may depend on  $g$  and  $j$ . We will write  $p_{gj}(t)$  as  $p_{gj}$  in the subsequent discussion and denote the  $(2^m - 1) \times m$  matrix of  $p_{gj}$ 's by  $P$ . Then, writing  $\tilde{\Lambda}_0^*(t)$  as the  $(2^m - 1) \times 1$  vector of  $\Lambda_{0g}^*(t)$ 's and  $\tilde{\Lambda}_0(t)$  as the  $m \times 1$  vector of the original cumulative baseline cause-specific hazards  $\Lambda_{0j}(t)$ 's, where  $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(u) du$ , we have, from (3),

$$\tilde{\Lambda}_0^*(t) = P \tilde{\Lambda}_0(t). \quad (10)$$

Then, following Dewanji and Sengupta (2003), we have a consistent estimate of  $\tilde{\Lambda}_0(t)$  as

$$\hat{\tilde{\Lambda}}_0(t) = \left( \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{P} \right)^{-1} \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{\tilde{\Lambda}}_0^*(t), \quad (11)$$

where  $\hat{\Sigma}(t)$  is the  $(2^m - 1) \times (2^m - 1)$  matrix of  $\hat{\sigma}_{gg'}(t)$ 's,  $\hat{P}$  denotes a consistent estimate of  $P$  and  $\hat{\tilde{\Lambda}}_0^*(t)$  is the  $(2^m - 1) \times 1$  vector of  $\hat{\Lambda}_{0g}^*(t)$ 's.

Note that  $\hat{\tilde{\Lambda}}_0(t) - \tilde{\Lambda}_0(t)$

$$\begin{aligned} &= \left( \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{P} \right)^{-1} \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{\tilde{\Lambda}}_0^*(t) - \left( \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{P} \right)^{-1} \left( \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{P} \right) \tilde{\Lambda}_0(t) \\ &= \hat{A}(t) \left( \hat{\tilde{\Lambda}}_0^*(t) - \tilde{\Lambda}_0^*(t) \right) + \hat{A}(t) \left( P - \hat{P} \right) \tilde{\Lambda}_0(t), \end{aligned}$$

where  $\hat{A}(t) = \left( \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{P} \right)^{-1} \hat{P}^T \hat{\Sigma}^{-1}(t)$  is a consistent estimate of  $\left( P^T \Sigma^{-1}(t) P \right)^{-1} P^T \Sigma^{-1}(t)$  with  $\Sigma(t)$  being the asymptotic covariance matrix of  $\hat{\tilde{\Lambda}}_0^*(t) - \tilde{\Lambda}_0^*(t)$ . Hence, using the weak convergence result of  $\hat{\tilde{\Lambda}}_0^*(t) - \tilde{\Lambda}_0^*(t)$  and the fact that  $\hat{P}$  is a consistent estimate of  $P$ , one can establish weak convergence of  $\hat{\tilde{\Lambda}}_0(t) - \tilde{\Lambda}_0(t)$  to a  $m$ -variate mean zero Gaussian process with the covariance matrix at time  $t$  given by  $\left( P^T \Sigma^{-1}(t) P \right)^{-1}$ , which can be consistently estimated by  $\left( \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{P} \right)^{-1}$ . This result on weak convergence to a Gaussian process similarly holds for the estimate of Dewanji and Sengupta (2003) and can be useful for nonparametric one- and  $k$ -sample tests for the cumulative cause-specific hazards in the spirit of Andersen and Borgan (1985, Section 5) and Andersen et al. (1993, Ch. V).

Note, however, that this estimate cannot be guaranteed to be non-decreasing, although it is expected to be so for large sample because of its consistency. In practice, one can use ‘‘pooling-the-adjacent-violators’’ algorithm to achieve monotonicity. If some of the  $\{N_g(t)\}$ 's are not observed to have any jump during the study, the corresponding  $\hat{\tilde{\Lambda}}_0^*(t)$ 's, and the associated entries of  $\hat{\Sigma}(t)$ , turn out to be zero; the corresponding rows

of  $P$  are also estimated, as given in Dewanji and Sengupta (2003), to be zero. The same estimation procedure goes through with the observed  $\{N_g(t)\}$ 's as long as the resulting  $\widehat{P}$  is of full column-rank. Even if the rank of  $\widehat{P}$  is less than  $m$ , some components of  $\widehat{\Lambda}_{\sim_0}(t)$  may be estimable.

### 3.3 Estimation of baseline cause-specific hazards

Smooth estimates of the baseline cause-specific hazards  $\lambda_{0g}^*(t)$ , for all  $g$ , may be obtained via kernel smoothing, as suggested by Andersen and Borgan (1985) and Andersen et al (1993, p507) and is given by

$$\widehat{\lambda}_{0g}^*(t) = \frac{1}{h} \int_0^1 K\left(\frac{t-u}{h}\right) d\widehat{\Lambda}_{0g}^*(u), \quad (12)$$

where  $\widehat{\Lambda}_{0g}^*(t)$  is as defined in (8). In the above expression the whole observation range is transformed into  $[0,1]$ , the kernel function  $K(\cdot)$  is a bounded function with support  $[-1,1]$  and having integral 1, and  $h$  is a positive constant denoting the window length. Following Andersen and Borgan (1985) and Andersen et al (1993, p507), when  $n \rightarrow \infty$  and window length  $h \rightarrow 0$  in such a way that  $nh \rightarrow \infty$ , the smooth estimates  $\widehat{\lambda}_{0g}^*(t)$ , for  $g \in \mathcal{G}$ , given by (12), are asymptotically independent and normally distributed with mean  $\lambda_{0g}^*(t)$  and variance that can be consistently estimated by

$$v_g^2(t) = \frac{1}{h^2} \int_0^1 K^2\left(\frac{t-u}{h}\right) \left(nS^{(0)}(\widehat{\beta}, u)\right)^{-2} dN_g(u),$$

so that the variance matrix of  $\widehat{\lambda}_{\sim_0}^*(t)$ , the vector of  $\widehat{\lambda}_{0g}^*(t)$ 's, is estimated by the diagonal matrix  $V(t) = \text{diag}(v_g^2(t))$ .

From (3) and (10), we have the relation  $\widehat{\lambda}_{\sim_0}^*(t) = P \widehat{\lambda}_{\sim_0}(t)$ , where  $\widehat{\lambda}_{\sim_0}(t)$  is the vector of original baseline cause-specific hazards. As in Dewanji and Sengupta (2003), an estimate of  $\widehat{\lambda}_{\sim_0}(t)$ , as in (11), is given by

$$\widehat{\lambda}_{\sim_0}(t) = \left(\widehat{P}^T V^{-1}(t) \widehat{P}\right)^{-1} \widehat{P}^T V^{-1}(t) \widehat{\lambda}_{\sim_0}^*(t). \quad (13)$$

The variance matrix of  $\widehat{\lambda}_{\sim_0}(t)$  is consistently estimated by  $\left(\widehat{P}^T V^{-1}(t) \widehat{P}\right)^{-1}$ . Asymptotic normality of  $\widehat{\lambda}_{\sim_0}(t)$  follows from that of  $\widehat{\lambda}_{\sim_0}^*(t)$  and (13) using the same argument as that used for weak convergence of  $\widehat{\Lambda}_{\sim_0}(t)$ , for fixed  $t$ .

### 3.4 Estimation of survival function

One may often be interested in estimating the survival function  $S(t, z_0)$  for individuals with a fixed covariate  $z_0$ . Note that the overall survival function  $S(t, z_0)$  is a function of the regression parameter  $\beta$  and the baseline cumulative hazard functions  $\Lambda_{0j}(t)$ 's, as given by  $S(t, z_0) = \exp[-e^{z_0\beta} \sum_{j=1}^m \Lambda_{0j}(t)]$ . Using (3) and (10), this can also be viewed as a function of the regression parameter  $\beta$  and the baseline cumulative hazard functions  $\Lambda_{0g}^*(t)$ 's for the modified competing risks problem. Noting that  $\sum_{j=1}^m \Lambda_{0j}(t) = \sum_{g \in \mathcal{G}} \Lambda_{0g}^*(t)$ , we have  $S(t, z_0) = \exp[-e^{z_0\beta} \sum_{g \in \mathcal{G}} \Lambda_{0g}^*(t)]$ . A natural estimate of  $S(t, z_0)$  is, therefore, given by

$$\hat{S}(t, z_0) = \exp[-e^{z_0\hat{\beta}} \sum_{g \in \mathcal{G}} \hat{\Lambda}_{0g}^*(t)], \quad (14)$$

where  $\hat{\beta}$  and  $\hat{\Lambda}_{0g}^*(t)$  are as given in subsections 3.1 and 3.2, respectively. Note that we have asymptotic normality for  $\hat{\beta}$  with estimated covariance matrix  $I^{-1}(\hat{\beta})$ . Also, for fixed  $t$ ,  $\{\hat{\Lambda}_{0g}^*(t), g \in \mathcal{G}\} = \underset{\sim}{\Lambda}_0^*(t)$  asymptotically follows a multivariate normal distribution with estimated covariance matrix  $\Sigma(t)$ . The asymptotic normality for the joint distribution of  $(\hat{\beta}, \{\hat{\Lambda}_{0g}^*(t), g \in \mathcal{G}\})$  can be obtained by using Theorem VII.2.3 of Andersen et al. (1993, p503-504). A straightforward application of delta method then establishes asymptotic normality of  $\hat{S}(t, z_0)$  with mean  $S(t, z_0)$  and variance that can be estimated as follows.

From Andersen et al. (1993, p505), the asymptotic covariance between  $\hat{\beta}$  and  $\hat{\Lambda}_{0g}^*(t)$ , for a fixed  $g$ , is consistently estimated by

$$c_g^*(t, \hat{\beta}) = -n^{-1} I^{-1}(\hat{\beta}) \int_0^t S^{(1)}(\hat{\beta}, u) S^{(0)}(\hat{\beta}, u)^{-2} dN_g(u) = -I^{-1}(\hat{\beta}) \int_0^t X(\hat{\beta}, u) dN_g(u).$$

In order to apply the delta method, we first note that the estimated asymptotic covariance matrix of  $(\hat{\beta}, \{\hat{\Lambda}_{0g}^*(t), g \in \mathcal{G}\})$  is given by

$$C^*(t, \hat{\beta}) = \begin{bmatrix} I^{-1}(\hat{\beta}) & \left( \underset{\sim}{c}^*(t, \hat{\beta}) \right)^T \\ \underset{\sim}{c}^*(t, \hat{\beta}) & \hat{\Sigma}(t) \end{bmatrix},$$

where  $\underset{\sim}{c}^*(t, \hat{\beta})$  is the  $(2^m - 1) \times 1$  vector of the covariance estimates  $c_g^*(t, \hat{\beta})$ 's, for  $g \in \mathcal{G}$ , as given above. The vector of first derivatives of  $S(t, z_0)$  with respect to  $(\beta, \underset{\sim}{\Lambda}_0^*(t))$  is given by

$$\left[ d^* \left( \beta, \underset{\sim}{\Lambda}_0^*(t) \right) \right]^T = S(t, z_0) \left[ z_0 \log[S(t, z_0)], -e^{z_0\beta}, \dots, -e^{z_0\beta} \right],$$

where the first component is the derivative with respect to  $\beta$  and the last  $(2^m - 1)$  components, each being equal to  $-e^{z_0\beta}S(t, z_0)$ , are the derivatives with respect to  $\Lambda_{0g}^*(t)$ , for  $g \in \mathcal{G}$ , the components of  $\tilde{\Lambda}_0^*(t)$ . The estimated asymptotic variance of  $\hat{S}(t, z_0)$  is, therefore, given by

$$\widehat{V}(\hat{S}(t, z_0)) = \left( d^* \left( \hat{\beta}, \hat{\tilde{\Lambda}}_0^*(t) \right) \right)^T C^*(t, \hat{\beta}) \left( d^* \left( \hat{\beta}, \hat{\tilde{\Lambda}}_0^*(t) \right) \right). \quad (15)$$

## 4 Estimation under Model 2

A similar method can be developed under Model 2. Using (3), the cause-specific hazard rate for failure at time  $t$  with  $g$  observed as the set of possible failure types, given covariate value  $z$ , is given by

$$\lambda_g^*(t, z) = \lambda_0(t) f_g(z, t, \theta), \quad (16)$$

where  $f_g(z, t, \theta) = f_g(z, t, \tilde{\gamma}, \tilde{\beta}) = \sum_{j \in g} p_{gj}(t) e^{\gamma_j + z\beta_j}$ , for all  $g \in \mathcal{G}$ . These have the similar semi-parametric form as those for the original cause-specific hazard rates in (5), except that the parametric component  $f_g(z, t, \theta)$ , for different  $g$ 's, are not of the simple exponential form. In particular, when the masking probabilities ( $p_{gj}(t)$ 's) are independent of time, the parametric component  $f_g(z, t, \theta)$  is also independent of time and is written as  $f_g(z, \theta)$ . Nevertheless, from Kalbfleisch and Prentice (1980, Sec. 7.2.3), the partial likelihood given in (17) is the most appropriate to estimate the vector of regression parameters  $\theta$ , in the absence of any knowledge on the baseline cause-specific hazards  $\lambda_0(t)$ , based on the 'modified' competing risks data which is available without any missingness.

### 4.1 Estimation of regression parameters

Let  $t_{(1)} < \dots < t_{(k)}$  be the  $k$  ordered observed failure times with covariates  $z_{(1)}, \dots, z_{(k)}$  and missing patterns  $g_{(1)}, \dots, g_{(k)}$ , respectively. Then, at each of these failure times, say  $t_{(i)}$ , we consider the conditional probability that the individual ( $i$ ) with covariate value  $z_{(i)}$  fails at time  $t_{(i)}$  with missing pattern  $g_{(i)}$ , given the history up to time  $t_{(i)}$ — and that

one failure occurs at time  $t_{(i)}$ . The method of partial likelihood then gives

$$\begin{aligned} L_{P2}(\hat{\theta}) &= \prod_{i=1}^k \left[ \frac{\lambda_{g(i)}^*(t_{(i)}, z_{(i)})}{\sum_{l \in R(t_{(i)})} \sum_{g \in \mathcal{G}} \lambda_g^*(t_{(i)}, z_l)} \right] \\ &= \prod_{i=1}^k \left[ \frac{f_{g(i)}(z_{(i)}, \hat{\theta})}{\sum_{l \in R(t_{(i)})} \sum_{g \in \mathcal{G}} f_g(z_l, \hat{\theta})} \right], \end{aligned} \quad (17)$$

where  $R(t_{(i)})$  is the risk set at time  $t_{(i)}$ . See Dewanji (1992) for a special case of this partial likelihood. An approximation may be needed to deal with tied failure times. This partial likelihood (17) is maximized to get an estimate of  $\theta$ , denoted by  $\hat{\theta}$ . Note that the model (16) cannot be written as the underlying multiplicative hazard competing risks model of Andersen et al. (1993, Ch. VII.2) for the asymptotic results therein to be readily available, as for the model in (6). However, similar asymptotic likelihood techniques can still be applied to this partial likelihood (17) and to the estimate  $\hat{\theta}$  for making inference on  $\theta$ . The proofs of the asymptotic results follow the similar steps, as those in Andersen and Gill (1982) and Andersen et al. (1993, Ch VII.2), with little modification, as worked out by Prentice and Self (1983) in the context of ordinary survival data with general relative risk form. In particular, under some regularity conditions as mentioned in Prentice and Self (1983), the asymptotic distribution of  $(\hat{\theta} - \theta^0)$ , where  $\theta^0$  denotes the true value of  $\theta$ , is approximately multivariate normal with mean zero and covariance matrix that may be consistently estimated by  $I^{-1}(\hat{\theta})$ , where  $-I(\theta)$  is the matrix of second order partial derivatives of  $\log L_{P2}(\theta)$ . A sketch of the proof is given in the Appendix.

## 4.2 Estimation of baseline cumulative cause-specific hazards

Estimation of  $\Lambda_0(t) = \int_0^t \lambda_0(u) du$  follows the similar argument as those in Andersen et al. (1993, p482) and, given the maximum partial likelihood estimate  $\hat{\theta}$ , a natural estimator for  $\Lambda_0(t)$  is

$$\hat{\Lambda}_0(t) = \int_0^t \frac{J(u)}{n \sum_{g \in \mathcal{G}} S_g^{(0)}(\hat{\theta}, u)} dN(u), \quad (18)$$

where  $N(t) = \sum_i^n \sum_{g \in \mathcal{G}} N_{gi}(t)$  and  $S_g^{(0)}(\theta, u) = n^{-1} \sum_{i=1}^n f_g(z_i, \theta) Y_i(u)$ . Under some regularity conditions, as required for asymptotic normality of  $\hat{\theta}$ , the process  $\sqrt{n} (\hat{\Lambda}_0(t) - \Lambda_0(t))$

converges weakly to a mean zero Gaussian process with variance function  $\sigma^2(t)$ , as given by (22), in the Appendix along with a consistent estimate  $\hat{\sigma}^2(t)$ , given in (23). The proof is similar to that in Andersen et al. (1993, p503-505) with little modification and is briefly sketched in the Appendix. The asymptotic covariance between  $\sqrt{n}(\hat{\theta} - \theta^0)$  and  $\sqrt{n}(\hat{\Lambda}_0(t) - \Lambda_0(t))$  is also obtained, as given by (24) in the Appendix, which can be consistently estimated by

$$b(\hat{\theta}, t) = -I^{-1}(\hat{\theta}) \int_0^t \left( \sum_g S_g^{(1)}(\hat{\theta}, u) \right) \left( \sum_g S_g^{(0)}(\hat{\theta}, u) \right)^{-2} dN(u).$$

### 4.3 Estimation of baseline cause-specific hazards

As in Section 3.3 for Model 1, a smooth estimate of baseline cause-specific hazard  $\lambda_0(t)$  can be obtained by kernel smoothing as given by

$$\hat{\lambda}_0(t) = \frac{1}{h} \int_0^1 K\left(\frac{t-u}{h}\right) d\hat{\Lambda}_0(u),$$

where  $\hat{\Lambda}_0(t)$  is as given in (18). Following Andersen and Borgan (1985) and Andersen et al. (1993, p507), when  $n \rightarrow \infty$  and  $h \rightarrow 0$  in such a way that  $nh \rightarrow \infty$ , then  $\hat{\lambda}_0(t)$  can be proved to be asymptotically normally distributed with mean  $\lambda_0(t)$  and variance that can be consistently estimated by

$$\frac{1}{h^2} \int_0^1 K^2\left(\frac{t-u}{h}\right) \left( n \sum_g S_g^{(0)}(\hat{\theta}, u) \right)^{-2} dN(u).$$

### 4.4 Estimation of survival function

Note that the overall survival function with a fixed covariate  $z_0$  is given by  $S(t, z_0) = \exp[-\Lambda_0(t) \sum_{j=1}^m e^{\gamma_j + z_0 \beta_j}]$ , which can be seen as a function of  $\Lambda_0(t)$  and  $\theta$ . A natural estimate of  $S(t, z_0)$  is, therefore, given by

$$\hat{S}(t, z_0) = \exp[-\hat{\Lambda}_0(t) \sum_{j=1}^m e^{\hat{\gamma}_j + z_0 \hat{\beta}_j}].$$

As in Section 3.4, using the asymptotic independence between  $\hat{\theta}$  and  $W(t)$ , given by (21) in the Appendix, one can establish asymptotic normality for the joint distribution

of  $\{\hat{\theta}, \hat{\Lambda}_0(t)\}$ . Then, an application of delta method establishes asymptotic normality of  $\hat{S}(t, z_0)$  with mean  $S(t, z_0)$  and variance that can be estimated as follows.

From the end of Section 4.2, the asymptotic covariance between  $\hat{\theta}$  and  $\hat{\Lambda}_0(t)$  is consistently estimated by  $n^{-1}b(\hat{\theta}, t)$ . Therefore, the estimated asymptotic covariance matrix of  $\{\hat{\theta}, \hat{\Lambda}_0(t)\}$  is given by

$$B^*(t, \hat{\theta}) = \begin{bmatrix} I^{-1}(\hat{\theta}) & n^{-1}b(\hat{\theta}, t) \\ n^{-1}b(\hat{\theta}, t) & n^{-1}\hat{\sigma}^2(t) \end{bmatrix},$$

where  $\hat{\sigma}^2(t)$  is as mentioned in Section 4.2. Writing  $e(\theta, z_0) = (e^{\gamma_2+z_0\beta_2}, \dots, e^{\gamma_m+z_0\beta_m})^T$ , the vector of first derivatives of  $S(t, z_0)$  with respect to  $\{\theta, \Lambda_0(t)\}$  is given by

$$\left[ d\left(\hat{\theta}, \hat{\Lambda}_0(t)\right) \right]^T = S(t, z_0) \left[ -\Lambda_0(t)e^T(\theta, z_0), -\Lambda_0(t)z_0 \otimes \left( e^{z_0\beta_1}, e^T(\theta, z_0) \right), -\sum_{j=1}^m e^{\gamma_j+z_0\beta_j} \right],$$

where the first component is the derivative with respect to  $\gamma = (\gamma_2, \dots, \gamma_m)^T$ , the second with respect to  $\beta = (\beta_1, \dots, \beta_m)^T$ , the last is with respect to  $\Lambda_0(t)$  and  $a \otimes b$ , for a  $p \times 1$  vector  $a$  and a  $q \times 1$  vector  $b$ , is the  $pq \times 1$  vector given by  $(a_1b_1, \dots, a_pb_1, \dots, a_1b_q, \dots, a_pb_q)^T$ . The estimated asymptotic variance of  $\hat{S}(t, z_0)$  is, therefore,

$$\left[ d\left(\hat{\theta}, \hat{\Lambda}_0(t)\right) \right]^T B^*(t, \hat{\theta}) \left[ d\left(\hat{\theta}, \hat{\Lambda}_0(t)\right) \right].$$

## 5 Simulation Studies

In order to investigate the finite sample properties of the estimates obtained in Sections 3 and 4, we carry out some simulation studies as described in the following two subsections for Model 1 and 2, respectively. For this purpose, we consider  $m = 3$  and the sample size  $n = 50, 100, 200$  and  $500$ . As in the previous work (Dewanji and Sengupta, 2003), given the true cause  $j$ , the probability of missing failure type is  $1 - p_{\{j\}j}$ , which is taken as constant  $\alpha$ , for all  $j$ , with values  $0.1, 0.3$  and  $0.5$ , representing different extent of missingness. Also, for  $g \ni j$  and  $g \neq \{j\}$ ,  $p_{gj}$  is taken as  $\alpha/3$ , since there are three such



$g$ 's. The true  $P$  matrix is, therefore, given by,

$$P^T = \begin{bmatrix} 1 - \alpha & 0 & 0 & \alpha/3 & \alpha/3 & 0 & \alpha/3 \\ 0 & 1 - \alpha & 0 & \alpha/3 & 0 & \alpha/3 & \alpha/3 \\ 0 & 0 & 1 - \alpha & 0 & \alpha/3 & \alpha/3 & \alpha/3 \end{bmatrix}$$

with the different  $g$ 's appearing in the order  $\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$ . For the sake of simplicity, we consider the case of a single covariate  $z$  generated from a  $\mathcal{U}[0, 10]$  distribution.

For each simulation, once the data is generated with sample size  $n$ , the regression parameter(s) and the cumulative baseline cause-specific hazards are estimated using the methods of Sections 3 and 4, respectively. While estimating, we use the true  $P$  matrix as given above and also an estimated  $P$  matrix using the method of Dewanji and Sengupta (2003) with the  $q$  matrix as given therein. The reported mean and standard error of the estimates are based on 1000 simulations. The standard error of the estimates are computed in two ways: (1) taking the mean of the individual standard errors, obtained by using the results in Sections 3 and 4, and (2) based on the estimates from the 1000 simulations. Since they give similar results, specially for large  $n$ , only those obtained by the latter method are reported (in parentheses). The Q-Q plots of the different parameter estimates over the 1000 simulations (not presented here) also show evidence in favour of asymptotic normality with the estimates of the cumulative baseline cause-specific hazards showing slower convergence, as expected.

## 5.1 Model 1

For the simulation model (see (4)), we consider constant baseline cause-specific hazards (that is,  $\lambda_{0j}(t) = \lambda_j$ ) taking values 0.1, 0.2 and 0.3, for  $j = 1, 2, 3$ , respectively. We also take the value of  $\beta$ , the regression coefficient, to be 0.01. For each simulation, after generating a value of  $z$ , the failure time  $T$  is generated from an exponential distribution with hazard rate  $(\sum_{j=1}^3 \lambda_j) \exp(z\beta) = 0.6 \exp(0.01z)$ . The true cause  $J$  is generated with probability  $\lambda_j / (\sum_{j'=1}^3 \lambda_{j'})$ , for  $j = 1, 2, 3$ . Then, the observed set of possible causes  $g$  is generated using the probabilities  $p_{gj}$ ,  $g \ni j$ . Therefore, the generated observed data for an individual consists of  $\{T, g, z\}$ . Generating such data for  $n$  individuals results in one simulated data set.

Table 1. Simulation based estimates of the regression coefficient  $\beta$  and the baseline cumulative cause-specific hazards  $\Lambda_{0j}(t)$  at 60th percentile ( $t_{60}$ ) of the true life distribution with mean covariate value 5 for Model 1. The true value of  $\beta$  is 0.01 and those of the  $\Lambda_{0j}(t_{60})$ ,  $j = 1, 2, 3$ , are 0.145, 0.291 and 0.436, respectively.

$n$	$\alpha$	Using true $P$			Using estimated $P$				
		$\hat{\beta}$	$\hat{\Lambda}_{0j}(t_{60})$			$\hat{\beta}$	$\hat{\Lambda}_{0j}(t_{60})$		
50	.1	.012	.035, .069, .192	.007	.075, .156, .229	(.054)	(.077, .145, .200)	(.055)	(.089, .172, .244)
	.3	.012	.069, .148, .247	.010	.090, .194, .298	(.054)	(.083, .140, .188)	(.055)	(.082, .156, .204)
	.5	.014	.094, .188, .283	.007	.113, .233, .363	(.055)	(.087, .138, .167)	(.052)	(.084, .139, .187)
100	.1	.011	.073, .147, .299	.011	.081, .170, .257	(.038)	(.081, .156, .175)	(.034)	(.077, .145, .206)
	.3	.012	.106, .214, .338	.010	.125, .258, .385	(.036)	(.061, .100, .139)	(.037)	(.061, .100, .130)
	.5	.009	.117, .240, .363	.011	.133, .268, .402	(.036)	(.062, .093, .118)	(.036)	(.056, .086, .117)
200	.1	.009	.116, .229, .384	.010	.111, .229, .355	(.025)	(.064, .117, .120)	(.024)	(.062, .110, .150)
	.3	.011	.128, .260, .393	.011	.137, .276, .416	(.025)	(.042, .070, .087)	(.026)	(.040, .063, .085)
	.5	.008	.131, .269, .403	.009	.140, .279, .422	(.026)	(.044, .067, .088)	(.025)	(.039, .062, .079)
500	.1	.010	.140, .278, .427	.009	.139, .283, .427	(.016)	(.029, .047, .058)	(.016)	(.027, .049, .058)
	.3	.010	.139, .280, .419	.010	.143, .286, .428	(.016)	(.026, .041, .054)	(.015)	(.026, .036, .048)
	.5	.010	.139, .279, .424	.010	.141, .285, .429	(.016)	(.029, .043, .056)	(.015)	(.026, .039, .050)

Table 1 presents the results (means and standard errors of the estimates) of the simulation study for Model 1. Although the baseline cumulative cause-specific hazards

$\Lambda_{0j}(t)$ 's are estimated at several time points (different quantiles of the true life distribution with mean covariate value 5), only the estimates at the 60th percentile ( $t_{60}$ ) are presented for space consideration. The results at other time points are qualitatively similar. The standard errors at higher quantiles are observed to be larger, as expected.

The estimate of the regression parameter  $\beta$  is close to the true value even for small sample sizes and the corresponding standard error decreases with the sample size  $n$ . This estimate does not seem to depend much on the missing proportion  $\alpha$  or the choice of  $P$  matrix (true or estimated). The estimates of the baseline cumulative cause-specific hazard functions also seem to have smaller error for larger  $n$ , but the reduction is slower, as expected. For large sample size ( $n = 500$ ), the performance of these estimates does not depend on  $\alpha$  or the choice of  $P$  matrix. For small sample sizes, scarcity of data makes these estimates poor as the available information gets divided into estimation of three baseline cumulative cause-specific hazards.

## 5.2 Model 2

For the simulation model (see (5)), we consider constant baseline hazard (that is,  $\lambda_0(t) = \lambda_0$ ) taking value 0.1. Note that  $\gamma_1 = 0$ , and  $\gamma_2$  and  $\gamma_3$  are so chosen to have the baseline cause-specific hazards as 0.1, 0.2 and 0.3, respectively, as for Model 1 in the previous subsection. This means that the values of  $\gamma_2$  and  $\gamma_3$  are taken as  $\log 2$  and  $\log 3$ , respectively. The three  $\beta$  parameters, the regression coefficients, are taken as 0.01, 0.02 and 0.03, respectively. For each simulation, after generating a value of  $z$ , the failure time  $T$  is generated from an exponential distribution with hazard rate  $\lambda_0 \sum_{j=1}^3 \exp(\gamma_j + z\beta_j) = 0.1[\exp(0.01z) + 2 \exp(0.02z) + 3 \exp(0.03z)]$ . The true cause  $J$  is generated with probability  $\exp(\gamma_j + z\beta_j) / [\sum_{j'=1}^3 \exp(\gamma_{j'} + z\beta_{j'})]$ , for  $j = 1, 2, 3$ . Then, the observed set of possible causes  $g$  is generated using the probabilities  $p_{gj}$ ,  $g \ni j$ . Therefore, as before, the generated observed data for an individual consists of  $\{T, g, z\}$ . Generating such data for  $n$  individuals results in one simulated data set.

Table 2 presents the results (means and standard errors of the estimates) of the simulation study for Model 2. Although the baseline cumulative cause-specific hazards  $\Lambda_0(t)$ 's are estimated at several time points (different quantiles of the true life distribution

Table 2. Simulation based estimates of the model parameters  $(\gamma_2, \gamma_3, \beta_1, \beta_2, \beta_3)$  and the baseline cumulative hazard  $\Lambda_0(t)$  at 60th percentile ( $t_{60}$ ) of the true life distribution with mean covariate value 5 for Model 2. The true value of the parameter vector is  $(\log 2 = 0.693, \log 3 = 1.099, 0.01, 0.02, 0.03)$  and that of  $\Lambda_0(t_{60})$  is 0.136, respectively.

$n$	$\alpha$	Using true $P$			Using estimated $P$			
		$\hat{\gamma}_2, \hat{\gamma}_3$	$\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$	$\hat{\Lambda}_0(t_{60})$	$\hat{\gamma}_2, \hat{\gamma}_3$	$\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$	$\hat{\Lambda}_0(t_{60})$	
50	.1	.823, 1.244	0.010, 0.020, 0.028	.168	.749, 1.201	.004, .023, .028	.166	
		(1.239, 1.214)	(.186, .103, .075)	(.075)	(1.116, 1.053)	(.173, .106, .079)	(.137)	
	.3	.894, 1.305	-.022, .015, .031	.166	.875, 1.312	.012, .024, .032	.184	
		(1.475, 1.394)	(.863, .112, .084)	(.161)	(5.266, 5.239)	(.648, .114, .080)	(.185)	
	.5	1.167, 1.615	.003, .021, .030	.215	.775, 1.241	-.030, .028, .029	.228	
		(6.383, 6.361)	(.786, .131, .092)	(.356)	(2.029, 1.961)	(.714, .119, .089)	(.478)	
100	.1	.714, 1.149	.009, .023, .031	.146	.685, 1.087	.004, .020, .031	.153	
		(.679, .642)	(.096, .067, .051)	(.035)	(.653, .634)	(.102, .065, .052)	(.084)	
	.3	.729, 1.176	.002, .023, .030	.152	.740, 1.118	.010, .016, .031	.152	
		(.827, .784)	(.127, .070, .055)	(.044)	(.734, .693)	(.115, .070, .055)	(.089)	
	.5	.729, 1.160	.007, .020, .030	.159	.712, 1.128	.011, .022, .031	.157	
		(.939, .862)	(.135, .075, .059)	(.107)	(.876, .814)	(.133, .082, .059)	(.096)	
200	.1	.711, 1.124	.009, .022, .031	.140	.686, 1.097	.008, .022, .031	.143	
		(.473, .437)	(.070, .044, .034)	(.027)	(.444, .417)	(.066, .044, .036)	(.052)	
	.3	.712, 1.133	.012, .023, .031	.140	.727, 1.096	.013, .019, .033	.141	
		(.535, .485)	(.073, .049, .038)	(.057)	(.472, .452)	(.074, .048, .039)	(.054)	
	.5	.756, 1.183	.015, .020, .029	.141	.689, 1.104	.013, .024, .031	.145	
		(.645, .604)	(.090, .054, .042)	(.069)	(.572, .520)	(.085, .054, .042)	(.063)	
500	.1	.693, 1.106	.008, .020, .029	.139	.686, 1.087	.008, .020, .031	.139	
		(.285, .275)	(.043, .028, .022)	(.023)	(.279, .259)	(.039, .028, .023)	(.031)	
	.3	.685, 1.092	.010, .021, .031	.140	.680, 1.085	.010, .021, .029	.141	
		(.336, .307)	(.047, .032, .023)	(.037)	(.301, .281)	(.046, .030, .023)	(.034)	
	.5	.666, 1.060	.008, .021, .031	.144	.681, 1.083	.011, .021, .029	.141	
		(.331, .314)	(.052, .032, .024)	(.038)	(.330, .309)	(.051, .031, .024)	(.037)	

with mean covariate value 5), only the estimates at the 60th percentile ( $t_{60}$ ) are presented for space consideration. The results at other time points are qualitatively similar. The standard errors at higher quantiles are observed to be larger, as expected.

The estimates of the three regression parameters  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  are close to their true values, specially for large  $n$ , regardless of the value of  $\alpha$  and the choice of  $P$  matrix with the standard errors decreasing with  $n$ , as expected. However, for small sample sizes, their performance in terms of both bias and precision seems to depend on  $\alpha$ , but not so much on the choice of  $P$ . For large sample size ( $n = 500$ ), the estimates of the parameters  $\gamma_2$  and  $\gamma_3$  are close to their true values; but the performance seems to depend slightly on  $\alpha$ . However, the performance is not as good for small sample sizes depending on both  $\alpha$  and the choice of  $P$  matrix. The estimation errors appear to reduce more slowly with sample size than in the case of estimates of the regression parameters. The estimate of baseline cumulative hazard  $\Lambda_0(t_{60})$  for large sample size ( $n = 500$ ) is close to the true value, but the precision depends slightly on  $\alpha$ . However, for small sample sizes, the performance is not as good with both bias and precision depending on both  $\alpha$  and the choice of  $P$ . This performance seems to be better than that of the estimates of the baseline cumulative cause-specific hazards in Model 1 (See Section 5.1 and Table 1), possibly due to the pooling of information in the estimation of  $\Lambda_0(t_{60})$ .

## 6 An Example

A large animal experiment with a total of 5000 rodents was conducted by the British Industrial Biological Research Association (Peto et al., 1984) to investigate the carcinogenicity of different nitrosamines administered in drinking water. Gart et al. (1986, p58–66) reported details of the data set for the occurrence of pituitary tumors in male rats given N-nitrosodimethylamine (NDMA) in sixteen different concentrations (in ppm) including a control group with 0 ppm. The other fifteen treated groups were given concentrations of 0.033 ppm, 0.066 ppm, 0.132 ppm, 0.264 ppm, 0.528 ppm, 1.056 ppm, 1.584 ppm, 2.112 ppm, 2.640 ppm, 3.168 ppm, 4.224 ppm, 5.280 ppm, 6.336 ppm, 8.448 ppm and 16.896 ppm, respectively. The control group comprised 192 animals and each treated group 48. In addition to this, the data consists of the time to death (in days) for each animal and some information on the cause of each death, described as the ‘context’ of death. There are essentially three causes of death: (1) death without tumor, (2) death due to tumor (fatal), and (3) death due to some other causes but with tumor present (incidental). Because of various pathological problems, observation on the actual cause

of death is sometimes missing. The ‘context’ of an observation gives this occasionally incomplete information. Out of seven such ‘contexts’, three give the exact causes of death, mentioned above. Two other ‘contexts’, *probably fatal* and *probably incidental*, are interpreted (for our analysis) as missing with  $\{2, 3\}$  being the set of possible causes, since presence of tumor is observed. In yet another ‘context’, the presence or absence of tumor is not ascertained but death is known *not* to be caused by tumor; hence  $\{1, 3\}$  is taken as the set of possible causes. In the last ‘context’, the cause of death is not at all ascertainable and, so,  $\{1, 2, 3\}$  is taken as the set of possible causes. Out of the 912 animals in this dataset, only 23 have missing causes with 12, 6 and 5 being observed with  $\{1, 3\}$ ,  $\{2, 3\}$  and  $\{1, 2, 3\}$  as the set of possible causes, respectively. The observed time to death (in days) in this dataset varies from 4 to 1234. The concentration of NDMA (in ppm) is taken as the only covariate in our analysis.

We analyze the data using methods of Section 3 and 4, respectively, for Model 1 and Model 2. Following the overall pattern of missingness in the entire data set, we use

$$P^T = \begin{bmatrix} .006 & .014 & .980 & 0 & 0 & 0 \\ .006 & .014 & 0 & .947 & .033 & 0 \\ .006 & 0 & 0 & 0 & .031 & .963 \end{bmatrix}$$

for the purpose of estimation. The estimates of the regression parameters and the baseline cumulative hazards at two different time points,  $t = 400$  and 1000 days, are presented in Table 3 along with the corresponding standard errors (denoted by SE).

Table 3. Estimates of the regression parameters and the baseline cumulative hazards at  $t = 400$  and 1000 days, for both the models.

Model 1			Model 2		
Parameter	Estimate	SE	Parameter	Estimate	SE
$\beta$	0.4304	0.0164	$\gamma_2$	-1.3555	0.1430
$\Lambda_{01}(400)$	0.0082	0.0017	$\gamma_3$	-1.1808	0.1317
$\Lambda_{01}(1000)$	0.6932	0.0432	$\beta_1$	0.5756	0.0205
$\Lambda_{02}(400)$	$7.0 \times 10^{-11}$	$1.1 \times 10^{-6}$	$\beta_2$	0.0961	0.0895
$\Lambda_{02}(1000)$	0.0793	0.0162	$\beta_3$	0.0937	0.0821
$\Lambda_{03}(400)$	$2.3 \times 10^{-11}$	$1.0 \times 10^{-6}$	$\Lambda_0(400)$	0.0025	$1.5 \times 10^{-5}$
$\Lambda_{03}(1000)$	0.1358	0.0186	$\Lambda_0(1000)$	0.5879	0.3621

We also compute the estimated survival probabilities at different time points in the sixteen different concentration levels using the methods of Sections 3.4 and 4.4 for Model 1 and Model 2, respectively. The estimates for the two models turn out to be very similar. The chemical NDMA seems to have significantly positive effect on the hazard rates (harmful effect on survival). Analysis under Model 2 suggests that this harmful effect on survival is evidently for the first cause of failure (that is, death without tumor). The estimates of the baseline cumulative hazards at later time points have larger standard errors, as expected.

## 7 Concluding Remarks

In this article, we have considered a general pattern of missingness in failure types while dealing with competing risks data. Without making the usual missing at random assumption, we have discussed the regression problem in which the cause-specific hazard rates may depend on certain covariates. We have considered two proportional hazards type regression models. The estimation of the regression parameters has been carried out using the partial likelihood approach. To estimate the baseline cumulative cause-specific hazard rates in Model 1, the non-parametric technique developed in Dewanji and Sengupta (2003) has been modified with the assumption that certain information on the diagnostic probabilities are available from the experimenter. In Model 2, this information was used to derive the cause-specific hazard rates for the ‘modified’ competing risks data. The corresponding baseline cumulative hazard rate has been estimated by using Breslow’s method. Simulation studies indicate reduction of bias and variance of the proposed estimates with sample size. This reduction is more pronounced in the case of the regression parameters as compared to the baseline cumulative hazards.

An important assumption in this work is the time constancy of the  $P$  matrix. This is, however, for simplicity and applicability in practice. If somehow this matrix can be estimated (or, known) at different time points, using extra information or otherwise, that can be readily incorporated in our method. The resulting estimates will naturally be more sensitive to this time-dependent specification of  $P$  matrix. Experimentalists may be able to give some information on the  $P$  or  $q$  matrix; when such information is available, this method gives an opportunity to utilize the same.

As remarked in Section 2, neither of the two models considered here is a special case of the other. While Model 1 permits arbitrary baseline cause-specific hazards, Model 2 assumes proportionality among them. On the other hand, Model 2 allows different effects of a particular covariate on the different cause-specific hazards, which Model 1 does not (See Section 6). Nevertheless, one may be interested to know which of the two gives a better fit for a particular dataset. In this regard, one may consider the Cox-Snell residuals for both the models, as given by  $r_i = \exp(z_i \hat{\beta}) \sum_{j=1}^m \hat{\Lambda}_{0j}(x_i)$  and  $r_i = \hat{\Lambda}_0(x_i) \sum_{j=1}^m \exp(\hat{\gamma}_j + z_i \hat{\beta}_j)$ , with  $\hat{\gamma}_1 = 0$ , for Model 1 and Model 2, respectively. These residuals, along with the corresponding  $\delta_i$ 's, can be used to check for model adequacy. The present work can readily be extended to time-dependent covariates. A more general model, as mentioned in Section 2, may be of interest. This may require a different technique and efforts in that direction is in order.

## References:

- Aalen, O. O. (1978). Non-parametric inference for a family of counting processes. *The Annals of Statistics* **6**, 701-726.
- Andersen, P. K. and Gill R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* **10**, 1100-1120.
- Andersen, P. K. and Borgan, Ø. (1985). Counting process models for life history data: a review. *Scandinavian Journal of Statistics* **12**, 97-158.
- Andersen, P. K., Borgan, Ø., Gill R. D. and Keiding N. (1993). *Statistical Models based on Counting Processes*. Springer-Verlag, New York.
- Basu, S. (2009). Inference About the Masking Probabilities in the Competing Risks Model, *Communications in Statistics Theory and Methods* **38**, 2677-2690.
- Chatterjee, N., Sinha S., Diver W. R and Feigelson, H. S. (2010). Analysis of cohort studies with multivariate and partially observed disease classification data. *Biometrika* **97**, 683-698.
- 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.



- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, series B* **39**, 1-22.
- Dewanji, A. (1992). A note on a test for competing risks with missing failure type. *Biometrika* **79**, 855-857.
- Dewanji, A. and Kalbfleisch, J. D. (1986). Nonparametric methods for survival/sacrifice experiments. *Biometrics* **42**, 325-341.
- Dewanji, A. and Sengupta D. (2003). Estimation of competing risks with general missing pattern in failure types. *Biometrics* **59**, 1065-1072.
- Dinse, G. E. (1982). Nonparametric estimation for partially-complete time and type of failure data. *Biometrics* **38**, 417-431.
- Dinse, G. E. (1986). Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause of death data. *Journal of the American Statistical Association* **81**, 328-336.
- Flehinger, B. J., Reiser, B. and Yashchin, E. (1998). Survival with competing risks and masked causes of failures, *Biometrika* **85**, 151-164.
- Gart, J. J., Krewski, D., Lee, P. N., Tarone, R. E. and Wahrendorf, J. (1986). *Statistical Methods in Cancer Research Volume III: The Design and Analysis of Long-Term Animal Experiments*. IARC Scientific Publications No. 79, International Agency for Research on Cancer, Lyon.
- Goetghebeur, E. and Ryan, L. (1990). A modified log rank test for competing risks with missing failure type. *Biometrika* **77**, 207-211.
- Goetghebeur, E. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing. *Biometrika* **82**, 821-833.
- Kalbfleisch J. D. and Prentice R. L. (1980). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York.
- Lagakos, S. W. and Louis, T. A. (1988). Use of tumour lethality to interpret tumorigenicity experiments lacking cause-of-death data. *Applied Statistics* **37**, 169-179.

- Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data (2nd edition)*. John Wiley & Sons, Hoboken.
- Lo, S. H. (1991). Estimating a survival function with incomplete cause-of-death data. *Journal of Multivariate Analysis* **39**, 217-235.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B* **44**, 226-233.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. John Wiley & Sons, New York.
- Lu, K. and Tsiatis, A. A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure *Biometrics* **57**, 1191-1197.
- Miyakawa, M. (1984). Analysis of incomplete data in competing risks model. *IEEE Transactions in Reliability* **33**, 293-296.
- Mukerjee, H. and Wang, J. L. (1993). Nonparametric maximum likelihood estimation of increasing hazard rate for uncertain cause-of-death data. *Scandinavian Journal of Statistics* **20**, 17-33.
- Peto, R., Gray, R., Brantom, P. and Grasso, P. (1984). Nitrosamine carcinogenesis in 5120 rodents: chronic administration of sixteen different concentrations of NDEA, MPYR and NPIP in the water of 4440 inbred rats, with parallel studies on NDEA alone of the effect of age of starting (3, 6 or 20 weeks) and of species (rats, mice or hamsters). In: O'Neill, I. K., Von Borstel, R. C., Miller, T. C., Long, J. and Bartsch, H., eds, *N-Nitroso Compounds: Occurrence, Biological Effects and Relevance to Human Cancer (IARC Scientific Publications No. 57)*, Lyon, International Agency for Research on Cancer, p627-665.
- Prentice, R. L. and Self, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *The Annals of Statistics* **11**, 804-813.
- Racine-Poon, A. H. and Hoel, D. G. (1984). Nonparametric estimation of the survival function when cause of death is uncertain. *Biometrics* **40**, 1151-1158.

- Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics* **11**, 453-466.
- Sen, A., Banerjee, M., Li, Y. and Noone, A. M. (2010). A Bayesian approach to competing risks analysis with masked cause of death, *Statistics in Medicine* **29** , 16811695.

## Appendix: Asymptotic Results for Model 2

Let  $\tau$  denote a large enough time representing the span of the study duration. Note that, from (17), we can write  $\log L_{P2}(\theta) = C(\theta, \tau)$ , where

$$C(\theta, \tau) = \sum_{i=1}^n \sum_{g \in \mathcal{G}} \int_0^\tau \log \left( f_g(z_i, \theta) \right) dN_{gi}(u) - \int_0^\tau \log \left( \sum_{i=1}^n \sum_{g \in \mathcal{G}} Y_i(u) f_g(z_i, \theta) \right) dN(u).$$

Write  $f_g^{(1)}(z, \theta) = \frac{\partial f_g(z, \theta)}{\partial \theta}$  and  $f_g^{(2)}(z, \theta) = \frac{\partial^2 f_g(z, \theta)}{\partial \theta \partial \theta^T}$ . Also, write

$$\begin{aligned} S_g^{(1)}(\theta, u) &= \frac{1}{n} \sum_{i=1}^n f_g^{(1)}(z_i, \theta) Y_i(u), \\ S_g^{(2)}(\theta, u) &= \frac{1}{n} \sum_{i=1}^n f_g^{(2)}(z_i, \theta) Y_i(u). \end{aligned}$$

Then, the corresponding score vector can be written as

$$\begin{aligned} U(\theta) &= \frac{\partial C(\theta, \tau)}{\partial \theta} \\ &= \sum_{i=1}^n \sum_{g \in \mathcal{G}} \int_0^\tau \frac{f_g^{(1)}(z_i, \theta)}{f_g(z_i, \theta)} dN_{gi}(u) - \int \frac{\sum_i \sum_g f_g^{(1)}(z_i, \theta) Y_i(u)}{\sum_i \sum_g f_g(z_i, \theta) Y_i(u)} dN(u), \\ &= \sum_{i=1}^n \sum_{g \in \mathcal{G}} \int_0^\tau H_{gi}(\theta, u) dM_{gi}(u), \end{aligned}$$

where

$$H_{gi}(\theta, u) = \frac{f_g^{(1)}(z_i, \theta)}{f_g(z_i, \theta)} - \frac{\sum_g S_g^{(1)}(\theta, u)}{\sum_g S_g^{(0)}(\theta, u)},$$

and  $M_{gi}(t) = N_{gi}(t) - \int_0^t \lambda_0(u) f_g(z_i, \theta) Y_i(u) du$  is a local square integrable martingale. The  $M_{gi}$ 's are orthogonal with the variance process given by  $\langle M_{gi}(t) \rangle = \int_0^t \lambda_0(u) f_g(z_i, \theta) Y_i(u) du$ . Note that the maximum partial likelihood estimator  $\hat{\theta}$  is a solution of  $U(\theta) = 0$ , which is assumed to exist. The negative of the matrix of second order partial derivatives of  $C(\theta, \tau)$  can be written as

$$I(\theta) = \sum_{i=1}^n \sum_{g \in \mathcal{G}} \int_0^\tau V_{gi}(\theta, u) dN_{gi}(u), \quad (19)$$

where

$$V_{gi}(\underline{\theta}, u) = \left[ \frac{\sum_g S_g^{(2)}(\underline{\theta}, u)}{\sum_g S_g^{(0)}(\underline{\theta}, u)} - \left( \frac{\sum_g S_g^{(1)}(\underline{\theta}, u)}{\sum_g S_g^{(0)}(\underline{\theta}, u)} \right)^{\otimes 2} \right] - \left[ \frac{f_g^{(2)}(z_i, \underline{\theta})}{f_g(z_i, \underline{\theta})} - \left( \frac{f_g^{(1)}(z_i, \underline{\theta})}{f_g(z_i, \underline{\theta})} \right)^{\otimes 2} \right],$$

and, for a column vector  $a$ ,  $a^{\otimes 2}$  denotes the matrix  $aa^T$ .

**Consistency of  $\hat{\theta}$ :** It can be shown that the process

$$X(\underline{\theta}, t) = \frac{1}{n} \left[ C(\underline{\theta}, t) - C(\underline{\theta}^0, t) \right]$$

has compensator given by  $\tilde{X}(\underline{\theta}, t) =$

$$\sum_{g \in \mathcal{G}} \int_0^t \left[ \frac{1}{n} \sum_{i=1}^n \log \left( \frac{f_g(z_i, \underline{\theta})}{f_g(z_i, \underline{\theta}^0)} \right) f_g(z_i, \underline{\theta}^0) Y_i(u) - \log \left( \frac{\sum_g S_g^{(0)}(\underline{\theta}, u)}{\sum_g S_g^{(0)}(\underline{\theta}^0, u)} \right) S_g^{(0)}(\underline{\theta}^0, u) \right] \lambda_0(u) du,$$

their difference being a local square integrable martingale with the predictable variation process given by  $\langle X(\underline{\theta}, \cdot) - \tilde{X}(\underline{\theta}, \cdot) \rangle (t) =$

$$\frac{1}{n^2} \sum_{g \in \mathcal{G}} \sum_{i=1}^n \int_0^t \left[ \log \left( \frac{f_g(z_i, \underline{\theta})}{f_g(z_i, \underline{\theta}^0)} \right) - \log \left( \frac{\sum_g S_g^{(0)}(\underline{\theta}, u)}{\sum_g S_g^{(0)}(\underline{\theta}^0, u)} \right) \right]^2 f_g(z_i, \underline{\theta}^0) Y_i(u) \lambda_0(u) du.$$

We now assume conditions similar to those in (7.2.1) and Condition VII.2.1(a)(b)(c) of Andersen et al. (1993, p497) on existence of the quantities  $s_g^{(h)}(\underline{\theta}, u)$ , for  $h = 0, 1, 2$  and  $g \in \mathcal{G}$ , satisfying continuity and boundedness conditions in a neighbourhood of  $\underline{\theta}^0$  such that  $S_g^{(h)}(\underline{\theta}, u) \xrightarrow{P} s_g^{(h)}(\underline{\theta}, u)$ , for all  $u \in [0, \tau]$ . In addition, we also assume that, for all  $g \in \mathcal{G}$  and  $u \in [0, \tau]$ , there exists a function  $s_g^{(3)}(\underline{\theta}, u)$  satisfying conditions like the  $s_g^{(h)}(\underline{\theta}, u)$ 's, for  $h = 0, 1, 2$ , such that

$$S_g^{(3)}(\underline{\theta}, u) = \frac{1}{n} \sum_{i=1}^n \log \left( \frac{f_g^{(1)}(z_i, \underline{\theta})}{f_g(z_i, \underline{\theta})} \right)^{\otimes 2} f_g(z_i, \underline{\theta}) Y_i(u) \xrightarrow{P} s_g^{(3)}(\underline{\theta}, u),$$

and that

$$\sup_{\underline{\theta}} \int_0^\tau n^{-2} \sum_{g \in \mathcal{G}} \sum_{i=1}^n \left\| \frac{f_g^{(2)}(z_i, \underline{\theta})}{f_g(z_i, \underline{\theta})} - \left( \frac{f_g^{(1)}(z_i, \underline{\theta})}{f_g(z_i, \underline{\theta})} \right)^{\otimes 2} \right\|^2 f_g(z_i, \underline{\theta}^0) Y_i(u) \lambda_0(u) du \xrightarrow{P} 0,$$

where  $\|\cdot\|$  denotes the supremum norm. Then, using the same argument as those in Prentice and Self (1983), one can prove that  $\langle X(\underline{\theta}, \tau) - \tilde{X}(\underline{\theta}, \tau) \rangle \xrightarrow{P} 0$ , so that  $X(\underline{\theta}, \tau)$  converges in probability to the same limit as does  $\tilde{X}(\underline{\theta}, \tau)$ . We now assume

$$\begin{aligned} S_g^{(4)}(\underline{\theta}, u) &= \frac{1}{n} \sum_{i=1}^n \log \left( \frac{f_g(z_i, \underline{\theta})}{f_g(z_i, \underline{\theta}^0)} \right) f_g(z_i, \underline{\theta}^0) Y_i(u) \xrightarrow{P} s_g^{(4)}(\underline{\theta}, u), \\ S_g^{(5)}(\underline{\theta}, u) &= \frac{1}{n} \sum_{i=1}^n \left( \frac{f_g^{(1)}(z_i, \underline{\theta})}{f_g(z_i, \underline{\theta})} \right) f_g(z_i, \underline{\theta}^0) Y_i(u) \xrightarrow{P} s_g^{(5)}(\underline{\theta}, u), \\ S_g^{(6)}(\underline{\theta}, u) &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{f_g^{(2)}(z_i, \underline{\theta})}{f_g(z_i, \underline{\theta})} - \left( \frac{f_g^{(1)}(z_i, \underline{\theta})}{f_g(z_i, \underline{\theta})} \right)^{\otimes 2} \right\} f_g(z_i, \underline{\theta}^0) Y_i(u) \xrightarrow{P} s_g^{(6)}(\underline{\theta}, u), \end{aligned}$$

and that  $s_g^{(h)}(\underline{\theta}, u)$ , for  $h = 4, 5, 6$ , satisfy the continuity and boundedness conditions, as mentioned earlier. Then, we have

$$\begin{aligned} \tilde{X}(\underline{\theta}, \tau) &\xrightarrow{P} \sum_{g \in \mathcal{G}} \int_0^\tau \left[ s_g^{(4)}(\underline{\theta}, u) - \log \left( \frac{\sum_g s_g^{(0)}(\underline{\theta}, u)}{\sum_g s_g^{(0)}(\underline{\theta}^0, u)} \right) s_g^{(0)}(\underline{\theta}^0, u) \right] \lambda_0(u) du \\ &= h(\underline{\theta}), \text{ say.} \end{aligned}$$

Note that

$$\frac{\partial h(\underline{\theta})}{\partial \underline{\theta}} = \sum_{g \in \mathcal{G}} \int_0^\tau \left[ s_g^{(5)}(\underline{\theta}, u) - \left( \frac{\sum_g s_g^{(1)}(\underline{\theta}, u)}{\sum_g s_g^{(0)}(\underline{\theta}, u)} \right) s_g^{(0)}(\underline{\theta}^0, u) \right] \lambda_0(u) du,$$

which is zero at  $\underline{\theta} = \underline{\theta}^0$ , since  $s_g^{(5)}(\underline{\theta}^0, u) = s_g^{(1)}(\underline{\theta}^0, u)$ . Also,

$$-\frac{\partial^2 h(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}^T} = \sum_{g \in \mathcal{G}} \int_0^\tau \left[ \left\{ \frac{\sum_g s_g^{(2)}(\underline{\theta}, u)}{\sum_g s_g^{(0)}(\underline{\theta}, u)} - \left( \frac{\sum_g s_g^{(1)}(\underline{\theta}, u)}{\sum_g s_g^{(0)}(\underline{\theta}, u)} \right)^{\otimes 2} \right\} s_g^{(0)}(\underline{\theta}^0, u) - s_g^{(6)}(\underline{\theta}, u) \right] \lambda_0(u) du,$$

which, at  $\underline{\theta} = \underline{\theta}^0$ , is equal to

$$\sum_{g \in \mathcal{G}} \int_0^\tau \left[ \frac{s_g^{(3)}(\underline{\theta}^0, u)}{s_g^{(0)}(\underline{\theta}^0, u)} - \left( \frac{\sum_g s_g^{(1)}(\underline{\theta}^0, u)}{\sum_g s_g^{(0)}(\underline{\theta}^0, u)} \right)^{\otimes 2} \right] s_g^{(0)}(\underline{\theta}^0, u) \lambda_0(u) du = \Sigma, \text{ say,}$$

which is assumed to be positive definite. Then, following the arguments of Prentice and Self (1983) and Andersen et al. (1993, p498), we have  $\hat{\underline{\theta}} \xrightarrow{P} \underline{\theta}^0$ .

**Asymptotic normality of  $\hat{\theta}$ :** Note that  $n^{-1/2}U(\hat{\theta})$  is a vector of stochastic integrals with corresponding matrix of predictable variance processes given by

$$n^{-1} \sum_{g \in \mathcal{G}} \sum_{i=1}^n \int_0^\tau \left( H_{gi}(\hat{\theta}, u) \right)^{\otimes 2} f_g(z_i, \hat{\theta}) Y_i(u) \lambda_0(u) du = \\ \int_0^\tau \left[ \sum_{g \in \mathcal{G}} S_g^{(3)}(\hat{\theta}, u) - \frac{\left( \sum_g S_g^{(1)}(\hat{\theta}, u) \right)^{\otimes 2}}{\sum_g S_g^{(0)}(\hat{\theta}, u)} \right] \lambda_0(u) du,$$

which converges in probability to  $\Sigma$ . The desired asymptotic normality for  $n^{-1/2}U(\hat{\theta})$  now follows from the Rebolledo central limit theorem, once it is proved that

$$\sum_{g \in \mathcal{G}} \sum_{i=1}^n \int_0^\tau \left( H_{gil}(\hat{\theta}, u) \right)^2 f_g(z_i, \hat{\theta}) Y_i(u) \lambda_0(u) \times I \left( |H_{gil}(\hat{\theta}, u)| > \epsilon \right) du \xrightarrow{P} 0, \quad (20)$$

for all  $\epsilon$  and  $l$ , where  $H_{gil}(\hat{\theta}, u)$  is the  $l$ th component of  $H_{gi}(\hat{\theta}, u)$ . For this, we assume, similar to Condition VII.2.2 of Andersen et al. (1993, p498), that there exists a  $\delta > 0$  such that

$$n^{-1/2} \sup_{g,i,t,l} \left| \frac{f_{gl}^{(1)}(z_i, \hat{\theta})}{f_g(z_i, \hat{\theta})} \right| Y_i(t) I \left( \log f_g(z_i, \hat{\theta}) > -\delta \mid \frac{f_{gl}^{(1)}(z_i, \hat{\theta})}{f_g(z_i, \hat{\theta})} \right) \xrightarrow{P} 0,$$

where  $f_{gl}^{(1)}(z_i, \hat{\theta})$  is the  $l$ th component of  $f_g^{(1)}(z_i, \hat{\theta})$ . Then, using the same technique as that of Andersen et al. (1993, p499-500) and the above condition, the convergence result in (20) can be proved.

In order to prove the convergence in probability of  $n^{-1}I(\hat{\theta})$  to  $\Sigma$ , let us write

$$\|n^{-1}I(\hat{\theta}) - \Sigma\| \leq \|A\| + \|B\| + \|C\| + \|D\|,$$

where

$$A = n^{-1} \sum_{g \in \mathcal{G}} \sum_{i=1}^n \int_0^\tau \frac{\sum_g S_g^{(2)}(\hat{\theta}, u)}{\sum_g S_g^{(0)}(\hat{\theta}, u)} dN_{gi}(u) - \sum_{g \in \mathcal{G}} \int_0^\tau \frac{\sum_g s_g^{(2)}(\hat{\theta}, u)}{\sum_g s_g^{(0)}(\hat{\theta}, u)} s_g^{(0)}(\hat{\theta}, u) \lambda_0(u) du, \\ B = n^{-1} \sum_{g \in \mathcal{G}} \sum_{i=1}^n \int_0^\tau \left( \frac{\sum_g S_g^{(1)}(\hat{\theta}, u)}{\sum_g S_g^{(0)}(\hat{\theta}, u)} \right)^{\otimes 2} dN_{gi}(u) - \sum_{g \in \mathcal{G}} \int_0^\tau \left( \frac{\sum_g s_g^{(1)}(\hat{\theta}, u)}{\sum_g s_g^{(0)}(\hat{\theta}, u)} \right)^{\otimes 2} s_g^{(0)}(\hat{\theta}, u) \lambda_0(u) du, \\ C = n^{-1} \sum_{g \in \mathcal{G}} \sum_{i=1}^n \int_0^\tau \frac{f_g^{(2)}(z_i, \hat{\theta})}{f_g(z_i, \hat{\theta})} dN_{gi}(u) - \sum_{g \in \mathcal{G}} \int_0^\tau s_g^{(2)}(\hat{\theta}, u) \lambda_0(u) du, \text{ and}$$

$$D = n^{-1} \sum_{g \in \mathcal{G}} \sum_{i=1}^n \int_0^\tau \left( \frac{f_g^{(1)}(z_i, \tilde{\theta})}{f_g(z_i, \tilde{\theta})} \right)^{\otimes 2} dN_{gi}(u) - \sum_{g \in \mathcal{G}} \int_0^\tau s_g^{(6)}(\tilde{\theta}^0, u) \lambda_0(u) du.$$

One can show, by the similar arguments as those in Andersen et al. (1993, p500-501), that each of these four terms converges to zero in probability for any  $\tilde{\theta} = \tilde{\theta}^*$  such that  $\tilde{\theta}^* \xrightarrow{P} \tilde{\theta}^0$ . Therefore, in particular,  $n^{-1}I(\hat{\tilde{\theta}}) \xrightarrow{P} \Sigma$ . The asymptotic normality of  $\hat{\tilde{\theta}}$  with mean  $\tilde{\theta}^0$  and variance  $\Sigma^{-1}$  then follows easily from the identity

$$n^{-1/2}U(\tilde{\theta}^0) = n^{-1} \frac{\partial U(\tilde{\theta}^*)}{\partial \tilde{\theta}^*} \sqrt{n} \left( \hat{\tilde{\theta}} - \tilde{\theta}^0 \right),$$

where  $\tilde{\theta}^*$  is a convex combination of  $\hat{\tilde{\theta}}$  and  $\tilde{\theta}^0$ , and the result that  $\hat{\tilde{\theta}} \xrightarrow{P} \tilde{\theta}^0$ .

**Asymptotic distribution of  $\hat{\Lambda}_0(t)$ :** Note that

$$\begin{aligned} \sqrt{n} \left( \hat{\Lambda}_0(t) - \Lambda_0(t) \right) &= \sqrt{n} \int_0^t J(u) \left[ \left( n \sum_g S_g^{(0)}(\hat{\tilde{\theta}}, u) \right)^{-1} - \left( n \sum_g S_g^{(0)}(\tilde{\theta}^0, u) \right)^{-1} \right] dN(u) \\ &\quad + \sqrt{n} \int_0^t J(u) \left[ \left( n \sum_g S_g^{(0)}(\tilde{\theta}^0, u) \right)^{-1} dN(u) - \lambda_0(u) du \right] \\ &\quad + \sqrt{n} \int_0^t [J(u) - 1] \lambda_0(u) du. \end{aligned}$$

Clearly, the third term  $\xrightarrow{P} 0$ . The second term, say  $\hat{W}(t)$ , is a stochastic integral with respect to a local square integrable martingale as given by  $\sqrt{n} \int_0^t J(u) \left( n \sum_g S_g^{(0)}(\tilde{\theta}^0, u) \right)^{-1} dM(u)$ , where  $M(t) = \sum_g \sum_i M_{gi}(t)$ . The first term, by Taylor series expansion, is

$$-\sqrt{n} \left( \hat{\tilde{\theta}} - \tilde{\theta}^0 \right)^T \int_0^t J(u) \left( \sum_g S_g^{(0)}(\tilde{\theta}^*, u) \right)^{-2} \left( \sum_g S_g^{(1)}(\tilde{\theta}^*, u) \right) \frac{dN(u)}{n},$$

where  $\tilde{\theta}^*$  is a convex combination of  $\hat{\tilde{\theta}}$  and  $\tilde{\theta}^0$ . It can be proved, as in Andersen et al. (1993, p504), that

$$\int_0^t J(u) \left( \sum_g S_g^{(0)}(\tilde{\theta}^*, u) \right)^{-2} \left( \sum_g S_g^{(1)}(\tilde{\theta}^*, u) \right) \frac{dN(u)}{n} \xrightarrow{P} \int_0^t \frac{\sum_g s_g^{(1)}(\tilde{\theta}^0, u)}{\sum_g s_g^{(0)}(\tilde{\theta}^0, u)} \lambda_0(u) du.$$

Also, note that, by writing  $U(\tilde{\theta}, t) = \partial C(\tilde{\theta}, t) / \partial \tilde{\theta}$ , we have  $\langle U(\tilde{\theta}^0, \cdot), \hat{W}(\cdot) \rangle (t) =$

$$\sqrt{n} \int_0^t \sum_g \sum_i \left[ \frac{f_g^{(1)}(z_i, \tilde{\theta}^0)}{f_g(z_i, \tilde{\theta}^0)} - \frac{\sum_g S_g^{(1)}(\tilde{\theta}^0, u)}{\sum_g S_g^{(0)}(\tilde{\theta}^0, u)} \right] \frac{J(u)}{n \sum_g S_g^{(0)}(\tilde{\theta}^0, u)} f_g(z_i, \tilde{\theta}^0) Y_i(u) \lambda_0(u) du$$



$$= \sqrt{n} \int_0^t \left[ \frac{\sum_g S_g^{(1)}(\tilde{\theta}^0, u)}{\sum_g S_g^{(0)}(\tilde{\theta}^0, u)} - \frac{\sum_g S_g^{(1)}(\theta^0, u)}{\sum_g S_g^{(0)}(\theta^0, u)} \right] J(u) \lambda_0(u) du = 0.$$

Hence, the local martingales  $U(\tilde{\theta}^0, t)$  and  $\hat{W}(t)$  are orthogonal. It follows that  $\sqrt{n} \left( \hat{\theta} - \tilde{\theta}^0 \right)$  and  $\hat{W}(t)$  are asymptotically independent. Then, the process

$$W(t) = \sqrt{n} \left( \hat{\Lambda}_0(t) - \Lambda_0(t) \right) + \sqrt{n} \left( \hat{\theta} - \tilde{\theta}^0 \right)^T \int_0^t \frac{\sum_g s_g^{(1)}(\tilde{\theta}^0, u)}{\sum_g s_g^{(0)}(\tilde{\theta}^0, u)} \lambda_0(u) du \quad (21)$$

and  $\sqrt{n} \left( \hat{\theta} - \tilde{\theta}^0 \right)$  are asymptotically independent. Note that, from the expression of  $\hat{W}(t)$ , the limiting distribution of  $W(t)$  is that of a mean zero Gaussian process with variance function given by  $\int_0^t \left( \sum_g s_g^{(0)}(\tilde{\theta}^0, u) \right)^{-1} \lambda_0(u) du$ . Hence,

$$\sqrt{n} \left( \hat{\Lambda}_0(t) - \Lambda_0(t) \right) = W(t) - \sqrt{n} \left( \hat{\theta} - \tilde{\theta}^0 \right)^T \int_0^t \frac{\sum_g s_g^{(1)}(\tilde{\theta}^0, u)}{\sum_g s_g^{(0)}(\tilde{\theta}^0, u)} \lambda_0(u) du$$

is a mean zero Gaussian process with variance function given by  $\sigma^2(t) =$

$$\int_0^t \left( \sum_g s_g^{(0)}(\tilde{\theta}^0, u) \right)^{-1} \lambda_0(u) du + \left( \int_0^t \frac{\sum_g s_g^{(1)}(\tilde{\theta}^0, u)}{\sum_g s_g^{(0)}(\tilde{\theta}^0, u)} \lambda_0(u) du \right)^T \Sigma^{-1} \left( \int_0^t \frac{\sum_g s_g^{(1)}(\tilde{\theta}^0, u)}{\sum_g s_g^{(0)}(\tilde{\theta}^0, u)} \lambda_0(u) du \right), \quad (22)$$

which can be consistently estimated by  $\hat{\sigma}^2(t) =$

$$n^{-1} \left[ \int_0^t \frac{dN(u)}{\left( \sum_g S_g^{(0)}(\hat{\theta}, u) \right)^2} + \left( \int_0^t \frac{\sum_g S_g^{(1)}(\hat{\theta}, u)}{\left( \sum_g S_g^{(0)}(\hat{\theta}, u) \right)^2} dN(u) \right)^T I^{-1}(\hat{\theta}) \left( \int_0^t \frac{\sum_g S_g^{(1)}(\hat{\theta}, u)}{\left( \sum_g S_g^{(0)}(\hat{\theta}, u) \right)^2} dN(u) \right) \right]. \quad (23)$$

It follows from the above results that the asymptotic covariance between  $\sqrt{n} \left( \hat{\theta} - \tilde{\theta}^0 \right)$  and  $\sqrt{n} \left( \hat{\Lambda}_0(t) - \Lambda_0(t) \right)$  is given by

$$-\Sigma^{-1} \int_0^t \frac{\sum_g s_g^{(1)}(\tilde{\theta}^0, u)}{\sum_g s_g^{(0)}(\tilde{\theta}^0, u)} \lambda_0(u) du. \quad (24)$$