

# Marginal Methods for Multivariate Failure Times Under Event-Dependent Censoring

Longyang Wu<sup>1</sup> & Richard J. Cook<sup>2</sup>

<sup>1</sup> Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA

<sup>2</sup> Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

Correspondence: Richard J. Cook, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON. N2L 3G1, Canada. E-mail: rjcook@uwaterloo.ca

Received: June 2, 2014 Accepted: June 25, 2014 Online Published: July 3, 2014

doi:10.5539/ijsp.v3n3p111

URL: <http://dx.doi.org/10.5539/ijsp.v3n3p111>

*The research is supported by the Natural Sciences and Engineering Research Council of Canada (RJC, Grant No. 101093) and the Canadian Institutes for Health Research (RJC, Grant No. 105099). Richard Cook is a Canada Research Chair in Statistical Methods for Health Research. The authors thank Novartis Pharmaceuticals for permission to use the data from the clinical trial*

## Abstract

Many chronic diseases put individuals at increased risk of several different types of adverse clinical events. Typically these events are combined to define composite events which are then used as the basis of treatment evaluation. A potentially more efficient approach is to conduct separate marginal assessments of the effect of treatment on each component and then to synthesize this information across each type of event. While there is considerable potential for more powerful tests of treatment effect in this setting, it is possible that dependent censoring can cause problems. This happens when the occurrence of one type of event increases the risk of withdrawal from a study and hence alters the probability of observing events of other types. The purpose of this article is to formulate a model which reflects this type of mechanism, to evaluate the effect on the asymptotic and finite sample properties of marginal estimates, and to examine the performance of estimators obtained using flexible inverse probability weighted marginal estimating equations. Data from a motivating study are used for illustration.

**Keywords:** dependent censoring, inverse weighting, marginal methods, multivariate failure time data

## 1. Introduction

Many chronic disease processes put individuals at risk for several types of events and it is often of interest to examine the effect of treatment on the risk of occurrence for each type (Dabrowska, 2006; Hougaard, 2000). In settings involving life history processes, multiple events can occur during a particular period of observation and composite endpoints are also routinely used as a basis for treatment assessment (Freemantle, Calvert, Wood, Eastaugh, & Griffin, 2003). In time to event data, a composite endpoint uses the time of the first event as the response, regardless of the type, and is appealing since it permits the use of standard methods for survival analysis (Kalbfleisch & Prentice, 2002; Lawless, 2003). It does not, however, lead to treatment comparisons based on a full characterization of the disease process. For this reason, in clinical trials investigators have increasingly turned to use of multiple endpoints and regulatory agencies are increasingly requiring demonstration of efficacy new interventions based on such analyses (Buzney & Kimball, 2008; Fleming & Lin, 2000; Freemantle & Calvert, 2007; Wei & Glidden, 1997).

There are three common frameworks for the analysis of multivariate failure time data including frailty-based models (Therneau & Grambsch, 2000), copula models (Liang, Self, Bandeen-Roche, & Zeger, 1995; Nelsen, 2006), and marginal methods (Wei, Lin, & Weissfeld, 1989). While frailty models and copula models yield multivariate distributions, they require assumptions regarding the frailty distribution or the copula function respectively. Such fully specified models are appealing if interest lies in estimating the degree of association between two or more event times or for prediction, but when assessing treatment effects in clinical trials it is generally desirable to make minimal assumptions and maintain robustness. The marginal approach of Wei et al. (1989) requires only

specification of a Cox regression model for each type of event but no dependence structure is specified for distinct failure times do not have a specific dependence structure. Simultaneous inference regarding the estimates of the marginal regression coefficients is carried out through use of a robust sandwich type variance estimator. This method is easily implemented in most major statistical software packages such as R/S-PLUS and SAS (Therneau & Grambsch, 2000) and is widely used in clinical trials (Lin, 1994).

The marginal approach of Wei et al. (1989) is based on a working independence assumption and the robust covariance matrix and hence has similarities with the generalized estimating equations approach of Liang and Zeger (1986) for dealing with clustered categorical data. A number of methodological advances have been made in the field of multivariate failure time data analysis which are based on a similar framework (Cai, Fan, Jiang, & Zhou, 2007; Cai, Fan, H. Zhou, & Y. Zhou, 2007; Cai & Prentice, 1995, 1997; Cai & Schaebel, 2004; Clegg, Cai, & Sen, 1999; Clegg, Cai, Sen, & Kupper, 2000; Kang & Cai, 2009; Lee, Wei, & Amato, 1992; Liang, Self, & Chang, 1993; Spiekerman & Lin, 1998; Greene & Cai, 2004; Yin & Cai, 2004, 2005). Since the marginal approach of Wei et al. (1989) is based on a partially specified model, however, it is only valid if censoring is completely independent of the failure time process. In studies of life history processes, when individuals are to be followed after the occurrence of events, it is common for censoring to be associated with occurrence of one or more particular types of events, yielding event-dependent censoring. For example, if the occurrence of the first event alerts a physician to the fact that the current treatment “failed” for a patient, it may increase the risk that they will be withdrawn from the study. In general, when marginal regression models are applied to multivariate failure times under such a dependent censoring scheme, biased (martingale) estimating equations are specified and the resulting estimators are inconsistent (Cook, Lawless, & Lee, 2007; Cook, Lawless, Lakhali-Chaieb, & Lee, 2009). The purpose of this article is to study the asymptotic bias that may result from a dependent censoring scheme of this sort under a working independence assumption. We also explore the utility of inverse probability of censoring weighted estimating equations developed by Rotnitzky and Robins (1995) and Robins and Finkelstein (2000). for marginal analysis of multivariate failure time data as a means of correcting for event-dependent censoring.

The remainder of this article is organized as follows. In Section 2 we provide the notation and formulate the models for the multivariate failure times and censoring processes. In Section 3 we investigate the asymptotic and empirical biases arising from a naive marginal approach. The inverse probability of censoring weighted marginal estimating equations are developed in Section 4, where we investigate the large sample properties of resulting estimators of treatment effect and the baseline hazard functions. Sections 5 and 6 contain simulation studies and an illustrative example respectively, and concluding remarks are made in Section 7.

## 2. Notation and Model Specification

### 2.1 Model Formulation for Multivariate Failure Times

Let  $T_k$  denote the time of the type  $k$  event and  $\{N_k(s), 0 < s\}$  denote the corresponding right-continuous counting process, where  $N_k(t) = I(T_k \leq t)$  indicates that the type  $k$  event has occurred at or before time  $t$ ,  $dN_k(t) = 1$  if a type  $k$  event occurs at time  $t$ , and  $dN_k(t) = 0$  otherwise. We further let  $N(t) = (N_1(t), \dots, N_K(t))'$  and remark that the multivariate counting process  $\{N(s), 0 < s\}$  is specified when interest lies in jointly modeling the occurrence of all event types. Suppose  $Z(t)$  is a vector of fixed, exogeneous or endogenous covariates and let  $\{Z(s), 0 < s\}$  denote the covariate process. The full history at  $t$  contains information on the number and times of events over  $[0, t)$  and covariate data over  $[0, t]$  and is denoted  $H(t) = \{N(s), 0 \leq s < t, Z(s), 0 \leq s \leq t\}$ .

The intensity function for type  $k$  events is

$$\lambda_k(t|H(t)) = \lim_{\Delta t \rightarrow 0} \frac{P(\Delta N_k(t) = 1|H(t))}{\Delta t},$$

where  $\Delta N_k(t) = N_k(t + \Delta t^-) - N_k(t^-)$  is the number of the events occurring over the interval  $[t, t + \Delta t)$ . The inclusion of a dependence on the history for process  $\ell$  in the intensity for type  $k$  events accommodates the association between processes. For continuous time processes where at most one event can occur at any time, these intensity functions fully define the multivariate counting processes (Gill, Keiding, & Andersen, 1993).

While this formulation completely specifies the multivariate process, in the context of clinical trials it is undesirable to assess treatment effects conditional on endogeneous variables (Kalbfleisch & Prentice, 2002) and hence intensity functions do not offer an appealing framework for analyses. Instead treatment effects are more naturally expressed in terms of marginal proportional hazards regression models of the form

$$\lambda_k(t|Z) = \lambda_{0k}(t) \exp(\beta_k Z) \quad (1)$$

where  $\lambda_{0k}(t)$  is an unspecified positive function,  $\beta_k$  is a regression parameter and  $Z$  is a fixed covariate which equals 1 for individuals receiving the experimental treatment and zero for those receiving a control therapy. The marginal hazard ratio reflecting the effect of treatment on type  $k$  events is then simply  $\exp(\beta_k)$ . The cumulative baseline hazard function is  $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(u)du$  and the marginal survivor function is  $\mathcal{F}_k(t|Z; \theta_k) = P(T_k \geq t|Z; \theta_k) = \exp(-\Lambda_{0k}(t)e^{\beta_k Z})$ , where  $\theta_k$  contains the regression coefficient  $\beta_k$  and the parameters indexing  $\lambda_{0k}(\cdot)$ . When marginal models of this type are specified it is necessary to address the association in the failure times differently than is done for intensity-based analyses. This is conveniently achieved using copula functions (Nelsen, 2006).

A copula function  $C_\phi(u_1, \dots, u_K)$  in  $K$  dimensions defines a multivariate distribution on the unit hypercube  $[0, 1]^K$  with uniform margins. Parametric copula functions are indexed by a parameter denoted by  $\phi$ , which characterizes the association between the components of the marginal quantities. Such functions offer a convenient way of constructing multivariate distributions with marginal distributions of a specified form. Specifically, the marginal probability integral transformation of each random variable can be applied to create a  $K$  dimensional vector of uniform random variables. These in turn are then viewed as the components of a multivariate uniform random variable with their joint distribution governed by a given copula. Thus the joint survival function  $\mathcal{F}_{12}(t_1, t_2|z) = P(T_1 \geq t_1, T_2 \geq t_2|z)$  can be specified by linking the two marginal survival functions via a copula function

$$\mathcal{F}_{12}(t_1, t_2|z; \Omega) = C_\phi(\mathcal{F}_1(t_1|z; \theta_1), \mathcal{F}_2(t_2|z; \theta_2)),$$

where  $\Omega = (\theta', \phi)'$  with  $\theta = (\theta'_1, \theta'_2)'$ . The Clayton copula is widely used in survival analysis and yields a joint survival distribution of the form

$$C_\phi(\mathcal{F}_1(t_1|z; \theta_1), \mathcal{F}_2(t_2|z; \theta_2)) = \left( [\mathcal{F}_1(t_1|z; \theta_1)]^{-\phi} + [\mathcal{F}_2(t_2|z; \theta_2)]^{-\phi} - 1 \right)^{-1/\phi}.$$

The degree of association between two failure times is often expressed in terms of Kendall's  $\tau$  which is given by  $\tau = \phi/(\phi + 2)$  ( $0 \leq \tau \leq 1$ ) for the Clayton copula where  $\tau = 0$  and  $\tau = 1$  correspond to the cases of independence and perfect association respectively.

## 2.2 A Model for Event-Dependent Censoring

When multiple clinical events arise investigators often withdraw patients from trials if there is a perception that the randomized treatment is no longer appropriate. If subjects are censored at the time of study withdrawal, this creates a type of event-dependent censoring which leads to inconsistent parameter estimates under partially specified models. Consider a setting in which the intention is to follow individuals over the interval  $[0, C^\dagger]$  where  $C^\dagger$  is a time of administrative censoring. Let  $C$  denote a random time of withdrawal where  $0 < C \leq C^\dagger$ . Let  $N^C(t) = I(C \leq t)$  and  $\{N^C(s), 0 < s\}$  be the counting process for the random censoring time where  $dN^C(t) = 1$  if random withdrawal occurs at time  $t$  and  $dN^C(t) = 0$  otherwise. Let  $Y^\dagger(s) = I(s \leq C^\dagger)$ ,  $Y(s) = I(s \leq C)$ ,  $\bar{Y}(s) = Y(s)Y^\dagger(s)$  and  $\bar{Y}_k(s) = \bar{Y}(s)I(s \leq T_k)$  indicate whether an individual is under observation and at risk of a type  $k$  event. Let  $d\bar{N}_k(t) = \bar{Y}_k(t)dN_k(t)$ ,  $\bar{N}_k(t) = \int_0^t d\bar{N}_k(s)$ , and  $\bar{N}(t) = (\bar{N}_1(t), \dots, \bar{N}_K(t))'$ . We observe  $\{(\bar{N}(s), N^C(s)), 0 < s \leq C^\dagger, Z\}$  and let  $\bar{H}(t) = \{(\bar{N}(s), N^C(s)), 0 < s < t, Z\}$  denote the observed history for the event and censoring processes.

The intensity for the random censoring time  $C$  is

$$\lambda^c(t|\bar{H}(t)) = \lim_{\Delta t \rightarrow 0} \frac{P(\Delta N^C(t) = 1|\bar{H}(t))}{\Delta t}, \quad (2)$$

which accommodates a dependence between the censoring, event times, and possibly the treatment assignment. When the analysis of the failure times adopted for multivariate failure time data (Wei et al., 1989) is often based on a working independence assumption, it is the dependence on the event times that is particularly problematic.

The dependence on the event history can take many forms, but in what follows we consider a particular model with the censoring intensity

$$\lambda^c(t|\bar{H}(t)) = \lambda_0^c(t) \exp(\alpha_1 N_1(t) + \alpha_2 N_2(t)), \quad (3)$$

where  $\lambda_0^c(t)$  is a baseline intensity for censoring and  $(\alpha_1, \alpha_2)'$  are regression coefficients which reflect how the risk of withdrawal changes upon the occurrence of type 1 and type 2 events; we write  $d\Lambda_0^c(t) = \lambda_0^c(t)dt$ . Thus  $\exp(\alpha_k)$  is the multiplicative factor by which the intensity of censoring increases upon the occurrence of a type  $k$  event,  $k = 1, 2$ , and if  $\alpha_1 = \alpha_2 = 0$ ,  $\min(C, C^\dagger)$  is an independent right-censoring time.

### 3. Asymptotic Biases of Marginal Estimators

In this section, we investigate the asymptotic bias caused by event-dependent censoring when marginal estimating equations are specified based on a working independence assumption, as is done for the multivariate approach of Wei et al. (1989). We first present the general framework and then study the one-sample estimates in detail. The data for a sample of  $n$  independent individuals are denoted by  $\{(\tilde{N}_i(s), N_i^C(s)), 0 < s \leq C^\dagger, Z_i, i = 1, \dots, n\}$  where we introduce the subscript  $i$  to index individuals. We assume that the marginal distribution of  $T_{ik}|z_i$  is exponential with  $\lambda_k(t|z_i) = \lambda_{ik}(t) = \lambda_k \exp(\beta_k z_i)$  and the joint distribution of  $(T_{i1}, T_{i2})|Z_i$  is defined through a Clayton copula. The naive marginal estimating equations are

$$U_{k1}(t) = \sum_{i=1}^n \bar{Y}_{ik}(t) (dN_{ik}(t) - d\Lambda_{ik}(t)) \quad (4)$$

$$U_{k2}(\beta) = \sum_{i=1}^n \int_0^\infty \bar{Y}_{ik}(u) (dN_{ik}(u) - d\Lambda_{ik}(u)) z_i \quad (5)$$

where  $d\Lambda_{ik}(t) = \exp(\beta_k z_i) \lambda_{0k}(t) dt$ ,  $k = 1, 2$ . The profile estimate of the cumulative baseline hazard for type  $k$  events is then

$$\tilde{\Lambda}_{0k}(t; \beta_k) = \int_0^t d\tilde{\Lambda}_{0k}(u; \beta_k) = \int_0^t \frac{\sum_{i=1}^n \bar{Y}_{ik}(u) dN_{ik}(u)}{\sum_{i=1}^n \bar{Y}_{ik}(u) \exp(\beta_k z_i)} \quad (6)$$

The estimate  $\hat{\beta}_k$  is obtained as the solution to

$$\sum_{i=1}^n \int_0^\infty \bar{Y}_{ik}(u) (dN_{ik}(u) - d\tilde{\Lambda}_{0k}(u; \beta_k) \exp(\beta_k z_i)) z_i = 0, \quad (7)$$

and upon substitution of  $\hat{\beta}_k$  into (6) the Breslow estimate  $\hat{\Lambda}_{0k}(t) = \tilde{\Lambda}_{0k}(t; \hat{\beta}_k)$  is obtained.

With completely independent censoring the above estimating equations yield consistent estimators of the cumulative baseline hazard function  $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(u) du$ , as well as the regression coefficient  $\beta_k$ ,  $k = 1, 2$ . If censoring is governed by an intensity featuring a dependence on the event history, (4) and (5) may yield inconsistent estimators (Robins & Finkelstein, 2000; Datta & Satten, 2001). The limiting value of the estimator of the cumulative baseline hazard function under a general censoring scheme is

$$\int_0^t d\Lambda_{0k}^*(u; \beta_k^*) = \int_0^t \frac{E(\bar{Y}_{ik}(u) dN_{ik}(u))}{E(\bar{Y}_{ik}(u) \exp(\beta_k^* z_i))} \quad (8)$$

where  $\beta_k^*$  is the limiting value of  $\hat{\beta}_k$  obtained as the implicit solution to

$$\int_0^\infty E[\bar{Y}_{ik}(t) z_i dN_{ik}(t)] - \frac{E[\bar{Y}_{ik}(t) \exp(\beta_k z_i) z_i]}{E[\bar{Y}_{ik}(t) \exp(\beta_k z_i)]} E[\bar{Y}_{ik}(t) dN_{ik}(t)] \quad (9)$$

The expectation  $E(\cdot)$  in (8) and (9) is taken with respect to the true process defined here in terms of the marginal distributions, the Clayton copula, and the event-dependent censoring intensity (3). Some details on these derivations are given in Appendix A.

To illustrate the bias of the naive marginal approach in estimation of the cumulative hazard function, we consider a separate analysis of two treatment groups and hence focus on the one-sample problem; in this case we restrict attention to (4) with  $z_i = 1$  for the treatment group and  $z_i = 0$  for the control group. The Nelson-Aalen estimator of  $\Lambda_{0k}(t)$  is

$$\hat{\Lambda}_{0k}(t|Z = z) = \int_0^t \frac{\sum_{i=1}^n \bar{Y}_{ik}(t) I(Z_i = z) dN_{ik}(t)}{\sum_{i=1}^n \bar{Y}_{ik}(t) I(Z_i = z)} \quad (10)$$

The bias of this estimator is investigated empirically by simulating a data set of  $n = 2000$  individuals having an equal risk of events with  $\lambda_k = 2$ ,  $k = 1, 2$  and  $C^\dagger$  chosen to give 14% administrative censoring. A Clayton copula model was used to induce an association between the failure times with  $\tau = 0.2$  and  $\tau = 0.6$ . The censoring intensity was based on (3) with  $\alpha_1 = \log 1.3$  and  $\alpha_2 = \log 3.5$ , and  $\lambda_0^c(t) = \lambda_0^c$  was chosen to give 50% net censoring for type 1 events. Figure 1 displays the true cumulative hazard function and naive estimates based on

the Nelson-Aalen estimator when  $\tau = 0.2$  and  $\tau = 0.6$ ; the left panel contains the results for type 1 events and the right panel for type 2 events. The plots demonstrate that the naive method yields a conservative estimate of the cumulative hazard function with the magnitude of this bias increasing with time. The stronger the association between the failure times, the greater the empirical bias. It is interesting to note that the bias is greater for the estimated cumulative hazard for type 1 events since the strength of the dependence between type 2 event times and censoring is greatest.

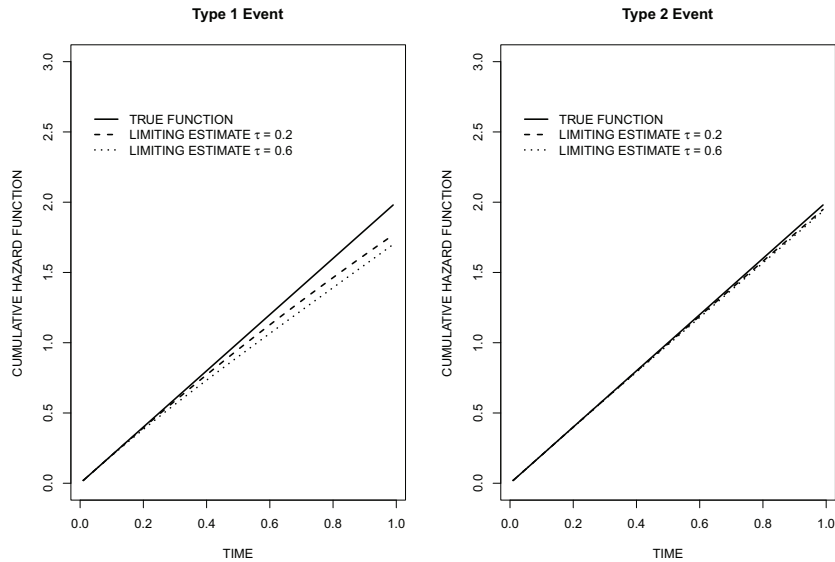


Figure 1. Plots of the true cumulative hazard function for type 1 (left panel) and type 2 (right panel) events with corresponding limiting values of naive (unweighted) Nelson-Aalen estimates when  $\tau = 0.2$  and  $\tau = 0.6$ ; bivariate failure time model defined by a Clayton copula with exponential margins ( $\lambda_1 = \lambda_2 = 2$ );  $C^\dagger$  chosen to give 10% administrative censoring; censoring intensity had  $\alpha_1 = \log 1.3$  and  $\alpha_2 = \log 3.5$  with  $\lambda_0^c$  chosen to give 50% net censoring rate

#### 4. IPCW Weighted Marginal Regression

##### 4.1 IPCW Weighted Estimating Equations

The estimating Equations (4) and (5) can be modified to yield consistent estimators by introducing inverse probability of censoring weights (Robins, 1993). If, as in (3), the censoring intensity for individual  $i$  at time  $t$  depends on the history  $\bar{H}_i(t)$ , then let  $G_i(t) = P(C_i \geq t | \bar{H}_i(t))$  which by the product integration (Gill et al., 1993) can be written as  $G_i(t) = \prod_{u < t} [1 - d\Lambda^c(u | \bar{H}_i(u))]$ . Furthermore let  $G(t)$  be survival function of the random right censoring time under the scenario of independent random censoring, in which case the censoring intensity is a hazard function with  $\lambda^c(t)dt = d\Lambda^c(t)$ ; note we drop the subscript 0 here since it is no longer a baseline censoring intensity. Then again by product integration we obtain  $G(t) = \prod_{u < t} [1 - d\Lambda^c(u)]$ .

The marginal estimating functions corresponding to (4) and (5) are then defined as

$$U_{k1}(t) = \sum_{i=1}^n \frac{G(t)\bar{Y}_i(t)}{G_i(t)} [dN_{ik}(t) - d\Lambda_{ik}(t)] \tag{11}$$

$$U_{k2}(\beta_k) = \sum_{i=1}^n \int_0^\infty \frac{G(u)\bar{Y}_i(u)}{G_i(u)} (dN_{ik}(u) - d\Lambda_{ik}(u)) z_i, \tag{12}$$

respectively. The weights are introduced to ensure that the resulting marginal estimating equations are unbiased and hence that consistent estimators are obtained. We show  $E(U_{k1}(t)) = 0$  by taking the expectation in stages. First we note that

$$E_{Y_i(t)}(U_{k1}(t) | H_i(t), dN_i(t)) = G(t)Y_i(t)(dN_{ik}(t) - d\Lambda_{ik}(t))$$

and then  $E_{dN_i(t)}[G(t)Y_i(t)(dN_{ik}(t) - d\Lambda_{ik}(t))|H_i(t)]$  is given by

$$G(t) [Y_i(t)E(dN_{ik}(t)|H_i(t)) - d\Lambda_{ik}(t)]$$

and taking the expectation of this with respect to the process history gives zero. By similar arguments it can be shown that the expectation of (12) is also zero.

From the derivations above it is clear that  $G(t)$  is not necessary to guarantee unbiasedness of the estimating equations. In fact it will have no role in estimation of the baseline hazard function since it cancels in the numerator and denominator of Breslow's estimator given in (11). Robins (1993) showed, however, that inclusion of  $G(t)$  in (12) yields estimators of  $\beta_k$  which are more efficient than those obtained when  $G(t) = 1$ . When  $G(t)$  is set to 1 we refer to these as regular IPCW weights and when  $G(t)$  is defined more generally we refer to these as stabilized weights.

In practise, of course, to use (11) and (12) the functions  $G(t)$  and  $G_i(t)$  must be consistently estimated. Let  $\Lambda^c(t) = \int_0^t d\Lambda^c(u)du$  where  $d\Lambda^c(u) = \lambda^c(u)du$  is the crude censoring hazard under the working independence assumption between the censoring and event processes. In this case  $\Lambda^c(u)$  is estimated simply as

$$\widehat{\Lambda}^c(t) = \int_0^t \frac{\sum_{i=1}^n \bar{Y}_i(u)dN_i^C(u)}{\sum_{i=1}^n \bar{Y}_i(u)}$$

which gives

$$\widehat{G}(t) = \prod_{u < t} [1 - d\widehat{\Lambda}^c(u)] ,$$

the usual Kaplan-Meier estimate of the survival function for the censoring distribution.

Correct specification of the model for  $G_i(t)$  is more crucial since it is what renders the inverse weighted estimating functions unbiased. If one believes the censoring intensity function (3) is correct, one can adopt it in the following. Alternatively, we prefer to relax the proportionality assumptions in (3) and consider a more robust stratified model for the censoring intensity with  $d\Lambda^c(t|H_i(t)) = d\Lambda^c(t|N_{i1}(t) = l, N_{i2}(t) = m) = d\Lambda_{lm}^c(t)$ , where  $l, m = 0, 1$ . The corresponding nonparametric estimate is then

$$d\widehat{\Lambda}^c(t|\bar{H}_i(t)) = \frac{\sum_{i=1}^n I(C_i = t, N_{i1}(t^-) = l, N_{i2}(t^-) = m)}{\sum_{i=1}^n I(C_i \geq t, N_{i1}(t^-) = l, N_{i2}(t^-) = m)} , \quad (13)$$

if  $N_i(t^-) = (l, m)'$ . Then again by product integration we have

$$\widehat{G}_i(t) = \prod_{u < t} [1 - d\widehat{\Lambda}^c(u|\bar{H}_i(u))] .$$

Upon substituting these estimates into (11), we obtain the weighted Breslow estimator of the cumulative baseline hazard function for type  $k$  events

$$\int_0^t d\widehat{\Lambda}_{0k}^w(u) = \int_0^t \frac{\sum_{i=1}^n \bar{Y}_{ik}(u)dN_{ik}(u)/\widehat{G}_i(u)}{\sum_{i=1}^n \bar{Y}_{ik}(u) \exp(\widehat{\beta}_k z_i)/\widehat{G}_i(u)} ,$$

where  $\widehat{\beta}_k$  is the estimate obtained from the weighted score function for the  $k$ th type of event:

$$\sum_{i=1}^n \int_0^\infty \frac{\widehat{G}(t)\bar{Y}_{ik}(t)}{\widehat{G}_i(t)} \left[ z_i - \frac{\sum_{i=1}^n \bar{Y}_{ik}(t) \exp(\beta_k z_i) z_i / \widehat{G}_i(t)}{\sum_{i=1}^n \bar{Y}_{ik}(t) \exp(\beta_k z_i) / \widehat{G}_i(t)} \right] dN_{ik}(t) . \quad (14)$$

The limiting distribution of estimated regression coefficients are given by the following theorem.

**Theorem 4.1** Under regularity conditions, if  $\widehat{G}(t)/\widehat{G}_i(t)$  is a consistent nonparametric estimate of  $G(t)/G_i(t)$ , then  $\sqrt{n}(\widehat{\beta}_k - \beta_k)$  converges in distribution to a vector of zero-mean normal random variable with a variance that can be consistently estimated by  $\widehat{I}_k^{-1} \widehat{\Sigma}_k \widehat{I}_k^{-1}$ , where  $\widehat{I}_k = -n^{-1} \partial U_k(\widehat{\beta}_k) / \partial \beta_k$  and  $\widehat{\Sigma}_k = n^{-1} \sum_{i=1}^n U_{ik}^2(\widehat{\beta}_k)$ .

**Remark 4.1** By the multivariate central limit theorem ( $\sqrt{n}(\widehat{\beta}_1 - \beta_1), \sqrt{n}(\widehat{\beta}_2 - \beta_2), \dots, \sqrt{n}(\widehat{\beta}_K - \beta_K)$ ) converges in distribution to a vector of zero-mean multivariate normal random variable with a covariance matrix that can be consistently estimated by  $\widehat{\Sigma}$  where the  $(l, m)$  element is  $n^{-1} \sum_{i=1}^n \widehat{I}_l^{-1} U_{il}(\widehat{\beta}_l) U_{im}(\widehat{\beta}_m) \widehat{I}_m^{-1}$ ,  $l, m = 1, \dots, K$ .

**Remark 4.2** The variance estimator is the usual robust sandwich estimator that can be directly obtained from R/S-PLUS or SAS using a suitably constructed dataframe in the counting process format.

**Remark 4.3** The global estimate of treatment is simply a linear combination of all component-specific estimates of treatment effect  $\widehat{\beta}_k, k = 1, \dots, K$ ,

$$\widehat{\beta}_c = \mathbf{c}(\widehat{\beta})' \widehat{\beta}, \tag{15}$$

where  $\widehat{\beta} = (\widehat{\beta}_1, \dots, \widehat{\beta}_K)$ , the weight  $\mathbf{c}(\widehat{\beta}) = \widehat{\Sigma}(\widehat{\beta})^{-1} \mathbf{J} [\mathbf{J}' \widehat{\Sigma}(\widehat{\beta})^{-1} \mathbf{J}]^{-1}$  is chosen to estimate the weight matrix to minimize the variance in the class of all linear estimators;  $\widehat{\Sigma}(\widehat{\beta})$  is the estimate for the variance-covariance matrix of  $\widehat{\beta}$  given in Remark 4.1, and  $\mathbf{J} = (1, \dots, 1)'$ . An asymptotically equivalent combined estimate can be obtained by fitting a single Cox model by stratifying on event type and constraining the coefficients to be the same for the different types of events (Therneau & Grambsch, 2000).

### 5. Empirical Investigation

Simulation studies were conducted to assess the finite-sample performance of the estimators obtained through the IPCW marginal estimating equations. The failure times  $T_1$  and  $T_2$  were generated using a Clayton copula with exponential margins given the treatment assignment. Without loss of generality we set  $C^\dagger = 1$ . For a given value of the association parameter  $\phi$ ,  $\lambda_1$  and  $\lambda_2$  were determined to give a particular stochastic ordering  $q = P(T_1 < T_2 | z = 0)$  and rate of administrative censoring  $p$  for  $T = \min(T_1, T_2)$  (i.e.  $P(T < C^\dagger) = p$ ).

We define the intensity for the random censoring time according to (3) with  $\lambda_0^c(t) = \lambda_0^c$ . For given  $(\alpha_1, \alpha_2)'$ ,  $\lambda_0^c$  is specified to ensure a prescribed probability of observing the first event in the control arm  $P(T_1 < C | z = 0) = \pi$  is satisfied. We set  $\beta_1 = \beta_2 = \log 0.80$ ,  $\tau = 0.4$  and varied  $q = P(T_1 < T_2 | z = 0)$  over 0.25, 0.50 and 0.75. We set  $\alpha_1 = \log 1.3$  and  $\alpha_2 = \log 3.5$  and set  $\lambda_0^c$  so that  $P(T_1 < C | z = 0) = 0.4$ . The regression coefficients were obtained by solving (4) and (5) to obtain unweighted estimates, and (11) and (12) to obtain weighted estimates with  $G(t) = 1$  or more generally. The estimate  $\widehat{\beta}_c$  of (15) is a pooled estimate of  $\widehat{\beta}_1$  and  $\widehat{\beta}_2$ .

Table 1. Empirical results from simulation studies examining the frequency properties of estimators of the marginal regression coefficients (Type 1 and Type 2) and global estimates (Global) under dependent censoring

Event	Weight	$P(T_1 < T_2   z = 0)$								
		0.25			0.50			0.75		
		BIAS	ESE	ASE	BIAS	ESE	ASE	BIAS	ESE	ASE
Type 1	None	0.015	0.145	0.146	0.006	0.141	0.140	0.000	0.136	0.134
	IPCW	0.001	0.172	0.172	-0.003	0.184	0.173	-0.004	0.168	0.159
	Stabilized IPCW	0.003	0.152	0.153	0.000	0.151	0.147	-0.000	0.142	0.138
Type 2	None	0.008	0.114	0.113	0.004	0.130	0.129	0.000	0.151	0.154
	IPCW	0.009	0.116	0.117	0.005	0.144	0.143	0.001	0.180	0.179
	Stabilized IPCW	0.009	0.113	0.113	0.004	0.130	0.129	-0.000	0.151	0.153
Global	None	0.011	0.109	0.110	0.005	0.116	0.117	0.001	0.122	0.122
	IPCW	0.006	0.120	0.121	0.002	0.140	0.143	-0.001	0.146	0.141
	Stabilized IPCW	0.007	0.112	0.113	0.003	0.122	0.120	-0.000	0.126	0.125

Note: Weighted estimates are obtained by inverse probability weighted estimating equations with weights  $1/G_i(t)$  and stabilized weights  $G(t)/G_i(t)$ .

The summary statistics of the estimated regression coefficients are reported in Table 1 including the empirical bias (Bias), the empirical standard error (ESE), and the average robust standard error (ASE) based on the large sample results. The simulations were conducted with 2000 samples each of  $n = 500$  individuals. There is generally very good agreement between the empirical and average asymptotic standard errors in all settings. While the biases are generally quite small in the unweighted analyses, it is apparent that the impact of dependent censoring is different for the two marginal parameters and the magnitude of bias is influenced by both the stochastic ordering of the events as well as the  $(\alpha_1, \alpha_2)'$  parameters. That is, the bias is greater for type 1 events since the dependence between the time of the type 2 event and censoring is greatest (this is what induces the dependent censoring from marginal analyses of type 1 events). The empirical biases of the weighted estimators are generally smaller

indicating the advantages of inverse weighting. The estimates obtained using the weight  $1/G_i(t)$  have considerably larger standard errors than the estimates obtain using the stabilized weight  $G(t)/G_i(t)$ , which are in turn much closer to the standard errors of the unweighted analyses. Thus the weight function  $G(t)/G_i(t)$  leads to estimator with the best performance in that it provides protection against bias induced by dependent censoring at the price of a relatively small increase in the standard error.

**6. Application**

Here we consider data from a trial of 380 breast cancer patients with skeletal metastases (Hortobagyi et al., 1998). The experimental treatment is a bisphosphonate which is studied for its palliative effect of reducing the incidence of fractures and the need for radiotherapy for the treatment of bone pain. Following randomization patients were followed for up to 24 months for the occurrence of these types of events and interest lies in the marginal analysis of the time to first fracture and the time to first round of radiotherapy, as well as a global assessment of the treatment effect using a Wei-Lin-Weissfeld analysis. To address the issue of the competing risks of death we adapt the marginal analyses to be based on fracture-free survival and radiation-free survival.

Figure 2 gives plots of the cumulative intensity for censoring by fracture status (left panel) and radiotherapy status (right panel). The slope of the cumulative intensity for censoring following the occurrence of the first fracture is considerably steeper than it is for fracture-free individuals, revealing a dependence between fracture status and censoring. The same pattern is seen in the right panel in that the rate of censoring for patients who have had one round of radiotherapy is higher (reflected by the steeper slope) than those who have not required radiotherapy. These plots are suggestive of a need to deal with dependent censoring for marginal analyses.

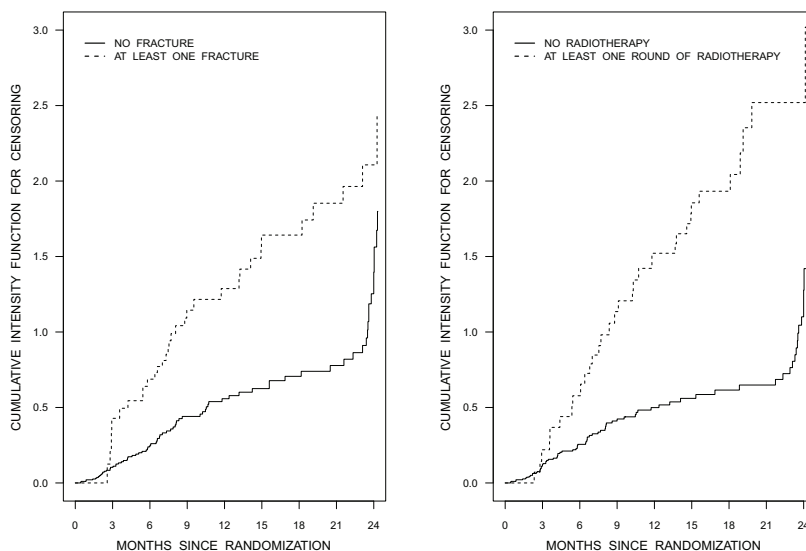


Figure 2. Plot of cumulative intensity function for censoring (13) by fracture status (left panel) and radiotherapy status (right panel) for placebo arm patients

The estimates of the cumulative baseline hazard for the analysis based on fracture-free survival and radiotherapy-free survival are given in Figure 3. There is empirical evidence of a greater effect of dependent censoring in the fracture-free survival analysis in that there is a bigger difference between the unweighted and weighted estimates than is seen for the radiotherapy-free survival analysis. This is compatible with the simulation results in that the larger difference between the cumulative intensities for censoring by radiotherapy status in Figure 2 (in comparison to the estimates of the censoring intensity by fracture status), suggests the dependence between censoring and radiotherapy is greater. This in turn will have a greater effect on the estimates related to the fracture-free survival endpoint. While the effects are not large, there is a suggestion that the unweighted analysis yields a conservative estimate of the event rates since the naive estimate is lower than either of the weighted estimates.



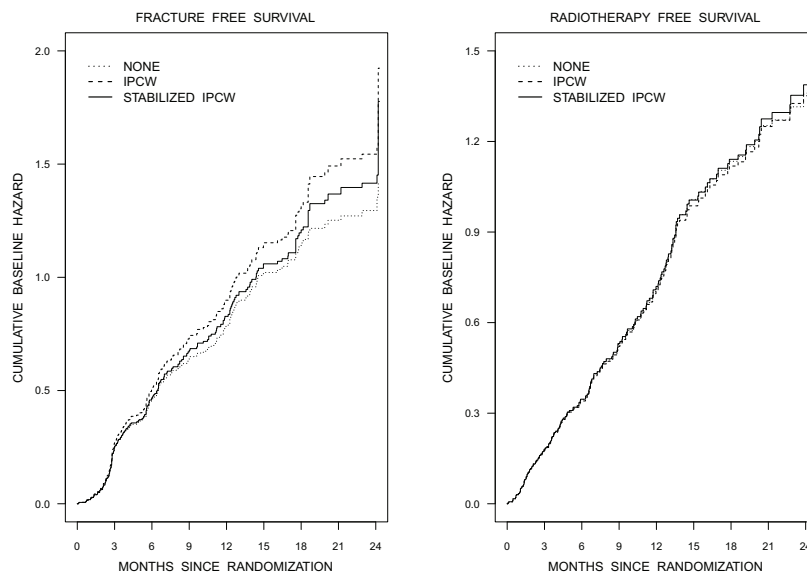


Figure 3. Plot of estimated cumulative baseline hazard function  $\int_0^\infty d\widehat{\Lambda}_{0k}^w(u; \widehat{\beta}_k)$

Table 2 contains the results of the marginal and global regression analyses. Here we see the estimates of the treatment effect from the use of stabilized weights are slightly larger than those obtained from an unweighted analysis. The relative risk reduction for fracture-free survival was 22.6% for the unweighted analysis compared to a 24.9% relative risk reduction from analysis using stabilized weights. Moreover, in contrast to the unweighted analysis, the results based on the stabilized weights give statistically significant evidence of a treatment benefit for fracture-free survival ( $p= 0.0465$ ). Very similar estimates are seen for the radiotherapy-free survival endpoint for unweighted and weighted analyses using stabilized weights. Finally, the use of stabilized weights incurs a relatively small penalty in terms of efficiency as the standard errors are very close to those of the unweighted analyses. For the global analysis, there is a 32.3% relative risk reduction from the unweighted analysis and a 33.2% reduction from the weighted analysis using stabilized weights.

Table 2. Estimates obtained by fitting separate marginal Cox models and using the global Wei-Lin-Weissfeld analysis (Global) in the analysis of data from the trial of breast cancer patients with skeletal metastases; unweighted and weighted analyses

Endpoint	Weight	EST	SE	HR	95% CI	p-value
Fracture	None	-0.256	0.142	0.774	(0.586, 1.023)	0.0714
	IPCW	-0.483	0.188	0.617	(0.427, 0.891)	0.0100
	Stabilized IPCW	-0.286	0.144	0.751	(0.567, 0.996)	0.0465
Radiation	None	-0.547	0.152	0.579	(0.430, 0.780)	0.0003
	IPCW	-0.505	0.158	0.604	(0.442, 0.823)	0.0014
	Stabilized IPCW	-0.550	0.154	0.577	(0.426, 0.781)	0.0004
Global	None	-0.393	0.104	0.675	(0.551, 0.827)	0.0002
	IPCW	-0.493	0.144	0.611	(0.461, 0.810)	0.0006
	Stabilized IPCW	-0.404	0.125	0.668	(0.523, 0.853)	0.0012

Note: Weighted estimates are obtained by inverse probability weighted estimating equations with weights  $1/G_i(t)$  and stabilized weights  $G(t)/G_i(t)$ .

### 7. Discussion

Multivariate failure time data are frequently encountered in clinical trials and observational studies. Frailty models are a popular choice for the analysis of multivariate failure time data, but they do not yield estimates of treatment

effect which have a simple marginal interpretation. Such models can be formulated using copula functions, but it is undesirable to base inferences on a particular parametric model and for this reason the marginal methods proposed by Wei et al. (1989) are preferred. Inference regarding regression coefficients in this framework are carried out by use of a robust sandwich-type variance estimate easily computed in SAS or R/S-PLUS (Therneau & Grambsch, 2000).

With multivariate failure time data, studies are usually designed to follow individuals for the occurrence of all types of events up until some administrative censoring time. In this setting however, event occurrence may cause investigators to withdraw patients from a trial if it is thought that following the protocol is no longer in the patients' best interests. This can occur, for example, when one or more clinical endpoints are observed. The simplicity of the marginal analysis of Wei et al. (1989) arises from the working independence assumption. This enables the use of standard software for point estimation, but the validity of this hinges on the censoring being independent of the event processes. When this is not satisfied inconsistent estimates are obtained for all marginal parameters including the cumulative baseline hazard functions and the regression coefficients from the marginal Cox models. Use of inverse probability of censoring weights are known to address this problem (Robins, 1993) and we have shown that this strategy can be put to good use in the context of multivariate failure time data. In this study, we proposed a marginal IPCW approach to analyze multivariate failure times with event-dependent censoring and demonstrated the effectiveness of the proposed approach using simulation studies. It offers a viable approach to correcting for event-dependent censoring when data from each type of event are to be used for analyses. An alternative, less efficient approach to this problem would be to artificially censor individuals after the occurrence of the first event. This creates a competing risk problem for which one could use standard cause-specific Cox regression models. Alternatively models based on the cumulative incidence function could be employed as developed by Fine and Gray (1999).

#### A. The Limiting Value of Unweighted Estimators

The numerator of (8),  $E(\bar{Y}_{ik}(t)dN_{ik}(t))$ , is calculated by noting

$$\begin{aligned} E(\bar{Y}_{ik}(t)dN_{ik}(t)) &= E_{Z_i}\{[E_{\bar{Y}_{ik}(t)|Z_i}E(dN_{ik}(t)|\bar{Y}_{ik}(t), Z_i)]\} \\ &= E_{Z_i}\{P(\bar{Y}_{ik}(t) = 1|Z_i)Pr(dN_{ik}(t) = 1|\bar{Y}_{ik}(t) = 1, Z_i)\} \end{aligned}$$

This gives

$$\sum_{z=0}^1 P(Z_i = z)P(dN_{ik}(t) = 1, \bar{Y}_{ik}(t) = 1|Z_i = z)$$

In the same way, the denominator  $E(\bar{Y}_{ik}(t) \exp\{\beta z_i\})$  can be obtained as

$$\sum_{z=0}^1 \exp(\beta_k z)P(Z_i = z)P(dN_{ik}(t) = 1, \bar{Y}_{ik}(t) = 1|Z_i = z).$$

The probabilities  $P(\bar{Y}_{ik}(t) = 1|z_i = z)$  can be obtained analytically under the marginal models, Clayton copula (2) and event-dependent censoring mechanism (3). Hence the limiting value of the estimator of the baseline cumulative hazard function  $d\Lambda_{0k}^*(t)$  can be obtained. The limiting value of estimators of the regression coefficients are obtained following these calculations from (9).

#### B. Proof of Theorem 4.1

The following derivations are provided in the context of a more general marginal Cox model with a vector of time-varying covariates, and we suppress the dependence on the type of event,  $k$ . For a random sample of  $n$  subjects, the observed data consist of  $\{\bar{N}_i(\cdot), \bar{Y}_i(\cdot), Z_i(\cdot), i = 1, \dots, n\}$ . We let  $S^{(k)}(\beta, t) = n^{-1} \sum_{i=1}^n \bar{Y}_i(t)Z_i(t)^{\otimes k} \exp\{Z_i'(t)\beta\}/G_i(t)$  and  $s^{(k)}(\beta, t) = E(S^{(k)}(\beta, t))$ , where  $k = 0, 1, 2$  and  $a^{\otimes 0} = 1$ ,  $a^{\otimes 1} = a$ , and  $a^{\otimes 2} = aa'$ . We also let  $\bar{Z}(\beta, t) = S^{(1)}(\beta, t)/S^{(0)}(\beta, t)$ ,  $\bar{z}(\beta, t) = s^{(1)}(\beta, t)/s^{(0)}(\beta, t)$  and

$$I_n = \frac{1}{n} \sum_{i=1}^n \int_0^{C^\dagger} \{Z_i(t) - \bar{Z}(\beta_0, t)\}^{\otimes 2} \bar{Y}_i(t)w_i(t)/(nS^{(0)}(\beta, t))dN_i(t).$$

Here we assume administrative censoring at  $C^\dagger$  and impose the following regularity conditions:

- 1)  $P(C_i \geq C^\dagger) > 0, i = 1, \dots, n;$

- 2)  $N_i(C^\dagger)$ ,  $i = 1, \dots, n$  are bounded by a constant;
- 3)  $|Z_{ji}(0)| + \int_0^{C^\dagger} |dZ_{ji}(t)| \leq K$  for all  $j = 1, \dots, p$  and  $i = 1, \dots, n$ , where  $Z_{ji}(t)$  is the  $j$ th component of  $Z_i(t)$  and  $K$  is a constant.
- 4)  $I = E \left[ \int_0^{C^\dagger} \{Z_i(t) - \bar{z}(\beta_0, t)\}^{\otimes 2} Y_i(t) w_i(t) \exp\{Z_i'(t)\beta_0\} d\Lambda_0(t) \right]$  is positive definite.
- 5)  $I_n = I + o_p(1)$ .

*Proof.* We first establish the asymptotic normality of  $\sqrt{n}(\widehat{\beta} - \beta_0)$  using the true weight function  $w_i(t) = G(t)/G_i(t)$ , and then prove that the effect of using a nonparametric estimation of  $w_i(t)$  can be ignored in the variance estimation. Consider the weighted Cox log partial likelihood that leads to the partial score function (14),

$$L(\beta, C^\dagger) = \sum_{i \leq n} \int_0^{C^\dagger} w_i(u) Y_i(u) \{Z_i'(u)\beta - \log R(\beta, u)\} dN_i(u), \quad (16)$$

where  $R(\beta, u) = nS^{(0)}(\beta, u)$ . By Lemma A2 of Hjort and Pollard (1993), we can expand  $\log R(\beta, u)$  around the true value  $\beta_0$

$$\log R(u, \beta_0 + x) - \log R(u, \beta_0) = \bar{Z}'(u)x + \frac{1}{2}x'V(u)x + v(x, u), \quad (17)$$

where  $V(u) = \sum_{i=1}^n w_i(u) Y_i(u) \exp\{Z_i'(u)\beta_0\} \{Z_i(u) - \bar{Z}(u)\}^{\otimes 2} / R(\beta_0, u)$ . The reminder term  $v(u, x)$  is bounded by  $\frac{4}{3} \max_{i \leq n} |(Z_i(u) - \bar{Z}(u))'x|$ .

Using (17), we expand  $L(\beta, C^\dagger)$  around the true value  $\beta_0$  to approximate

$$L(\beta_0 + t/\sqrt{n}) - L(\beta_0) \quad (18)$$

by

$$\sum_{i \leq n} \int_0^{C^\dagger} \left[ n^{-\frac{1}{2}} (Z_i(u) - \bar{Z}(u))'t - \frac{1}{2} n^{-1} t'V(u)t - v\left(\frac{t}{\sqrt{n}}, u\right) \right] w_i(t) Y_i(u) dN_i(u),$$

which can be further written as

$$U_n't - \frac{1}{2} t'I_n t - r_n(t), \quad (19)$$

where  $U_n = n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^{C^\dagger} (Z_i(u) - \bar{Z}(u)) w_i(u) Y_i(u) dN_i(u)$  and  $r_n(t) = \sum_{i \leq n} \int_0^{C^\dagger} v_n(t/\sqrt{n}, u) dN_i(u)$ , which is bounded by  $\sum_{i \leq n} \int_0^{C^\dagger} \frac{4}{3} (2K)^3 |t|^3 n^{-\frac{3}{2}} dN_i(u)$  where  $K$  is the absolute bound on the covariates. The latter term is  $O(n^{-\frac{1}{2}})$  and goes to zero as  $n \rightarrow \infty$ . Hence, (18) can be approximated by  $U_n't - \frac{1}{2} t'I_n t$ , which can be maximized at  $t = I_n^{-1} U_n$ . Note that its concavity in  $t$  follows from the convexity of  $\log R(\beta, u)$  in  $\beta$ . Suppose  $\widehat{\beta}$  is a solution to the estimating equation (14) that maximizes the log partial likelihood (16), then  $\sqrt{n}(\widehat{\beta} - \beta_0)$  maximizes (18). By the assumption of (5) and the extension of the Basic Corollary of Hjort and Pollard (1993),

$$\sqrt{n}(\widehat{\beta} - \beta_0) = I^{-1} U_n + o_p(1), \quad (20)$$

and the asymptomatic normality of  $\widehat{\beta}$  can be established if the asymptomatic normality of  $U_n$  is established.

We now follow the arguments in Lin, Wei, Yang, and Ying (2000) to establish the asymptomatic normality of  $U_n$ . Let

$$M_i(t) = \int_0^t w_i(u) \bar{Y}_i(u) dN_i(u) - \int_0^t w_i(u) \bar{Y}_i(u) \exp\{Z_i'(u)\beta_0\} d\Lambda_0(u),$$

then write the partial score function as

$$U_n(\beta_0, t) = n^{-\frac{1}{2}} \bar{M}_Z(t) - \int_0^t \bar{Z}_n(u) d\bar{M}(u), \quad (21)$$

where  $\bar{M}(t) = \sum_{i \leq n} M_i(t)$  and  $\bar{M}_Z(t) = \sum_{i \leq n} \int_0^t Z_i(u) dM_i(u)$ . For fixed time  $t$ ,  $\bar{M}(t)$  and  $\bar{M}_Z(t)$  are sum of iid zero-mean random variables. By the multivariate central limit theorem,  $(n^{-\frac{1}{2}} \bar{M}(t), n^{-\frac{1}{2}} \bar{M}_Z(t))$  converges in finite

dimensional distribution to a zero-mean Gaussian processes  $(\mathcal{W}_M, \mathcal{W}_{M_Z})$ . Note that  $M_i(t)$  is the difference of two monotone functions. The bounded variation assumption (3) implies that  $Z_i(\cdot)$  is bounded and we may assume without loss of generality that  $Z_i(\cdot) \geq 0$ ; otherwise  $Z_i(\cdot)$  can be written as difference of two nonnegative, non-decreasing functions by the Jordan decomposition. Thus each component of  $\int_0^t Z_i(u) dM_i(u)$  is also a difference of two monotone functions in  $t$ . Therefore, by the weak convergence of the monotone class as in the example of 2.11.16 in van der Vaart and Wellner (1996),  $(n^{-\frac{1}{2}}\bar{M}(t), n^{-\frac{1}{2}}\bar{M}_Z(t))$  is tight and converges weakly to  $(\mathcal{W}_M, \mathcal{W}_{M_Z})$  and it can be verified that both satisfy Kolmogorov-Chentsov criterion (e.g. Corollary 16.9 in Kallenberg, 2010) so that they have continuous sample path with respect to the Euclidean distance.

By the Skorokhod strong embedding theorem (Shorack & Wellner, 1986, pp. 47-48), there exists a probability space in which  $(n^{-\frac{1}{2}}\bar{M}(t), n^{-\frac{1}{2}}\bar{M}_Z(t), S^{(1)}(\beta_0, t), S^{(0)}(\beta_0, t))$  converges almost surely to  $(\mathcal{W}_M, \mathcal{W}_{M_Z}, s^{(1)}(\beta_0, t), s^{(0)}(\beta_0, t))$ . Note that  $Z_i(\cdot) \geq 0$  ( $i = 1, \dots, n$ ) is a monotone function by assumption and  $1/G_i(t)$  is nonnegative and nondecreasing function in  $t$ ; therefore,  $S^{(0)}(\beta_0, t)$  and  $S^{(1)}(\beta_0, t)$  are nonnegative monotone functions in  $t$ . Then, we can apply the Lemma 1 in Lin et al. (2000) twice to show that

$$n^{-\frac{1}{2}} \int_0^t \frac{S^{(1)}(\beta_0, u)}{S^{(0)}(\beta_0, u)} d\bar{M}(u) \rightarrow \int_0^t \frac{s^{(1)}(\beta_0, u)}{s^{(0)}(\beta_0, u)} d\mathcal{M}_M(u) \quad (22)$$

uniformly in  $t$  almost surely. Combining this result with the convergence of  $n^{-\frac{1}{2}}\bar{M}_Z$  to  $\mathcal{M}$  yields the uniform convergence of  $U_n(\beta_0, t)$  to  $\mathcal{W}_{M_Z}(t) - \int_0^t \bar{z}(\beta_0, u) d\mathcal{W}_M(u)$  almost surely in the new probability space and thus weakly in the original probability space. This limiting Gaussian process has covariance function  $\Sigma(s, t) = E[\int_0^s \{Z_i(u) - \bar{z}(\beta_0, u)\} dM_i(u) \int_0^t \{Z_i(u) - \bar{z}(\beta_0, u)\}^T dM_i(u)]$ ,  $0 \leq s, t \leq C^\dagger$ , between times  $s$  and  $t$ . Then by the Basic Corollary of Hjort and Pollard (1993),  $\sqrt{n}(\hat{\beta} - \beta_0)$  converges in distribution to a multivariate normal distribution  $MVN(0, I^{-1}\Sigma I^{-1})$ .

We now prove that the nonparametric estimation of the weight function  $w_i(t)$  can be ignored in the variance estimation. Note that the weight  $w_i(t)$  is a generic weight function  $W(\cdot)$  evaluated at time  $t$  based on the history  $H_i(t)$  of subject  $i$ , where  $H(t)$  is defined in Section 2.1. Therefore, we can write  $w_i(t) = W(t, H_i(t))$ . We then suppress the arguments of the function  $W$  for notational convenience. Let  $U(N_i(t), Y_i(t), \beta, W)$  be the partial score functions corresponding to (14). Then  $E_{H(t)}(U(N_i(t), Y_i(t), \beta_0, W_0)) = 0$  for true value  $\beta_0$  and the true weight function  $W_0$  as showed in Section 4. The estimating function (14) is to solve

$$\frac{1}{n} \sum_{i \leq n} U(N_i(t), Y_i(t), \beta, \widehat{W}) = 0 \quad (23)$$

for  $\beta$ , by plugging in a nonparametric estimate  $\widehat{W}$ . Here the  $\widehat{W}$  is obtained by a stratified Kaplan-Meier estimator. Note that the partial score function  $U(N_i(t), Y_i(t), \beta, W)$  is obtained by  $\partial L(N_i(t), Y_i(t), \beta, W) / \partial \beta$ , where  $L(\cdot)$  is the log partial likelihood function in (16). By using the accumulated Kullback-Leiber information for partial likelihood functions as in Wong (1986), we can show that the true weight function  $W_0$  maximizes  $E_{H(t)}(L(N_i(t), Y_i(t), \beta, W))$  over the set of weight functions  $W$ , where  $E_{H(t)}(\cdot)$  is the expectation taken with respect to the history  $H(t)$ . This indicates that the criterion of 3.11 in Newey (1994) is satisfied. Therefore, by the Proposition 2 of Newey (1994), the nonparametric estimation of the weight function  $\widehat{W}$  can be ignored in calculating the asymptotic variance of  $\hat{\beta}$ ; that is, the variance estimate will be the same as if  $\widehat{W} = W_0$ .

## References

- Buzney, E. A., & Kimball, A. B. (2008). A critical assessment of composite and coprimary endpoints: A complex problem. *Journal of the American Academy of Dermatology*, 59, 890-896. <http://dx.doi.org/10.1016/j.jaad.2008.05.021>
- Cai, J., Fan, J., Jiang, J., & Zhou, H. (2007). Partially linear hazard regression for multivariate survival data. *Journal of American Statistical Association*, 102, 538-551. <http://dx.doi.org/10.1198/016214506000001374>
- Cai, J., Fan, J., Zhou, H., & Zhou, Y. (2007). Hazard models with varying coefficients for multivariate failure time data. *Annals of Statistics*, 35, 324-354. <http://www.jstor.org/stable/25463557>
- Cai, J., & Prentice, P. L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika*, 82, 151-164. <http://dx.doi.org/10.1093/biomet/82.1.151>
- Cai, J., & Prentice, P. L. (1997). Regression estimation using multivariate failure time data and a common baseline hazard function model. *Lifetime Data Analysis*, 3, 197-213. <http://dx.doi.org/10.1023/A:1009613313677>

- Cai, J., & Schaubel, D. E. (2004). Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Analysis*, 10, 121-138. <http://dx.doi.org/10.1023/B:LIDA.0000030199.23383.45>
- Clegg, L. X., Cai, J., & Sen, P. K. (1999). A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics*, 55, 805-812. <http://dx.doi.org/10.1111/j.0006-341X.1999.00805.x>
- Clegg, L. X., Cai, J., Sen, P. K., & Kupper, L. L. (2000). Misspecification of marginal hazards model in multivariate failure time data. *Sankhya: The Indian Journal of Statistics (Series B)*, 62, 25-42. <http://www.jstor.org/stable/25053118>
- Cook, R. J., Lawless, J. F., Lakhali-Chaieb, L., & Lee, K-A. (2009). Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination: application to skeletal complications in cancer metastatic to bone. *Journal of the American Statistical Association*, 104, 60-75. <http://dx.doi.org/10.1198/jasa.2009.0004>
- Cook, R. J., Lawless, J. F., & Lee, K-A. (2007). A copula-based mixed Poisson model for bivariate recurrent events under event-dependent censoring. *Statistics in Medicine*, 29, 694-707. <http://dx.doi.org/10.1002/sim.3830>
- Dabrowska, D. M. (2006). Multivariate Survival Analysis. In N. Keiding & P. K. Andersen (Eds.), *Survival and Event History Analysis*. Wiley Reference Series in Biostatistics. Hoboken, NJ: John Wiley & Sons.
- Datta, S., & Satten, G. A. (2001). Estimation of integrated transition hazards and stage occupation probabilities from non-Markov systems under dependent censoring. *Biometrics*, 52, 355-363. <http://dx.doi.org/10.1111/j.0006-341X.2002.00792.x>
- Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446), 496-509. <http://dx.doi.org/10.2307/2670170>
- Fleming, T. R., & Lin, D. Y. (2000). Survival analysis in clinical trials: past developments and future directions. *Biometrics*, 56, 971-983. <http://dx.doi.org/10.1111/j.0006-341X.2000.0971.x>
- Freemantle, N., & Calvert, M. (2007). Weighing the pros and cons for composite outcomes in clinical trials. *Journal of Clinical Epidemiology*, 60, 658-659. <http://dx.doi.org/10.1016/j.jclinepi.2006.10.024>
- Freemantle, N., Calvert, M., Wood, J., Eastaugh, J., & Griffin, C. (2003). Composite outcomes in randomized trials: greater precision but with greater uncertainty? *Journal of American Medical Association*, 289, 2554-2559. <http://dx.doi.org/10.1001/jama.289.19.2554>
- Gill, R. D., Keiding, N., & Andersen, P. K. (1993). *Statistical Models Based on Counting Processes*. New York, NY: Springer-Verlag.
- Greene, W. F., & Cai, J. (2004). Measurement error in covariates in the marginal hazards model for multivariate failure time data. *Biometrics*, 60, 987-996. <http://dx.doi.org/10.1111/j.0006-341X.2004.00254.x>
- Hjort, N. L., & Pollard, D. (1993). Asymptotics for minimisers of convex processes. *Technical Report*. Department of Statistics, Yale University, New Haven, CT.
- Hortobagyi, G. N., Theriault, R. L., Lipton, A., Porter, L., Blayney, D., Sinoff, C., ... Reitsma, D. J. (1998). Long-term prevention of skeletal complications of metastatic breast cancer with pamidronate. *Journal of Clinical Oncology*, 16, 2038-2044.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York, NY: Springer Science + Business Media, Inc.
- Kalbfleisch, J. D., & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Hoboken, NJ: John Wiley & Sons.
- Kallenberg, O. (2010). *Foundations of Modern Probability*. New York, NY: Springer.
- Kang, S., & Cai, J. (2009). Marginal Hazards Regression for Retrospective Studies within Cohort with Possibly Correlated Failure Time Data. *Biometrics*, 65, 405-414. <http://dx.doi.org/10.1111/j.1541-0420.2008.01077.x>
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Lee, E. W., Wei, L. J., & Amato, D. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In J. P. Klein, & P. K. Goel (Eds.), *Survival Analysis: State of the Art*,

- (pp. 237-247). Dordrecht, Netherlands: Kluwer Academic Publisher.
- Liang, K. Y., Self, S. G., & Chang, Y. C. (1993). Modelling Marginal Hazards in Multivariate Failure Time Data. *Journal of the Royal Statistical Society, Series B*, 55, 441-453. <http://www.jstor.org/stable/2346205>
- Liang, K-Y, Self, S. G., Bandeen-Roche, K. J., & Zeger, S. L. (1995). Some recent development for regression analysis of multivariate failure time data. *Lifetime Data Analysis*, 1, 403-415. <http://dx.doi.org/10.1007/BF00985452>
- Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22. <http://dx.doi.org/10.1093/biomet/73.1.13>
- Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in Medicine*, 13, 2233-2247. <http://dx.doi.org/10.1002/sim.4780132105>
- Lin, D. Y., Wei, L. J., Yang, I., & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of Royal Statistical Society (Series B)*, 62, 711-730. <http://dx.doi.org/10.1111/1467-9868.00259>
- Nelsen, R. B. (2006). *An Introduction to Copulas*. New York, NY: Springer-Verlag.
- Newey, W. K. (1994). The Asymptotic Variance of Semiparametric Estimators. *Econometrica*, 62, 1394-1382. <http://dx.doi.org/10.2307/2951752>
- Robins, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate marker. *Proceedings of the Biopharmaceutical Section, American Statistical Association*, 24-33.
- Robins, J. M., & Finkelstein, D. M. (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*, 56, 779-788. <http://dx.doi.org/10.1111/j.0006-341X.2000.00779.x>
- Rotnitzky, A., & Robins, J. M. (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika*, 82, 805-820. <http://dx.doi.org/10.1093/biomet/82.4.805>
- Shorack, G. R., & Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. New York, NY: Wiley.
- Spiekerman, C. F., & Lin, D. Y. (1998). Marginal Regression Models for Multivariate Failure Time Data. *Journal of American Statistical Association*, 93, 1164-1175. <http://dx.doi.org/10.1080/01621459.1998.10473777>
- Therneau, T., & Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer-Verlag.
- Van der Vaart, A. W., & Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York, NY: Springer.
- Wei, L. J., & Glidden, D. V. (1997). An overview of statistical methods for multiple failure time data in clinical trials. *Statistics in Medicine*, 16, 833-839. [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19970430\)16:8<833::AID-SIM538>3.0.CO;2-2](http://dx.doi.org/10.1002/(SICI)1097-0258(19970430)16:8<833::AID-SIM538>3.0.CO;2-2)
- Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *Journal of American Statistical Association*, 84, 1065-1073. <http://dx.doi.org/10.1080/01621459.1989.10478873>
- Wong, W. H. (1986). Theory of Partial Likelihood. *Annual of Statistics*, 14, 88-123. Retrieved from <http://www.jstor.org/stable/2241269>
- Yin, G., & Cai, J. (2004). Additive hazards model with multivariate failure time data. *Biometrika*, 91, 801-818. <http://dx.doi.org/10.1093/biomet/91.4.801>
- Yin, G., & Cai, J. (2005). Quantile regression models with multivariate failure time data. *Biometrics*, 61, 151-161. <http://dx.doi.org/10.1111/j.0006-341X.2005.030815.x>

### **Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).