Published in final edited form as:

Lifetime Data Anal. 2014 October ; 20(4): 563–583. doi:10.1007/s10985-013-9288-y.

Joint Modeling Approach for Semicompeting Risks Data with Missing Nonterminal Event Status

Chen Hu and

Radiation Therapy Oncology Group (RTOG) Statistical Center, 1818 Market Street, Suite 1600
Philadelphia, PA 19103

Alex Tsodikov

Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor,
Michigan 48109-2029, U.S.A

Chen Hu: chu@acr.org; Alex Tsodikov: tsodikov@umich.edu

Abstract

Semicompeting risks data, where a subject may experience sequential non-terminal and terminal events, and the terminal event may censor the non-terminal event but not vice versa, are widely available in many biomedical studies. We consider the situation when a proportion of subjects' non-terminal events is missing, such that the observed data become a mixture of "true" semicompeting risks data and partially observed terminal event only data. An illness-death multistate model with proportional hazards assumptions is proposed to study the relationship between non-terminal and terminal events, and provide covariate-specific global and local association measures. Maximum likelihood estimation based on semiparametric regression analysis is used for statistical inference, and asymptotic properties of proposed estimators are studied using empirical process and martingale arguments. We illustrate the proposed method with simulation studies and data analysis of a follicular cell lymphoma study.

Keywords

Semicompeting risks data; Multistate model; Missing non-terminal event; Proportional hazards; Dependent censoring; Survival analysis

1 Introduction

In many biomedical studies, patients may experience multiple distinct causes of failures. With competing risks data, the first occurrence of any cause censors the occurrences of all other causes. When only one of the two failure causes (terminal) can censor the other but not vice versa, the risks are called semicompeting (Fine et al, 2001). Semicompeting risks data is more informative than competing risks data. An example of semicompeting risks data involves time to disease progression (non-terminal event) and mortality (terminal event).

When non-terminal event can be missing, different underlying models may lead to seemingly identical competing risks observations. The observation of terminal event such as death is usually well defined and objective. However, absence of non-terminal event may be interpreted as right censoring when the event has not occurred yet, or left censoring when

the non-terminal event occurred but was not registered. Further, upon absence of the non-terminal event it may be unknown which of the two possibilities takes place. In cancer, recurrence or progression and death often represent semicompeting risks data. Here disease progression may be established through a series of clinical and/or radiologic measurements. Observation of such non-terminal events may be missing (unobserved) dependent on what procedures are performed and how they are registered in the database. As a result, an unknown proportion of subjects may have their non-terminal events occurred but unobserved, such that the observed semicompeting risks data become a mixture of “true” semicompeting risks data and terminal event only data with unknown non-terminal event status. This situation may occur by design when a mortality database that does not have recurrence information is combined with a clinical dataset where such information is partially or fully available.

Because non-terminal and terminal events are usually driven by the same disease and thus correlated, censoring of the non-terminal event by the terminal event is informative. Ignoring the dependent censoring structure or the association may lead to biased estimation or efficiency loss (Zheng and Klein, 1995; Huang and Zhang, 2008).

Understanding the association between the events is interesting in itself as an important component of the disease natural history useful in prognostic and therapeutic studies. Learning about the disease natural history through an observation of a non-terminal event and adapting treatment to this knowledge is an interesting problem.

In statistical literature, semicompeting risks have been addressed in a semiparametric instead of a fully nonparametric framework due to the non-identifiability issue (Peng and Fine, 2007). A copula model for the two event times has been repeatedly considered under a restriction on the upper wedge $T_1^* \leq T_2^*$ of the positive quadrant, where T_1^*, T_2^* are potential times to non-terminal and terminal events, respectively. In the one-sample setting, Fine et al (2001) proposed a semiparametric estimation approach based on Clayton copula (Clayton, 1978), and Wang (2003) extended this framework to more general copula models. This approach has also been extended to the regression setting by postulating separate marginal regression models for non-terminal and terminal events (Peng and Fine, 2007; Hsieh et al, 2008; Chen, 2012). It is noted that all of these models proceed from latent failure times and are similar in spirit to the competing risks models in the same class, e.g., Fine and Gray (1999). Therefore, the interpretation of the marginal distribution of the non-terminal event in this framework is hypothetical.

Alternatively, multistate illness-death models (Kalbfleisch and Prentice, 2002; Hougaard, 2000), Figure 1, can be adopted. Associations can be incorporated through effects on transition intensity functions. Let $\lambda_{ij}(t/t_i)$ be the intensity of transition from state i to state j , where t_i is the time spent in state i . Markov or semi-Markov illness-death models represent specific assumption on the form of λ_{ij} . A variety of proportional hazards (PH) models have been considered by Andersen et al (1991) and Kalbfleisch and Prentice (2002)(p.270) (Markov models); Andersen et al (2000) and Shu et al (2007) (semi-Markov models).

In this paper we extend the PH framework to incorporate a possibility that non-terminal event is unobserved either because it is censored or because the observation is missing, and it is generally unknown which of the two possibilities takes place.

A shared-frailty effect (Nielsen et al, 1992; Govindarajulu et al, 2011) has been a popular tool to induce positive dependence. Xu et al (2010) incorporates gamma frailty into the illness-death model to model semicompeting risks data. However, conditional specification of the model where non-terminal event affects the risk for the terminal event may be more attractive when the dependence cannot be assumed positive. Specifically we use an idea similar to so-called Markov modulated models based on intensity processes (e.g., (Kalbfleisch and Prentice, 2002) Section 9.2, (Cook and Lawless, 2007) Section 5.3) used with recurrent data analysis. This leads us to an approach similar to transformation models in survival analysis (Chen et al, 2002; Tsodikov, 2003; Chen, 2009).

Partial-likelihood-based approaches conditioning on the past history become inappropriate when the past history is not fully observable (Andersen and Keiding, 2002). Therefore our paper is based on a full likelihood approach.

The paper is organized as follows. In Section 2, we describe the illness-death model and its properties in terms of bivariate survival distributions, the data structure to be used throughout the paper, as well as the likelihood construction when missing non-terminal event is possible. The maximum likelihood estimation procedure and asymptotic properties of the estimators for regression coefficients are discussed in Section 3 and 4. We present Monte Carlo simulation studies and a data analysis of follicular cell lymphoma patients in Section 5, and conclude the article with a discussion in Section 6.

2 Model and Data Structure

2.1 Notation and Model

Let T_1^* and T_2^* be the potential times to non-terminal and terminal events, respectively; $Z(t)$ be the covariate vector of dimension d ; $N_1^*(t) = I(T_1^* < t)$ be the counting processes for non-terminal event. Throughout the paper, suppressing subscripts, we denote S , f and λ as survival, density and hazard functions, respectively. We define the joint distribution of T_1^* and T_2^* through the marginal and conditional hazard functions, respectively:

$$\lambda_1(t|Z(t)) = \lim_{\Delta \rightarrow 0} P(T_1^* \in [t, t+\Delta] | T_1^* \geq t, T_2^* \geq t, Z(t)) / \Delta = h_0(t)\eta(Z(t)), \quad (1)$$

$$\lambda_2(t|Z(t), N_1^*(t)) = \lim_{\Delta \rightarrow 0} P(T_2^* \in [t, t+\Delta] | T_2^* \geq t, Z(t), N_1^*(t)) / \Delta = h_0^*(t)\theta(Z(t))\mu(Z(t))^{N_1^*(t)}, \quad (2)$$

where $\eta(Z(t)) = e^{\beta\eta(Z(t))}$, $\theta(Z(t)) = e^{\beta\theta(Z(t))}$, $\mu(Z(t)) = e^{\beta\mu(Z(t))}$, and h_0, h_0^* are the baseline hazard functions. For brevity, we consider time-independent covariate Z from now on, and keep in mind that the proposed model may accommodate external time-dependent $Z(t)$ as well.

The proposed specification can be viewed as a combination of two commonly used models: 1) a Cox proportional hazard (PH) model for T_1^* ; and 2) a Cox PH model for T_2^* with the counting process for T_1^* as a time-dependent covariate. Alternatively, one may view the proposed model in terms of an illness-death process (Figure 1), where λ_1 is the transition hazard between state 0 and 1 (λ_{01}), λ_2 is the path transition hazard between state 0 and 2 which may depend on whether state 1 occurs (λ_{12}) or not (λ_{02}). Furthermore, $\mu(Z)$ may be interpreted as the hazard ratio between λ_{02} and λ_{12} , which defines the dependence between non-terminal and terminal events, as discussed in Section 1. $\mu(Z) = 1$ corresponds to no dependence between T_1^* and T_2^* ; $\mu(Z) > 1$ or $\mu(Z) < 1$ represent positive or negative dependence between T_1^* and T_2^* the occurrence of T_1^* accelerating or decelerating the occurrence of T_2^* .

The model (1) and (2) defines a bivariate distribution for (T_1^*, T_2^*) on the positive quadrant of $(t_1 \geq 0, t_2 \geq 0)$. For semicompeting risks (observed) data, observations are restricted to the upper wedge $0 \leq t_1 \leq t_2$ and are positively correlated even if the potential times are not. Following Xu et al (2010), define $T_1^* = \infty$ if terminal event occurs before non-terminal event, such that there is no probability mass in the lower wedge $t_2 < t_1 < \infty$. The probability model for (T_1^*, T_2^*) is taken to be absolutely continuous with joint density $f_1(t_1, t_2)$ defined in the upper wedge $0 \leq t_1 \leq t_2$, and continuous along the line at $t_1 = \infty$ with density $f_\infty(t_2)$, $t_2 > 0$, such that $\int_0^\infty \int_{t_1}^\infty f_1(t_1, t_2) dt_2 dt_1 + \int_0^\infty f_\infty(t_2) dt_2 = 1$. This definition provides validity for the marginal distribution of T_1^* , which is not well defined in semicompeting risks data. We further denote S_1 and S_∞ as the survival functions correspond to f_1 and f_∞ , respectively.

In the following we assume proportionality relationship between T_1^* and T_2^* , i.e., $h_0 = h_0^* = h$. In cancer studies this assumption can be justified through a common tumor growth process driving the risks of observed events proportional to a common baseline hazard serving as a surrogate of the growth pattern. We will discuss approaches to relax this assumption in Section 6. By denoting $\eta = \eta(Z)$, $\theta = \theta(Z)$, $\mu = \mu(Z)$, $\bar{\mu} = 1 - \mu$, $H(t) = \int_0^t h(x) dx$, and

$D_k = -\frac{\partial}{\partial t_k}$ for $k = 1, 2$, we can subsequently derive the joint probability density functions (f) and survival functions (S) of T_1^* and T_2^* on the upper wedge $0 \leq t_1 \leq t_2$ as

$$\begin{aligned} S_1(t_1, t_2) &= S(t_1, t_2, t_1 \leq t_2) = \frac{\eta\theta\bar{\mu}}{\eta+\theta\bar{\mu}} \left\{ \frac{e^{-H(t_1)(\eta+\theta\bar{\mu})-H(t_2)\theta\mu}}{\theta\mu} - \frac{e^{-H(t_2)(\eta+\theta)}}{\eta+\theta} \right\}, \\ f_1(t_1, t_2) &= D_1 D_2 S_1(t_1, t_2) = h(t_1)h(t_2)\eta\theta\bar{\mu}e^{-H(t_1)(\eta+\theta\bar{\mu})-H(t_2)\theta\mu}, \\ D_1 S_1(t_1, t_2) &= h(t_1)\eta e^{-H(t_1)(\eta+\theta\bar{\mu})-H(t_2)\theta\mu}, \\ D_2 S_1(t_1, t_2) &= \frac{h(t_2)\eta\theta\bar{\mu}}{\eta+\theta\bar{\mu}} \left\{ e^{-H(t_1)(\eta+\theta\bar{\mu})-H(t_2)\theta\mu} - e^{-H(t_2)(\eta+\theta)} \right\}. \end{aligned}$$

These model quantities are derived by integrating the densities presented in Supplemental Materials Section A. We note that the integration goes over H as an argument due to the proportionality assumption. When the subject fails before the non-terminal event occurs, i.e., $t_2 < t_1$ (t_1 redefined as $= \infty$ in this case) the probability density and survival functions are

$$f_{\infty}(t_2)=h(t_2)\theta e^{-H(t_2)(\theta+\eta)}, \quad S_{\infty}(t_2)=\frac{\theta}{\theta+\eta}e^{-H(t_2)(\theta+\eta)}.$$

Kendall's τ is a popular global measure of dependence between bivariate survival times. One can show that the Kendall's τ on the upper wedge $0 \leq t_1 \leq t_2$ is given by:

$$\tau=4 \times \int_0^{\infty} \int_0^y S_1(x, y) f_1(x, y) dx dy - 1 = \frac{\theta \mu}{\eta + \theta + \theta \mu}.$$

Remarkably, this measure does not depend on the baseline hazard function, a consequence of the proportionality assumption.

The crossratio function (Clayton, 1978; Oakes, 1989) is commonly used to measure the local dependence between bivariate survival times. For the upper wedge of $0 \leq t_1 \leq t_2$, the crossratio function $\gamma(t_1, t_2)$ is as follows:

$$\begin{aligned} \gamma(t_1, t_2) &= \frac{S_1(t_1, t_2) f_1(t_1, t_2)}{\{D_1 S_1(t_1, t_2)\} \{D_2 S_1(t_1, t_2)\}} = \frac{\lambda_2(t_2 | T_1 = t_1)}{\lambda_2(t_2 | T_1 > t_1)} = \frac{\lambda_1(t_1 | T_2 = t_2)}{\lambda_1(t_1 | T_2 > t_2)} \\ &= 1 + \frac{\eta + \theta \mu}{\eta + \theta} \frac{1}{e^{(H(t_2) - H(t_1))(\eta + \theta \mu)} - 1}. \end{aligned}$$

We notice that $\gamma(t_1, t_2)$ cannot be written in terms of $S(t_1, t_2)$ alone, and thus the corresponding copula does not belong to the Archimedean class (Oakes, 1989).

2.2 Observed Data Structure and Counting Process Notation

2.2.1 Conventional Semicompeting Risks Data—Let V^* be the censoring time independent of (T_1^*, T_2^*) , given Z ; ξ be the maximum follow-up time in the study. Suppose $(T_{1i}^*, T_{2i}^*, V_i^*, Z_i)$, $i = 1, \dots, n$ are n independent and identically distributed replicates of (T_1^*, T_2^*, V^*, Z) , such that the observed semicompeting risks data for subject i , $i = 1, \dots, n$, consists of $\{T_{1i}^*, \tilde{T}_i, \tilde{\delta}_{ki}^*, \tilde{\delta}_{2i}^*, Z_i; k = 1, 2, 0 \leq t \leq \xi\}$, where $T_i^* = \min(T_{1i}^*, T_{2i}^*, V_i)$, $\tilde{T}_i = \min(T_{2i}^*, V_i)$ with $V_i = \min(V_i^*, \xi)$, $\delta_{ki}^* = I(T_i = T_{ki}^*)$, $k = 1, 2$ are competing risks cause of failure indicators, $\tilde{\delta}_{2i}^* = \delta_{1i}^* I(\tilde{T}_i = T_{2i}^*)$ is an indicator of terminal event observed after the non-terminal one, and where $I(\cdot)$ is the indicator function. Because $0 \leq T_i \leq \tilde{T}_i$, the observations are restricted to the upper wedge. Furthermore, if $\delta_{1i}^* = 0, \delta_{2i}^* = 1, \tilde{\delta}_{2i}^* = 0$, then $T_i = \tilde{T}_i = T_{2i}^*$ and $T_{1i}^* = \infty$ by definition.

2.2.2 Semicompeting Risks with Missing Non-terminal Events—When non-terminal event can be unobservable (missing), the observed semi-competing risks data become a mixture of subjects whose disease history is completely or partially observed. Throughout the paper, we distinguish the observed status of one's non-terminal event from the missing status indicator showing whether one's non-terminal event status *can* be observed. The former is an observed event status (i.e., δ_1^*); the latter is assumed to be a binary unobserved missing data indicator variable following a missing at random (MAR) mechanism (Little and Rubin, 2002). Let $M_i = 1$ if the non-terminal event is unobservable,

and $M_i = 0$ if it is observable. It is natural to assume $\Pr(M_i=1)=p=\frac{\exp(\beta_p)}{1+\exp(\beta_p)}$, where p is following a logistic model (or other link functions as applicable). Covariates can be easily added to this formulation. The complete data is given by $(T_{1i}^*, T_{2i}^*, V_i^*, M_i, Z_i)$, $i = 1, \dots, n$ independent and identically distributed replicates of $(T_1^*, T_2^*, V^*, M, Z)$. The observed semicompeting risks data in presence of missing non-terminal event are similar $\{T_i, \tilde{\delta}_{ki}, \tilde{\delta}_{2i}, Z_i; k = 1, 2, 0 \leq t \leq \xi\}$, where $\tilde{T}_i = \min(T_{2i}^*, V_i)$, $\delta_{2i} = I(T_i = T_{2i}^*)$, but $T_i^*, \tilde{\delta}_{1i}^*, \delta_{2i}^*$ are modified into

$T_i = \min(T_{1i}^*, T_{2i}^*, V_i)I(M_i=0) + \min(T_{2i}^*, V_i)I(M_i=1)$, $\delta_{1i} = I(T_i = T_{1i}^*)I(M_i=0)$, and $\delta_{2i} = \delta_{1i}I(\tilde{T}_i = T_{2i}^*)$. In other words, observing non-terminal event ($\delta_{1i} = 1$) implies $M_i = 0$; not observing non-terminal event ($\delta_{1i} = 0$) indicates either (i) non-terminal event has occurred but is unobservable ($M_i = 1$), or (ii) non-terminal event has not occurred yet.

In counting process notation, for subject i , let $N_{ki}^* = I(T_{ki}^* \leq t)$, $k = 1, 2$, be the underlying counting processes for non-terminal ($k = 1$) and terminal ($k = 2$) events, $Y_i(t) = I(T_i \geq t)$ be the at-risk process for the first-occurring event regardless of event type, $\tilde{Y}_i(t) = I(T_i \geq t > T_i)$ be the at-risk process for the terminal event occurring after the observed non-terminal one, and let $N_{1i}(t) = \delta_{1i}N_{1i}^*(t)$ and $N_{2i}(t) = \delta_{2i}N_{2i}^*(t)$ be the observed counting processes of the first-occurring non-terminal and terminal events, respectively, and $\tilde{N}_{2i}(t) = \int_0^t \tilde{Y}_i(x) dN_{2i}^*(x) = \tilde{\delta}_{2i}N_{2i}^*(t)$ be the observed counting process of terminal event occurring after non-terminal event. In addition, we assume $\Pr(T_1^* = T_2^*) = 0$.

The observed semicompeting risks data are thus expressed through the three counting processes $N_{1i}(\cdot), N_{2i}(\cdot), \tilde{N}_{2i}(\cdot)$, and two corresponding at-risk processes $Y_i(\cdot), \tilde{Y}_i(\cdot)$ for subject i . By definition, $N_{1i}(\cdot) + N_{2i}(\cdot) = 1$, and $\tilde{N}_{2i}(\cdot) = 0$ if $\delta_{2i} = 1$ or $M_i = 1$.

2.3 Likelihood with Missing non-terminal Event Status

Following Section 2, the likelihood contribution of subject i will take one of the following four forms (where the $\Pr$ notation is informal and context dependent, and is understood as a density where appropriate):

- If both non-terminal and terminal events are observed sequentially, i.e., $(T_i < T_i, \tilde{\delta}_{1i} = 1, \delta_{2i} = 0, \tilde{\delta}_{2i} = 1)$:

$$\begin{aligned}\mathcal{L}_{1i}^* &= \Pr(T_{1i}^* = T_i, T_{2i}^* = \tilde{T}_i, M_i = 0) \\ &= \Pr(M_i = 0) \Pr(T_{1i}^* = T_i, T_{2i}^* = \tilde{T}_i | M_i = 0) \\ &= (1-p) \times f_1(T_i, \tilde{T}_i) = (1-p) \times h(T_i) h(\tilde{T}_i) \eta_i \theta_i \mu_i e^{-H(T_i)(\eta_i + \theta_i \bar{\mu}_i) - H(\tilde{T}_i) \theta_i \mu_i}.\end{aligned}$$

- If non-terminal event is observed, but terminal event is censored, i.e., $(T_i < T_i, \tilde{\delta}_{1i} = 1, \delta_{2i} = 0, \tilde{\delta}_{2i} = 0)$:

$$\begin{aligned}\mathcal{L}_{2i}^* &= \Pr(T_{1i}^* = T_i, T_{2i}^* \geq \tilde{T}_i, M_i = 0) \\ &= \Pr(M_i = 0) \Pr(T_{1i}^* = T_i, T_{2i}^* \geq \tilde{T}_i | M_i = 0) \\ &= (1-p) \times D_1 S_1(t_1, t_2) |_{(T_i, \tilde{T}_i)} = (1-p) \times h(T_i) \eta_i e^{-H(T_i)(\eta_i + \theta_i \bar{\mu}_i) - H(\tilde{T}_i) \theta_i \mu_i}.\end{aligned}$$

- If non-terminal event is not observed, and terminal event is observed, i.e., $(T_i = T_i^*, \delta_{1i} = 0, \delta_{2i} = 1, \delta_{2i}^* = 0)$:

$$\begin{aligned}\mathcal{L}_{3i} &= \Pr(T_{1i}^* < T_i, T_{2i}^* = T_i, M_i = 1) + \Pr(T_{1i}^* > T_i, T_{2i}^* = T_i) \\ &= p \times D_2 S_1(t_1, t_2)|_{(0, T_i)} + f_\infty(T_i) \\ &= p \times h(T_i) \frac{\eta_i \theta_i \mu_i}{\eta_i + \theta_i \mu_i} \{e^{-H(T_i) \theta_i \mu_i} - e^{-H(T_i) (\eta_i + \theta_i)}\} + h(T_i) \theta_i e^{-H(T_i) (\eta_i + \theta_i)}\end{aligned}$$

- If neither event is observed, i.e., $(T_i = T_i^*, \delta_{1i} = 0, \delta_{2i} = 0, \delta_{2i}^* = 0)$:

$$\begin{aligned}\mathcal{L}_{4i} &= \Pr(T_{1i}^* \leq T_i, T_{2i}^* \geq T_i, M_i = 1) + \Pr(T_{1i}^* \geq T_i, T_{2i}^* \geq T_i) \\ &= p \times \frac{\eta_i}{\eta_i + \theta_i \mu_i} \{e^{-H(T_i) \theta_i \mu_i} - e^{-H(T_i) (\eta_i + \theta_i)}\} + e^{-H(T_i) (\eta_i + \theta_i)}.\end{aligned}$$

Note that we can alternatively express $\mathcal{L}_{1i}^*$ and $\mathcal{L}_{2i}^*$ as

$$\begin{aligned}\mathcal{L}_{1i}^* &= \Pr(T_{1i}^* = T_i, T_{2i}^* = \tilde{T}_i, M_i = 0) = \mathcal{L}_{1i} \times \tilde{\mathcal{L}}_{1i}, \\ \mathcal{L}_{2i}^* &= \Pr(T_{1i}^* = T_i, T_{2i}^* \geq \tilde{T}_i, M_i = 0) = \mathcal{L}_{1i} \times \tilde{\mathcal{L}}_{2i},\end{aligned}$$

where

$$\mathcal{L}_{1i} = \Pr(M_i = 0) \Pr(T_{1i}^* = T_i | M_i = 0) = (1 - p) \times h(T_i) \eta_i e^{-H(T_i) (\eta_i + \theta_i)},$$

$$\tilde{\mathcal{L}}_{1i} = \Pr(T_{2i}^* = \tilde{T}_i | T_{1i}^* = T_i, M_i = 0) = h(\tilde{T}_i) \theta_i \mu_i e^{-[H(\tilde{T}_i) - H(T_i)] \theta_i \mu_i},$$

$$\tilde{\mathcal{L}}_{2i} = \Pr(T_{2i}^* = \tilde{T}_i | T_{1i}^* = T_i, M_i = 0) = e^{-[H(\tilde{T}_i) - H(T_i)] \theta_i \mu_i},$$

such that the contribution from non-terminal event is separated from the terminal event. Therefore, the contribution of subject i in the log-likelihood is

$$\begin{aligned}\ell_i &= \delta_{1i} \tilde{\delta}_{2i} \log(\mathcal{L}_{1i}^*) + \delta_{1i} (1 - \tilde{\delta}_{2i}) \log(\mathcal{L}_{2i}^*) + \delta_{2i} \log(\mathcal{L}_{3i}) + (1 - \delta_{1i} - \delta_{2i}) \log(\mathcal{L}_{4i}) \\ &= \{\delta_{1i} \log(\mathcal{L}_{1i}) + \delta_{2i} \log(\mathcal{L}_{3i}) + (1 - \delta_{1i} - \delta_{2i}) \log(\mathcal{L}_{4i})\} + \delta_{1i} \{\tilde{\delta}_{2i} \log(\tilde{\mathcal{L}}_{1i}) + (1 - \tilde{\delta}_{2i}) \log(\tilde{\mathcal{L}}_{2i})\}.\end{aligned}\quad (3)$$

Using notation $\ell_{1i} = \delta_{1i} \log(\mathcal{L}_{1i}) + \delta_{2i} \log(\mathcal{L}_{3i}) + (1 - \delta_{1i} - \delta_{2i}) \log(\mathcal{L}_{4i})$, $\ell_{2i} = \delta_{1i} \{\tilde{\delta}_{2i} \log(\tilde{\mathcal{L}}_{1i}) + (1 - \tilde{\delta}_{2i}) \log(\tilde{\mathcal{L}}_{2i})\}$, and $\ell_1 = \sum_{i=1}^n \ell_{1i}$, $\ell_2 = \sum_{i=1}^n \ell_{2i}$, we can write the full log-likelihood as

$$\ell = \sum_{i=1}^n \ell_i = \sum_{i=1}^n \{\ell_{1i} + \ell_{2i}\} = \ell_1 + \ell_2.$$

Such likelihood partition can be easily recognized from the illness-death model formulation, where ℓ_1 corresponds to the contributions from first-occurring events (a competing risks type likelihood), and ℓ_2 corresponds to the additional likelihood contributions from the terminal events if the preceding non-terminal events are observed.

3 Nonparametric Maximum Likelihood Estimation

3.1 Martingale Theory

Let us denote $\beta = (\beta_{\eta}, \beta_{\theta}, \beta_{\mu}, \beta_p)$ and the full parameter set $\Omega = (\beta, H(\cdot))$, where β is finite-dimensional and $H(\cdot)$ is the infinite-dimensional parameter. Using the NPMLE approach, we treat $H(\cdot)$ as a nondecreasing step function with jumps dH only at the time points where events are observed (Tsodikov, 2003; Zeng and Lin, 2006; Chen, 2009). Therefore $\{dH\}$ is viewed as the collection of the jump sizes of H at the observed event times.

The following martingales (at the true model) can be constructed based on observed counting processes with respect to filtration

$$\mathcal{F}(t-) = \sigma\{N_{1i}(x), N_{2i}(x), \tilde{N}_{2i}(x), Y_i(x), \tilde{Y}_i(x), Z_i: x \in [0, t], i=1, \dots, n\},$$

$$\begin{aligned} dM_{ki}(t) &= dN_{ki}(t) - Y_i(t) \Theta_{ki}(H(t), \beta) dH(t), k=1, 2, \\ d\tilde{M}_{2i}(t) &= d\tilde{N}_{2i}(t) - \tilde{Y}_i(t) \tilde{\Theta}_{2i}(\beta) dH(t), \end{aligned}$$

where

$$\begin{aligned} \Theta_{1i}(H(t); \beta) &= \frac{(1-p) \times \eta_i e^{-H(t)} (\eta_i + \theta_i)}{p \times \frac{\eta_i}{\eta_i + \theta_i \mu_i} \{e^{-H(t) \theta_i \mu_i} - e^{-H(t) (\eta_i + \theta_i)}\} + e^{-H(t) (\eta_i + \theta_i)}}, \\ \Theta_{2i}(H(t); \beta) &= \frac{p \times \frac{\eta_i \theta_i \mu_i}{\eta_i + \theta_i \mu_i} \{e^{-H(t) \theta_i \mu_i} - e^{-H(t) (\eta_i + \theta_i)}\} + e^{-H(t) (\eta_i + \theta_i)}}{p \times \frac{\eta_i}{\eta_i + \theta_i \mu_i} \{e^{-H(t) \theta_i \mu_i} - e^{-H(t) (\eta_i + \theta_i)}\} + e^{-H(t) (\eta_i + \theta_i)}}, \\ \tilde{\Theta}_{2i}(\beta) &= \theta_i \mu_i. \end{aligned}$$

$\Theta_{ki}(H(t); \beta)$, $k = 1, 2$ and $\tilde{\Theta}_{2i}(\beta)$ can be derived through the following probabilistic argument,

$$\begin{aligned} E\{dN_{ki}(t) | \mathcal{F}_i(t-)\} &= Y_i(t) \Pr(dN_{ki}(t)=1 | Y_i(t)=1) = Y_i(t) \Theta_{ki}(H(t); \beta) dH(t), \\ E\{d\tilde{N}_{2i}(t) | \mathcal{F}_i(t-)\} &= \tilde{Y}_i(t) \Pr(d\tilde{N}_{2i}(t)=1 | \tilde{Y}_i(t)=1) = \tilde{Y}_i(t) \tilde{\Theta}_{2i}(\beta) dH(t), \end{aligned}$$

where $k = 1, 2$, and

$$\begin{aligned} \Pr(dN_{1i}(t)=1 | Y_i(t)=1) &= \frac{\Pr(T_{1i}^* = t, T_{2i}^* \geq t, M_i=0)}{\Pr(T_{1i}^* \leq t, T_{2i}^* \geq t, M_i=1) + \Pr(T_{1i}^* \geq t, T_{2i}^* \geq t)}, \\ \Pr(dN_{2i}(t)=1 | Y_i(t)=1) &= \frac{\Pr(T_{1i}^* \leq t, T_{2i}^* = t, M_i=1) + \Pr(T_{1i}^* > t, T_{2i}^* = t)}{\Pr(T_{1i}^* \leq t, T_{2i}^* \geq t, M_i=1) + \Pr(T_{1i}^* \geq t, T_{2i}^* \geq t)}, \\ \Pr(d\tilde{N}_{2i}(t)=1 | \tilde{Y}_i(t)=1) &= \frac{\Pr(T_{1i}^* = t^*, T_{2i}^* = t, M_i=0)}{\Pr(T_{1i}^* = t^*, T_{2i}^* \geq t, M_i=0)}. \end{aligned}$$

3.2 Score Function and Estimating Equation

For any functional J that depends on the path $\bar{H}(x)$ of the function $H(y)$, $y \in [0, x]$, define a derivative as

$$\frac{\partial J(\bar{H}(x))}{\partial dH(t)} = \frac{dJ(\bar{H}(x) + a \times f)}{da} \Big|_{a=0, f=1(y-t)},$$

where $1(x) = 1$ when $x \geq 0$ and 0 otherwise, and the function $f(y) = 1(y - t)$ serves as a direction of perturbation of the argument of J . For example, using this definition

$\frac{\partial H(x)}{\partial dH(t)} = 1(x - t)$, $\frac{\partial dH(x)}{\partial dH(t)} = d1(x - t) = \delta_D(x - t)dx$, where δ_D is a Dirac's delta function representing a limit of normal density when variance approaches zero. Suppose now that $J(\bar{H}(\tau)) = \int_0^\tau \varphi(H(x))dH(x)$, where $\varphi(y)$ is a differentiable function of its argument. Differentiating J as a composite functional gives

$$\frac{\partial J(\bar{H}(\tau))}{\partial dH(t)} = I(t < \tau) \int_t^\tau \varphi'(H(x))dH(x) + \varphi(H(t)).$$

Also this definition of the functional derivative corresponds to taking a derivative with respect to a jump of H at time t if H is a step-function.

Denote the functional derivative with respect to $\{dH(t)\}$ and β of $\Theta_{ki}(H(t); \beta)$ and $\tilde{\Theta}_{2i}(\beta)$ as:

$$\Theta_{ki,H}(H(t); \beta) = \partial \Theta_{ki}(H(t); \beta) / \partial dH(t), \\ \Theta_{ki,\beta}(H(t); \beta) = \partial \Theta_{ki}(H(t); \beta) / \partial \beta, \quad \tilde{\Theta}_{2i,\beta}(\beta) = d\tilde{\Theta}_{2i}(\beta) / d\beta$$

In order to link the martingales $M_{ki}(t)$ and $M_{2i}(t)$ with the likelihood, we note that the denominator of $\Theta_{ki}(H(t); \beta)$, $k = 1, 2$ represents the probability of neither of the events occurring by the time t in the presence of the missing data mechanism (compare to the contribution $\mathcal{L}_{4i}$). As this is a survival function, we can express it through the respective cumulative hazard as

$$p \times \frac{\eta_i}{\eta_i + \theta_i \bar{\mu}_i} \{e^{-H(t)\theta_i \mu_i} - e^{-H(t)(\eta_i + \theta_i)}\} + e^{-H(t)(\eta_i + \theta_i)} = \exp \left\{ - \int_0^t \sum_{k=1}^2 \Theta_{ki}(H(x); \beta) dH(x) \right\}.$$

Applying this expression to the terms of the log-likelihood $\ell = \ell_1 + \ell_2$ in Section 2.3, and moving to the counting process notation, we get the compact counting process forms

$$\ell_1 = \sum_{i=1}^n \sum_{k=1}^2 \left\{ \int_0^\xi [\log dH(x) + \log \Theta_{ki}(H(x); \beta)] dN_{ki}(x) - Y_i(x) \Theta_{ki}(H(x); \beta) dH(x) \right\} \quad (4)$$

$$\ell_2 = \sum_{i=1}^n \left\{ \int_0^\xi [\log dH(x) + \log \tilde{\Theta}_{2i}(\beta)] d\tilde{N}_{2i}(x) - \tilde{Y}_i(x) \tilde{\Theta}_{2i}(\beta) dH(x) \right\}, \quad (5)$$

By taking derivative of the re-expressed log-likelihood with respect to $\{dH\}$ using the differentiation rules defined above, we obtain the following score function for $H(t)$:

$$U_{H(t)} = \sum_{i=1}^n \int_0^t \left\{ \sum_{k=1}^2 [dN_{ki}(x) - Y_i(x) \omega_{ki}(\bar{H}(\xi), x; \beta) \Theta_{ki}(H(x); \beta) dH(x)] + d\tilde{N}_{2i}(x) - \tilde{Y}_i(x) \tilde{\Theta}_{2i}(\beta) dH(x) \right\}, \quad (6)$$

where

$$\omega_{ki}(\bar{H}(\xi), t; \beta) = 1 - \frac{\int_{t+}^\xi \psi_{ki}(H(x); \beta) dM_{ki}(x)}{\Theta_{ki}(H(t); \beta)}; \psi_{ki}(H(t); \beta) = \frac{\partial \log \Theta_{ki}(H(t); \beta)}{\partial dH(y)} = \frac{\Theta_{ki,H}(H(t); \beta)}{\Theta_{ki}(H(t); \beta)}, y < t.$$

We note that ω is a functional of H that invokes its full path $H(\bar{\xi})$. The score function for β is

$$U_\beta = \sum_{i=1}^n \int_0^\xi \left\{ \sum_{k=1}^2 \left[\frac{\Theta_{ki,\beta}(H(x); \beta)}{\Theta_{ki}(H(x); \beta)} dN_{ki}(x) - Y_i(x) \Theta_{ki,\beta}(H(x); \beta) dH(x) \right] \frac{\tilde{\Theta}_{2i,\beta}(\beta)}{\tilde{\Theta}_{2i}(\beta)} d\tilde{N}_{2i}(x) - \tilde{Y}_i(x) \tilde{\Theta}_{2i,\beta}(\beta) dH(x) \right\}. \quad (7)$$

Solutions to the above estimating equations (6) and (7) set to zero, gives an NPMLE of $\{d\hat{H}\}$ and $\hat{\beta}$.

The NPMLE is obtained by maximizing the log-likelihood ℓ or equivalently, solving the system of score functions (6) and (7). Note that the proposed method is semiparametric and thus the dimension of the parameter $\Omega = (\beta, \{dH\})$ is of the same order as the sample size.

We can write the NPMLE of H based on $U_{H(t)}$ as

$$\hat{H}(t) = \sum_{i=1}^n \int_0^t \frac{\sum_{k=1}^2 dN_{ki}(x) + d\tilde{N}_{2i}(x)}{S_0(\bar{H}(\xi), x; \beta)}, \text{ where} \quad (8)$$

$$S_0(\bar{H}(\xi), x; \beta) = \sum_{i=1}^n \left[\sum_{k=1}^2 Y_i(t) \omega_{ki}(\bar{H}(\xi), t; \beta) \Theta_{ki}(H(t); \beta) + \tilde{Y}_i(t) \tilde{\Theta}_{2i}(\beta) \right].$$

Note that $\hat{H}$ can be viewed as a weighted Breslow-type estimator (Breslow, 1975), and has a close connection to the estimator proposed by Chen (2009) for semiparametric transformation models. The weights ω_{ki} have unit expectation at the true model parameters by martingale properties of score function. Therefore, maximization of the likelihood with respect to H can be done by an iterative reweighting algorithm (Chen, 2009). Given β , starting with initial weights $\omega^{(0)} = 1$ and initial values $dH^{(0)}(t)$ (for example based on the Nelson-Aalen estimator), we repeat the following steps until convergence:

1. Fix weights $\omega^{(k)}$ and solve the estimating equation (8) and obtain the solution $H^{(k+1)}$. The weights being fixed, S_0^ω in (8) only depends on the current value $H(t)$ rather than its full path, and the problem of solving for H is recurrent where a value

of H for the next point of jump is found by solving a univariate algebraic equation, the solutions for all previous jumps available from the previous steps.

2. Update weights $\omega^{(k+1)}$ using $H^{(k+1)}$.

By replacing $dH(t)$ and $H(t)$ in the loglikelihood by the point of convergence $d\hat{H}(t)$ and $\hat{H}(t)$ of the above procedure, we obtain the profile likelihood $\ell_{\text{pr}}(\beta)$ that can be maximized by conventional finite-dimensional methods.

4 Asymptotic Properties

We use a combination of empirical process and martingale theory following a general line of (Zeng and Lin, 2007, 2010; Chen, 2009, 2010).

Regularity conditions are listed in the Supplemental Materials Section C. Based on the score functions (6) and (7), we can write the score equations for $\Omega = (\beta, \{dH\})$ in martingale form

$$U_{\beta} = \sum_{i=1}^n \int_0^{\xi} \left\{ \sum_{k=1}^2 \frac{\Theta_{ki,\beta}(H(x);\beta)}{\Theta_{ki}(H(x);\beta)} dM_{ki}(x) + \frac{\tilde{\Theta}_{2i,\beta}(\beta)}{\tilde{\Theta}_{2i}(\beta)} d\tilde{M}_{2i}(x) \right\}, \quad (9)$$

and

$$U_{H(t)} = \sum_{i=1}^n \int_0^t \left\{ \sum_{k=1}^2 \left[dM_{ki}(x) + \int_{x+}^{\xi} [\psi_{ki}(H(u);\beta) dM_{ki}(u)] dH(x) \right] + d\tilde{M}_{2i}(x) \right\}.$$

After exchanging the integration over u and x , it is not difficult to show:

$$U_{H(t)} = \sum_{i=1}^n \int_0^{\xi} \left\{ \sum_{k=1}^2 \varepsilon_{ki}(u, t; \beta, H) dM_{ki}(u) + I(u \leq t) d\tilde{M}_{2i}(u) \right\}, \quad (10)$$

where $\varepsilon_{ki}(u, t; \beta, H) = I(u \leq t) + \varphi_{ki}(H(u); \beta)H(u \wedge t)$. As we show in the Supplementary Materials Section B, the linear transform $\int_0^{\xi} \varepsilon_{ki}(u, t; \beta, H) dM_{ki}(u)$ and $\int_0^{\xi} I(u \leq t) d\tilde{M}_{2i}(u)$ is a martingale when $\varepsilon_{ki}(u, t; \beta, H)$ does not depend on t for $u \leq t$ as is the case here. Therefore, the score functions U_{β} and $U_{H(t)}$ are both martingales under the true model.

In the following we present the consistency and weak convergence results for the proposed NPMLE $\hat{\Omega} = (\hat{\beta}, \hat{H}(t))$ with details given in the Supplementary Materials Section C.

Theorem 1 Assuming regularity conditions hold, with probability one, $\hat{\beta}$ converges to β^0 , $\hat{H}(t)$ converges to $H^0(t)$ uniformly in the interval $[0, \xi]$, where $H^0(t)$, β^0 are the true values of $H(t)$, β .

Consider a linear functional

$$n^{1/2} \left\{ a^T (\hat{\beta} - \beta^0) + \int_0^\xi b(t)^T d(\hat{H}(t) - H^0(t)) \right\}, \quad (11)$$

where a is real vector, $b(t)$ is a function with bounded total variation in $[0, \xi]$, and let B be the vector consisting of the values of $b(t)$ evaluated at the observed failure times corresponding to the set $\{dH\}$, and $\varepsilon^T = (a^T, B^T)$. We have

Theorem 2 Assuming regularity conditions hold, $n^{1/2}\{\hat{\beta} - \beta^0, \hat{H}(t) - H^0(t)\}$ converges weakly to a zero-mean Gaussian process. In addition, $n\varepsilon^T (\mathcal{I}_n)^{-1} \varepsilon$ converges in probability to the asymptotic variance-covariance function of the linear functional (11), where $\mathcal{I}_n$ is the negative Hessian matrix of the observed log-likelihood function with respect to $\hat{\Omega}$.

For a differentiable functional $F(\Omega)$ of Ω , based on the functional delta method (Andersen et al, 1993) Section II.8), $n^{1/2}\{F(\hat{\Omega}) - F(\Omega^0)\}$ converges weakly to a zero-mean Gaussian process with variance-covariance function $\dot{F}(\Omega)^T (\mathcal{I}^0)^{-1} \dot{F}(\Omega)$, where $\dot{F}(\Omega)$ is the gradient of $F(\Omega)$ with respect to Ω , $\mathcal{I}^0$ can be consistently estimated by $n^{-1} \mathcal{I}_n$, with the explicit expression of $\mathcal{I}_n$ is derived in the Supplementary Materials Section C.3.

5 Numerical Examples

5.1 Simulation Studies

Data on two event times T_1^* and T_2^* are simulated based on the marginal and conditional specifications. We consider the following true models: (i) all non-terminal events are observable (scenario I, $p = 0$); or (ii) 30% subjects' non-terminal events are unobservable due to the MAR missingness mechanism (scenario II, $p = 0.3$). For scenario II, it is important to distinguish that while 30% subjects' non-terminal events are unobservable, the actual observed non-terminal event proportion may vary depending on the censoring distribution. Both scenarios assume a common baseline hazard $h(t) = 0.1$, with two covariates Z_1 and Z_2 included in regression models of T_1^* and T_2^* , where Z_1 is generated from a uniform distribution $U(0, 1)$, and Z_2 is generated from a binomial distribution with a success probability of 0.5. The true logarithmic hazard ratios for T_1^* and T_2^* are $\beta_{\eta 1} = 0.5$, $\beta_{\eta 2} = 1$, $\beta_{\mu 1} = -0.5$, $\beta_{\mu 2} = 0.5$, and $\beta_{\mu 0} = 1$, $\beta_{\mu 1} = 0.5$, $\beta_{\mu 2} = 0$. Under independent right censoring time from a uniform distribution $U(0, 10)$, in scenario I, 31.1% of subjects have both non-terminal and terminal events, 14.9% only have a non-terminal event and censored terminal event, 22.6% have only a terminal event, and 31.5% have neither non-terminal nor terminal events; in scenario II, on average 30% of subjects' non-terminal events are missing, the proportions with different types of events being 22.1%, 10.9%, 31.2%, and 35.8%, respectively. For each simulation scenario, 1,000 datasets are generated.

Under both scenarios, the simulated datasets were analyzed by the proposed method with and without accounting for missing non-terminal events (Model I and II, respectively), as well as by the Markov illness-death model based on partial likelihood (Kalbfleisch and Prentice, 2002), p.270). The performance of all these methods was assessed under sample sizes of $n = 250$ and $n = 400$. The Markov model assumes a set of baseline hazards, one for each transition intensity, $\lambda_{jk}(t) = \lambda_{jk0}(t) \exp(\beta_{jk} Z_j)$, for $j < k$, $j = 0, 1$, $k = 1, 2$. Compared with the Markov model, Model II is more parsimonious since it assumes proportional baseline

hazards, such that (i) β_η and β_θ have same interpretations as β_{01} and β_{02} ; (ii) $\beta_{\mu 0}$ in our proposed model can be viewed as the log hazard ratio between λ_{120} and λ_{020} ; (iii) β_μ can be viewed as the interaction between Z and the occurrence of T_1^* , and $\beta_\mu = \beta_{12} - \beta_{02}$. For comparison purpose, β_{12} in the Markov model is expressed in terms of β_μ based on the above relationship. In addition, $\beta_p = \text{logit}(p)$ is expressed in p and the corresponding asymptotic standard deviation is obtained through Delta method.

In the computation of hazard estimators, initial values for the algorithm are taken from a Nelson-Aalen estimator. Convergence criterion was specified as the magnitude of relative improvement in the full likelihood falling below tolerance of 10^{-6} . Tolerance for the solution of the hazard equation was taken one order of magnitude smaller (10^{-7}) compared to the likelihood maximization over the finite-dimensional parameter vector. The computation aspects of the iterative reweighting algorithm are discussed and compared with the direct maximization approach (Zeng and Lin, 2007) in Supplementary Materials Section D.

The simulation results are summarized in Tables 1 and 2, respectively.

When the observed data do not contain missing non-terminal events (scenario I, Table 1), as expected, we find Model II performs best. Slight small-sample bias of the NPMLE for regression coefficients β reduces as sample size increases. The estimated mean asymptotic standard errors obtained from the observed information matrix are quite close to the empirical standard errors, and the resultant 95% confidence intervals attain appropriate coverage probabilities. We also find that the method allowing for missing non-terminal events (Model I) also achieves almost unbiased estimates, and maintains the 95% coverage probability at nominal levels, especially with large sample size of $n = 400$. The model with no missing non-terminal events, $p = 0$, in Model I gives a reasonably small bias of 0.04–0.05, considering the fact that $p = 0$ ($\text{logit}(p) = -\infty$) is on the boundary of parameter space for this model. The estimates from the Markov model, while unbiased, are less efficient than results from Model II, since the more parsimonious model is correctly specified.

When missing non-terminal events (30%) are present (scenario II), our proposed method incorporating the missing mechanism (Model I) is superior. As sample size increases, we find diminishing bias of the proposed estimators, good correspondence between asymptotic standard errors and empirical standard errors, and correct 95% coverage probability at nominal levels. Neither method assuming complete data can match its performance, as expected. The Markov model allowing different baseline hazards for different transition intensities has slightly smaller bias than Model II due to its higher flexibility. Both methods ignoring missing data provide biased estimates and unreliable statistical inference.

5.2 Data Analysis

We apply the proposed method to a follicular cell lymphoma dataset collected at the Princess Margaret Hospital, Toronto, between 1967 and 1996 (Pintilie, 2006). Follicular cell lymphoma is a common type of non-Hodgkin lymphoma (NHL), a slow-growing lymphoma arising from B-cells. It mainly affects older adults with few early warning signs, and is

usually difficult to cure (unlike the Hodgkin's lymphoma). It is characterized by a relatively high cancer relapse rate even with very good response to treatment.

In this observational study cohort, 517 patients with early stage disease (I or II) were treated with either radiation alone (RT, 80.0%) or with radiation and chemotherapy (RT+CMT, 20.0%). Patients who had complete treatment response are included in the analysis. The times to cancer relapse and death are available as outcome variables, and are measured in years from the date of cancer diagnosis. Other variables assessed at the time of diagnosis include patients' age (mean 56.6 years, sd 14.0 years) and haemoglobin level (mean 139.1 g/l, sd 15.3 g/l). The median follow-up time is 8.9 years. Among the 354 (68.5%) and 163 (31.5%) patients with stage I and II cancers, respectively, 301 (85.0%) and 113 (69.3%) were treated by RT alone ($p < .0001$). Two-sample t-tests find the distributions of patients' age are significantly different between treatment groups ($p = 0.032$) and clinical stages ($p = 0.0002$), while the distributions of hemoglobin level are similar between treatment groups ($p = 0.469$) and clinical stages ($p = 0.202$).

Among the 517 patients, 162 (31.3%) had prior relapse before death, 74 (14.3%) died without a prior recorded relapse, 88 (17.0%) had relapse and were censored afterwards, and 193 (37.3%) were censored for both events. Like in many other cancer studies, it is plausible to expect that patients who died without recorded relapse or were censored for both events may have experienced a relapse that went missing.

We first fit a model for time to relapse and time to death separately: (i) a Cox PH model for time to relapse under independent censoring assumption for death; and (ii) a Cox PH model on time to death with relapse as a time-dependent covariate. The results are summarized in Table 3. The baseline covariates included for regression analysis are treatment group, clinical stage, and patients' age (centered at its sample mean 56.6 years and standardized by its sample standard deviation 14.0 years). We find the effect of patients' hemoglobin level at baseline is always insignificant for all methods, and therefore it is not included in the final model. After adjusting for other covariates, separate analysis finds that stage II, older age patients treated by RT had shorter time to relapse, and those diagnosed at older age also had shorter time to death without relapse. In addition, patients with relapse prior to time t are expected to have about 10 times ($HR = \exp(2.33), p < .0001$) the risk of dying at subsequent times as compared with a patient who has not experienced a relapse, and older patients may live longer than younger patients if both had relapse before ($p = 0.085$). This finding suggests a strong positive association between relapse and death, as expected.

Our proposed illness-death models taking account of missing relapse was applied to the follicular cell lymphoma dataset (Table 3). We estimated that, among all patients who had experienced cancer relapse, 29.5% ($se = 3.6\%$, $p < .0001$) were not recorded as relapsed. With this important correction to the observed semicompeting risks data, we find that cancer relapse is positively associated with death, as the risk of death for a patients with prior relapse is about 5 times that of a patient without relapse ($HR = \exp(1.634), p < .0001$). Compared with the other models, not only our model finds similar effects of clinical stage, patient's age, a stronger effect of treatment on time to relapse, but also concludes that stage I, younger patients receiving RT lived longer if they did not have relapse. Furthermore, a

patient with stage I cancer and prior relapse is expected to die sooner ($p=0.019$). This result is not surprising, as many early-stage cancers become more aggressive and lethal once they relapse. In fact, our alternative analysis of time to death using a Cox PH model with relapse as a time-dependent covariate also suggests a similar but insignificant finding ($p=0.470$).

One important feature of our proposed joint model for semicompeting risks data is a measure of dependence between times to relapse and death based on covariate-specific Kendall's τ (Figure 2). The estimated Kendall's τ from Model I and Model II show similar trends but are otherwise slightly different. Older patients, except for those stage II patients who received RT+CMT, tend to have stronger dependence between relapse and death; stage II patients who received RT+CMT seem to have the strongest dependence among all patients.

In summary, our proposed method is in general agreement with separate analyses identified additional significant covariate effects. In particular, our proposed model supports a positive correlation between relapse and death, especially for stage I patients. Based on our simulation results in Section 5.1, the nontrivial proportion of unobserved patients' relapse status (29.5%) was estimated. Failing to recognize missing relapses may lead to diluted covariate effects and misleading biased estimates.

6 Discussion

We have developed an illness-death model for semicompeting risks that recognizes that an unknown proportion of non-terminal events may be missing at random (left censored) rather than right censored, if not observed prior to death or a censoring event. The inherent difficulty on top of the semicompeting risks problem is that, both the time to non-terminal event and its status may be unobserved in this situation. In addition to offering a general treatment for the missingness of non-terminal event, the proposed illness-death model incorporates positive and negative dependence between non-terminal and terminal events. Dependent censoring issues were taken into account by the proposed joint model in a natural way.

Time to non-terminal event may be viewed as a mixture of exact, left and right censored data with partially missing status and dependent censoring. The PH assumption of common baseline hazard allowed a methodology similar to the one used with transformation models avoiding irregular convergence rates of current status data. The same assumption of proportionality between hazards for non-terminal and terminal events yielded close-form expressions for the model and likelihood contributions facilitating an explicit evaluation of the properties of the model such as measures of local and global dependence.

The conventional methodology for Markov multistate models, including illness-death models and progressive multistate models, are mainly based on partial likelihood. While these methods are flexible in analyzing multiple outcome data like semicompeting risks and recurrent event data, limited research has been conducted to address the situations when states are *not* always observable.

One advantage of our proposal is that the score functions retain a martingale structure, such that the martingale central limit theorems can be used to establish weak convergence. Empirical processes (a Glivenko-Cantelli argument) and functional differentiation are used to establish uniform law of large numbers implying consistency.

Relaxing the missing at random assumption does not appear to be difficult if an alternative model for the missing data mechanism is identified.

Under the proportional hazards assumption between the non-terminal and terminal risks, even if the non-terminal event is never observed ($p=1$), the model remains identifiable. Identifiability conditions can be expressed as an intergo-differential equation with the cumulative incidence of terminal event on the one side and its model-based counterpart on the other side. Uniqueness of solution can be guaranteed by regularity conditions for such equations. With non-terminal event unobservable, it represents a frailty variable whose distribution is parameterized by the same hazard H . Formal identifiability result can then be established similar to how this is done with univariate frailty models. By implication, a more informative partially observable non-terminal event setting will keep the model identifiable.

The PH assumption can be relaxed without altering the basic framework of this paper by linking the baseline hazards for the terminal and non-terminal events using a parametric transformation. While we can justify the PH assumption using mechanistic models of cancer, for other diseases models parameterized by two infinite dimensional parameters represent an interesting issue for future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the associate editor, and two referees for their helpful comments, which have greatly improved the manuscript. This research was supported by National Cancer Institute grant CA157224 (CISNET).

References

- Andersen P, Keiding N. Multi-state models for event history analysis. *Statistical Methods in Medical Research*. 2002; 11(2):91–115. [PubMed: 12040698]
- Andersen P, Hansen L, Keiding N. Non-and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous markov process. *Scandinavian Journal of Statistics*. 1991:153–167.
- Andersen, P.; Borgan, Ø.; Gill, R.; Keiding, N. *Statistical models based on counting processes*. Springer; 1993.
- Andersen P, Esbjerg S, Sørensen T. Multi-state models for bleeding episodes and mortality in liver cirrhosis. *Statistics in medicine*. 2000; 19(4):587–599. [PubMed: 10694738]
- Breslow N. Analysis of survival data under the proportional hazards model. *International Statistical Review/Revue Internationale de Statistique*. 1975; 43(1):45–57.
- Chen K, Jin Z, Ying Z. Semiparametric analysis of transformation models with censored data. *Biometrika*. 2002; 89(3):659–668.
- Chen Y. Weighted Breslow-type and maximum likelihood estimation in semiparametric transformation models. *Biometrika*. 2009; 96(3):591–600.

- Chen Y. Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2010; 72(2):235–251.
- Chen YH. Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. *Lifetime data analysis*. 2012; 18(1):36–57. [PubMed: 21850528]
- Clayton D. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*. 1978; 65(1):141.
- Cook R, Lawless J. *The statistical analysis of recurrent events*. 2007
- Fine J, Gray R. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. 1999:496–509.
- Fine J, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001; 88(4):907–919.
- Govindarajulu U, Lin H, Lunetta K, D'Agostino R. *Frailty models: Applications to biomedical and genetic studies*. *Statistics in Medicine*. 2011
- Hougaard P. *Analysis of multivariate survival data*. 2000
- Hsieh J, Wang W, Adam Ding A. Regression analysis based on semi-competing risks data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2008; 70(1):3–20.
- Huang X, Zhang N. Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach. *Biometrics*. 2008; 64(4):1090–1099. [PubMed: 18266895]
- Kalbfleisch, J.; Prentice, R. *The statistical analysis of failure time data*. Vol. 360. Wiley-Interscience; 2002.
- Kosorok, M. *Introduction to empirical processes and semiparametric inference*. Springer Verlag; 2008.
- Little R, Rubin D. *Statistical analysis with missing data*. 2002; 333
- Nielsen G, Gill R, Andersen P, Sørensen T. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*. 1992:25–43.
- Oakes D. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*. 1989; 84(406):487–493.
- Peng L, Fine J. Regression modeling of semicompeting risks data. *Biometrics*. 2007; 63(1):96–108. [PubMed: 17447934]
- Pintilie, M. *Competing risks: a practical perspective*. Vol. 22. Wiley; 2006.
- Shu Y, Klein J, Zhang M. Asymptotic theory for the Cox semi-Markov illness-death model. *Lifetime Data Analysis*. 2007; 13(1):91–117. [PubMed: 17115258]
- Tsodikov A. Semiparametric models: a generalized self-consistency approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2003; 65(3):759–774.
- Van Der Vaart, A.; Wellner, J. *Weak convergence and empirical processes*. Springer Verlag; 1996.
- Wang W. Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2003; 65(1):257–273.
- Xu J, Kalbfleisch J, Tai B. Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics*. 2010; 66(3):716–725. [PubMed: 19912171]
- Zeng D, Lin D. Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*. 2006; 93(3):627–640.
- Zeng D, Lin D. Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2007; 69(4):507–564.
- Zeng D, Lin D. A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. *Statistica Sinica*. 2010; 20(2):871. [PubMed: 20577580]
- Zheng M, Klein J. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*. 1995; 82(1):127.

Fig. 1.
The three-state illness-death model

Fig. 2.

Estimated Kendall's τ . Model I: accounting for possibly missing relapse; Model II: assuming no missing relapse except due to censoring

Fig. 3.

Precision of likelihood maximization by the computation time (precision is measured as the L2 norm of the difference between the limiting value of the likelihood as the tolerances tend to $-\infty$ and the maximal likelihood value achieved under a stopping rule): Iterative Reweighting (IW), Direct Maximization (DM)

Fig. 4. Numerical efficiency by different sample sizes at tolerance level $1e-6$: Iterative Reweighting (IW), Direct Maximization (DM) 28 Chen Hu, Alex Tsodikov

Table 1

Simulation results for a model with two covariates; no missing non-terminal events except by right censoring

	Parameter						
	$\beta_{\eta} = 0.5$	$\beta_{\eta} = 1$	$\beta_{\eta} = -0.5$	$\beta_{\eta} = 0.5$	$\beta_{\eta} = 1$	$\beta_{\eta} = 0.5$	$p = 0$
<i>n = 250, Model I</i>							
Bias	0.052	0.039	-0.102	-0.072	-0.028	0.096	0.092
SE	0.301	0.194	0.446	0.314	0.375	0.625	0.412
ASD	0.300	0.192	0.463	0.301	0.366	0.627	0.399
CP	94.5%	94.2%	95.2%	93.7%	95.1%	94.2%	96.2%
<i>n = 250, Model II</i>							
Bias	-0.011	0.005	-0.019	-0.010	-0.009	0.003	0.021
SE	0.307	0.192	0.406	0.273	0.376	0.606	0.373
ASD	0.288	0.184	0.391	0.257	0.354	0.565	0.358
CP	93.7%	93.4%	93.7%	93.3%	94.0%	93.5%	93.7%
<i>n = 250, Markov Model</i>							
Bias	0.017	0.013	-0.006	0.000	-	-0.006	0.009
SE	0.341	0.204	0.489	0.298	-	0.653	0.397
ASD	0.329	0.199	0.487	0.285	-	0.638	0.384
CP	94.8%	94.8%	94.6%	94.7%	-	94.6%	94.1%
<i>n = 400, Model I</i>							
Bias	0.050	0.029	-0.086	-0.061	-0.017	0.087	0.069
SE	0.227	0.153	0.357	0.236	0.282	0.481	0.304
ASD	0.236	0.151	0.364	0.233	0.285	0.491	0.309
CP	94.8%	94.4%	96.0%	94.6%	95.4%	95.0%	95.3%
<i>n = 400, Model II</i>							
Bias	-0.014	-0.003	-0.024	-0.008	-0.006	0.009	0.013
SE	0.245	0.152	0.339	0.211	0.289	0.477	0.285
ASD	0.226	0.145	0.307	0.202	0.276	0.441	0.280
CP	93.6%	93.4%	93.2%	93.6%	93.5%	92.8%	93.8%
<i>n = 400, Markov Model</i>							
Bias	0.005	0.003	0.000	0.007	-	-0.003	-0.006

Parameter							
	$\beta_{\eta_1} = 0.5$	$\beta_{\eta_2} = 1$	$\beta_{\eta_1} = -0.5$	$\beta_{\eta_2} = 0.5$	$\beta_{\eta_0} = 1$	$\beta_{\eta_1} = 0.5$	$\beta_{\eta_2} = 0$ $p = 0$
SE	0.263	0.161	0.373	0.229	-	0.484	0.291
ASD	0.257	0.156	0.381	0.223	-	0.497	0.300
CP	94.8%	94.2%	96.0%	94.2%	-	96.2%	95.9%

Model I: Proposed model with missingness

Model II: Proposed model without missingness

SE: Empirical Standard Errors

ASD: Asymptotic Standard Deviations

CP: 95% Coverage Probability

Table 2

Simulation results a model with two covariates, missing non-terminal events present

	Parameter						
	$\beta_{\eta} = 0.5$	$\beta_{\eta} = 1$	$\beta_{\eta} = -0.5$	$\beta_{\eta} = 0.5$	$\beta_{\eta} = 1$	$\beta_{\eta} = 0.5$	$p = 0.3$
<i>n</i> = 250, <i>Model I</i>							
Bias	-0.098	-0.017	0.015	-0.004	0.014	-0.044	0.022
SE	0.382	0.249	0.632	0.383	0.497	0.862	0.525
ASD	0.394	0.254	0.605	0.367	0.475	0.835	0.509
CP	93.7%	94.8%	93.2%	94.6%	94.1%	93.8%	94.9%
<i>n</i> = 250, <i>Model II</i>							
Bias	-0.335	-0.162	0.489	0.248	0.130	-0.462	-0.214
SE	0.356	0.217	0.335	0.221	0.449	0.622	0.383
ASD	0.325	0.210	0.343	0.222	0.411	0.597	0.373
CP	81.3%	87.0%	70.0%	80.4%	91.9%	87.5%	89.8%
<i>n</i> = 250, <i>Markov Model</i>							
Bias	-0.069	-0.100	0.250	0.199	-	-0.232	-0.174
SE	0.406	0.237	0.401	0.244	-	0.658	0.398
ASD	0.391	0.235	0.403	0.237	-	0.640	0.392
CP	94.8%	93.3%	91.0%	87.4%	-	92.6%	92.4%
<i>n</i> = 400, <i>Model I</i>							
Bias	-0.041	-0.025	0.029	0.000	0.029	-0.040	0.007
SE	0.302	0.212	0.488	0.332	0.372	0.678	0.424
ASD	0.305	0.206	0.473	0.327	0.362	0.659	0.419
CP	94.5%	94.3%	93.7%	94.9%	94.3%	93.8%	94.8%
<i>n</i> = 400, <i>Model II</i>							
Bias	-0.325	-0.177	0.493	0.249	0.148	-0.507	-0.224
SE	0.278	0.177	0.273	0.177	0.328	0.482	0.285
ASD	0.256	0.165	0.269	0.175	0.321	0.468	0.292
CP	74.5%	78.8%	54.3%	69.6%	92.1%	80.2%	88.7%
<i>n</i> = 400, <i>Markov Model</i>							
Bias	-0.062	-0.116	0.252	0.200	-	-0.270	-0.186

Parameter							
	$\beta_{\eta_1} = 0.5$	$\beta_{\eta_2} = 1$	$\beta_{\eta_1} = -0.5$	$\beta_{\eta_2} = 0.5$	$\beta_{\eta_0} = 1$	$\beta_{\eta_1} = 0.5$	$\beta_{\eta_2} = 0$ $p = 0.3$
SE	0.314	0.194	0.301	0.194	-	0.485	0.295
ASD	0.308	0.185	0.316	0.186	-	0.499	0.305
CP	93.1%	88.4%	87.3%	80.1%	-	92.7%	91.0%

Model I: Proposed model with missingness

Model II: Proposed model without missingness

SE: Empirical Standard Errors

ASD: Asymptotic Standard Deviations

CP: 95% Coverage Probability

Table 3

Results of separate Cox PH analyses, and the proposed model accounting for possibly missing relapse

Coefficients	Separate Analysis			Proposed Model		
	Estimate	SE	P-value	Estimate	SE	P-value
<i>Time to relapse</i>						
RT vs RT+CMT	0.641	0.196	0.001	1.554	0.227	<.0001
Stage I vs II	-0.463	0.139	.0009	-0.451	0.166	0.007
Age	0.269	0.068	<.0001	0.329	0.087	<.0001
<i>Time to death without relapse</i>						
RT vs RT+CMT	0.219	0.351	0.533	-1.868	0.695	0.007
Stage I vs II	-0.440	0.277	0.112	-1.400	0.522	0.007
Age	1.098	0.147	<.0001	1.097	0.289	<.0001
<i>Time to death with relapse</i>						
Relapse	2.332	0.413	<.0001	1.634	0.315	<.0001
RT vs RT+CMT	-0.630	0.431	0.144	1.304	0.745	0.080
Stage I vs II	0.235	0.325	0.470	1.296	0.554	0.019
Age	-0.299	0.173	0.085	-0.373	0.310	0.229
Prob. of missing relapse	-	-	-	0.295	0.036	<.0001