

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

Anup Dewanji<sup>1</sup>, P. G. Sankaran<sup>2</sup>, Debasis Sengupta<sup>1</sup> and Bappa Karmakar<sup>1</sup>

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

2. *Department of Statistics, Cochin University of Science and Technology, Cochin-22, 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. We also consider an example from an animal experiment to illustrate our methodology.

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

## 1 Introduction

In survival studies, it is common that death of individuals may be attributable to more than one cause or factor. Competing risks models are usually employed to analyse such type of data. In competing risks set up. For each individual, we observe a random vector  $(T, J)$  where  $T$  is possibly censored survival time and  $J$  represents cause for death( exactly one of say  $m$  possible causes). However, due to inadequacy on the diagnostic mechanism one often uncertain about the true failure type or it is reluctant to reveal the value of  $J$  for certain individuals. For example, in carcinogenicity studies, death of individuals can be classified into death s with tumour or deaths due to other reasons. However, there is uncertainty in assigning these causes of death even if the presence of tumour can be ascertained. See Dinse(1986) and Lagakos and Louis (1988) for details. In certain cases, one cannot even distinguish deaths with tumour and death without tumour because they are totally cannibalized.

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.

## 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

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

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

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

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  $p_{gj}(t, z) = p_{gj'}(t, z)$ , 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

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

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

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

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

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

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

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

where  $\lambda_{0g}^*(t) = \sum_{j \in \mathcal{G}} p_{gj}(t) \lambda_{0j}(t)$ , 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, the 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  $t_{(g_i)}$ , we consider the conditional probability that the individual  $(g_i)$  with covariate value  $z_{(g_i)}$  fails at time  $t_{(g_i)}$ , given the history up to time  $t_{(g_i)} -$  and that one failure with missing pattern  $g$  occurs at time  $t_{(g_i)}$ . The method of partial likelihood then gives

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

where  $R(t_{(g_i)})$  is the risk set at time  $t_{(g_i)} -$  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

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

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

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

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)

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

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  $\hat{\sigma}_{gg'}(t) =$

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

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  $\Lambda_{\sim_0}^*(t)$  as the  $(2^m - 1) \times 1$  vector of  $\Lambda_{0g}^*(t)$ 's and  $\Lambda_{\sim_0}(t)$  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),

$$\Lambda_{\sim_0}^*(t) = P \Lambda_{\sim_0}(t). \tag{10}$$

Then, following Dewanji and Sengupta (2003), we have a consistent estimate of  $\Lambda_{\sim_0}(t)$  as

$$\hat{\Lambda}_{\sim_0}(t) = \left( \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{P} \right)^{-1} \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{\Lambda}_{\sim_0}^*(t), \tag{11}$$

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  $\hat{\Lambda}_{\sim_0}^*(t)$  is the  $(2^m - 1) \times 1$  vector of  $\hat{\Lambda}_{0g}^*(t)$ 's.

Note that  $\hat{\Lambda}_{\sim_0}(t) - \Lambda_{\sim_0}(t)$

$$\begin{aligned} &= \left( \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{P} \right)^{-1} \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{\Lambda}_{\sim_0}^*(t) - \left( \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{P} \right)^{-1} \left( \hat{P}^T \hat{\Sigma}^{-1}(t) \hat{P} \right) \Lambda_{\sim_0}(t) \\ &= \hat{A}(t) \left( \hat{\Lambda}_{\sim_0}^*(t) - \Lambda_{\sim_0}^*(t) \right) + \hat{A}(t) \left( P - \hat{P} \right) \Lambda_{\sim_0}(t), \end{aligned}$$

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  $(P^T \Sigma^{-1}(t) P)^{-1} P^T \Sigma^{-1}(t)$  with  $\Sigma(t)$  being the asymptotic covariance matrix of  $\hat{\Lambda}_{\sim_0}^*(t) - \Lambda_{\sim_0}^*(t)$ . Hence, using the weak convergence result of  $\hat{\Lambda}_{\sim_0}^*(t) - \Lambda_{\sim_0}^*(t)$  and the fact that  $\hat{P}$  is a consistent estimate of  $P$ , one can establish weak convergence of  $\hat{\Lambda}_{\sim_0}(t) - \Lambda_{\sim_0}(t)$  to a  $m$ -variate mean zero Gaussian process with the covariance matrix at time  $t$  given by  $(P^T \Sigma^{-1}(t) P)^{-1}$ , 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  $\hat{\Lambda}_{0g}^*(t)$ '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  $\hat{P}$  is of full column-rank. Even if the rank of  $\hat{P}$  is less than  $m$ , some components of  $\hat{\Lambda}_{\sim_0}(t)$  may be estimable.

## 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

$$\lambda_g^*(t, z) = \lambda_0(t) f_g(z, t, \theta), \tag{12}$$

where  $f_g(z, t, \theta) = f_g(z, t, \gamma, \beta) = \sum_{j \in g} p_{gj}(t) e^{\gamma_j + z \beta_j}$ , 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 in (13) 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

$$\begin{aligned} L_{P2}(\theta) &= \prod_{i=1}^k \left[ \frac{\lambda_{g_{(i)}}^*(t_{(i)}, z_{(i)})}{\sum_{l \in R(t_{(i)})} \sum_{g \in \mathcal{G}} \lambda_g^*(t_{(i)}, z_l)} \right] \\ &= \prod_{i=1}^k \left[ \frac{f_{g_{(i)}}(z_{(i)}, \theta)}{\sum_{l \in R(t_{(i)})} \sum_{g \in \mathcal{G}} f_g(z_l, \theta)} \right], \end{aligned} \tag{13}$$

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 (13) is maximized to get an estimate of  $\theta$ , denoted by  $\hat{\theta}$ . Note that the model (12) 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 (13) 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 a 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  $\log L_{P2}(\theta)$ .

### 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

$$\hat{\Lambda}_0(t) = \int_0^t \frac{J(u)}{n \sum_{g \in \mathcal{G}} S_g^{(0)}(\hat{\theta}, u)} dN(u), \tag{14}$$

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  $\sigma^2(t)$  along with a consistent estimate  $\hat{\sigma}^2(t)$  which are given by

$$\sigma^2(t) = \int_0^t \left( \sum_g S_g^{(0)}(\theta^0, u) \right)^{-1} \lambda_0(u)du + \left( \int_0^t \frac{\sum_g S_g^{(1)}(\theta^0, u)}{\sum_g S_g^{(0)}(\theta^0, u)} \lambda_0(u)du \right)^T \Sigma^{-1} \left( \int_0^t \frac{\sum_g S_g^{(1)}(\theta^0, u)}{\sum_g S_g^{(0)}(\theta^0, u)} \lambda_0(u)du \right), \tag{15}$$

and

$$\hat{\sigma}^2(t) = n^{-1} \left[ \int_0^t \frac{dN(u)}{\left( \sum_g S_g^{(0)}(\hat{\theta}, u) \right)^2} + \left( \int_0^t \frac{\sum_g S_g^{(1)}(\hat{\theta}, u)}{\left( \sum_g S_g^{(0)}(\hat{\theta}, u) \right)^2} dN(u) \right)^T I^{-1}(\hat{\theta}) \left( \int_0^t \frac{\sum_g S_g^{(1)}(\hat{\theta}, u)}{\left( \sum_g S_g^{(0)}(\hat{\theta}, u) \right)^2} dN(u) \right) \right]. \tag{16}$$

The proof is similar to that in Andersen et al. (1993, p503-505) with little modification. The asymptotic covariance between  $\sqrt{n} \left( \hat{\theta} - \theta^0 \right)$  and  $\sqrt{n} \left( \hat{\Lambda}_0(t) - \Lambda_0(t) \right)$  is also obtained, which can be consistently estimated by

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

**References:**

Andersen, P. K. and Borgan, Ø. (1985). Counting process models for life history data: a review. *Scandinavian Journal of Statistics* **12**, 97-158.

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.

Dewanji, A. (1992). A note on a test for competing risks with missing failure type. *Biometrika* **79**, 855-857.

Dewanji, A. and Sengupta D. (2003). Estimation of competing risks with general missing pattern in failure types. *Biometrics* **59**, 1065-1072.

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.

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

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.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. John Wiley & Sons, New York.

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**, 1681-1695.