Biometrics 66, 716–725
September 2010
DOI: 10.1111/j.1541-0420.2009.01340.x

Statistical Analysis of Illness–Death Processes and Semicompeting Risks Data

Jinfeng Xu,[1,*] John D. Kalbfleisch,[2] and Beechoo Tai[3]

[1]Department of Statistics and Applied Probability, Risk Management Institute, National University of Singapore, Singapore 117546, Singapore
[2]Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, Michigan 48109-2029, U.S.A.
[3]Department of Epidemiology and Public Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore
[*]*email:* staxj@nus.edu.sg

Summary. In many instances, a subject can experience both a nonterminal and terminal event where the terminal event (e.g., death) censors the nonterminal event (e.g., relapse) but not vice versa. Typically, the two events are correlated. This situation has been termed semicompeting risks (e.g., Fine, Jiang, and Chappell, 2001, *Biometrika* **88,** 907–939; Wang, 2003, *Journal of the Royal Statistical Society, Series B* **65,** 257–273), and analysis has been based on a joint survival function of two event times over the positive quadrant but with observation restricted to the upper wedge. Implicitly, this approach entertains the idea of latent failure times and leads to discussion of a marginal distribution of the nonterminal event that is not grounded in reality. We argue that, similar to models for competing risks, latent failure times should generally be avoided in modeling such data. We note that semicompeting risks have more classically been described as an illness–death model and this formulation avoids any reference to latent times. We consider an illness–death model with shared frailty, which in its most restrictive form is identical to the semicompeting risks model that has been proposed and analyzed, but that allows for many generalizations and the simple incorporation of covariates. Nonparametric maximum likelihood estimation is used for inference and resulting estimates for the correlation parameter are compared with other proposed approaches. Asymptotic properties, simulations studies, and application to a randomized clinical trial in nasopharyngeal cancer evaluate and illustrate the methods. A simple and fast algorithm is developed for its numerical implementation.

Key words: Copula; Dependent censoring; Frailty; Illness–death model; Proportional hazards; Semicompeting risks data; Terminal event.

1. Introduction

In clinical trials comparing therapeutic interventions, a subject may experience several distinct types of failures. If aside from censoring, the observation time ends upon the occurrence of the first failure, such data are commonly referred to as competing risks data. If all types of failures can occur in any order until possible censoring, this gives rise to multivariate failure time data. In this article, we consider the situation where a subject may experience a nonterminal event, such as disease recurrence, and/or a terminal event, such as death, where the terminal event censors the nonterminal event but not vice versa. This is sometimes referred to as semicompeting risks data (Fine, Jiang, and Chappell, 2001). A typical example is provided by a randomized clinical trial of nasopharyngeal cancer, which was recently conducted to compare radiation therapy alone (RT) with a combination chemotherapy and radiation therapy (CRT) treatment (Wee et al., 2005). Two endpoints were of interest—time to tumor recurrence and time to mortality. Large percentages of deaths were observed among the patients with recurrence in both the RT and CRT groups. In this and in many similar applications, it is to be expected that the time to the nonterminal event is strongly correlated with the time to the terminal event. The specific aims of the study are then to estimate covariate and treatment effects on the rates of terminal and nonterminal events, and also to evaluate the dependence or correlation of these events.

Semicompeting risks data without covariates have previously been modeled by assuming that the joint survivor function of the two event times follows a copula with two margins and with observations available only in the upper wedge, $t_1 \leqslant t_2$. For example, Fine et al. (2001) define the Clayton copula model,

$$S(T_1, T_2) = P(T_1 > t_1, T_2 > t_2) = \{S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} - 1\}^{-1/\theta},$$

where T_1, T_2 are times to nonterminal and terminal events, S_1, S_2 are their respective marginal survival functions, and $\theta \geqslant 0$ is a parameter measuring the correlation. They extended the concordance estimator of Oakes (1982) to estimate and test hypotheses about the dependency parameter θ. A pseudo partial likelihood estimator of θ was proposed by Clayton (1978) and considered by Day, Bryant, and Lefkopoulou (1997) who established its consistency. The asymptotic

normality of this estimator was shown by Wang (2003), who also considered an extension to more general copula models.

All of these models and approaches involve similar assumptions to those underlying multiple decrement models for latent failure times in some analyses of competing risks data. In particular, the interpretation of the marginal distribution of the nonterminal event is hypothetical as are the marginal models in the multiple decrement formulation of competing risks. Another difficulty with these models for semicompeting risks is that they complicate covariance analysis. Currently there is limited literature on regression analysis of semicompeting risks data, except for Peng and Fine (2007) who postulated separate marginal regression models for the time to nonterminal and terminal events.

In this article, we suggest that semicompeting risks data are better modeled using an illness–death compartment model (Andersen et al., 1993; Kalbfleisch and Prentice, 2002). In fact, we essentially argue (as pointed out by an associate editor) that "semicompeting risks" is nothing but the more than 50-year-old illness–death model (Fix and Neyman, 1951; Sverdrup, 1965). In the illness–death model, a subject can either transit directly to the terminal event or first to the nonterminal event and then to the terminal event. The model is completely specified by the transition intensity functions for the three distinct transitions. In particular, we consider a class of illness–death models with a shared frailty. In its restricted form, this model is essentially equivalent to the Clayton model mentioned above. In the shared frailty models, covariance analysis can be incorporated either through the conditional (on the frailty) or marginal (after integrating out the frailty) transition intensities. We use nonparametric maximum likelihood estimation (NPMLE) for inference and develop a simple iterative procedure for its numerical implementation. The maximum likelihood estimators for the regression coefficients and the dependence parameter are shown to be efficient.

The remainder of this article is organized as follows. In Section 2, we describe a general illness–death model with a shared frailty and consider a special case that is essentially equivalent to the copula model. We consider further this very special case in Section 3, where statistical analyses based on maximum likelihood, pseudo partial likelihood (Day et al., 1997), and concordance (Fine et al., 2001) are considered and compared. Section 4 returns to the general shared frailty type model and discusses covariance analysis, and Section 5 presents results of simulation studies. The methods are applied to data from a randomized trial of nasopharyngeal cancer in Section 6, and this article concludes with some discussion in Section 7. All proofs are given in the Web Supplementary Materials.

2. The Model

Let C_i and T_{i2} be the censoring and terminal event times and T_{i1} be the nonterminal event time for the ith subject, $i = 1, \ldots, n$. If the subject fails before the nonterminal event occurs, then we define $T_{i1} = \infty$. We suppose that there is a p-dimensional vector of covariates, x_i, measured on the ith individual. We assume that the censoring time C_i is independent of T_{i1}, T_{i2} given x_i. We first consider the homogeneous case (without covariates) and covariance analysis will be considered in Section 4. The observations can be simply described as $(Y_{i2} = T_{i2} \wedge C_i, \delta_{i2} = I(T_{i2} \leqslant C_i), Y_{i1} = T_{i1} \wedge Y_{i2}, \delta_{i1} = I(T_{i1} \leqslant Y_{i2}), x_i, i = 1, \ldots, n)$. Because $0 \leqslant Y_{i1} \leqslant Y_{i2}$, the observations are restricted to the upper wedge. Note, for example, that if $\delta_{i1} = 0$ and $\delta_{i2} = 1$, then $Y_{i1} = Y_{i2} = T_{i2}$ and $T_{i1} = \infty$.

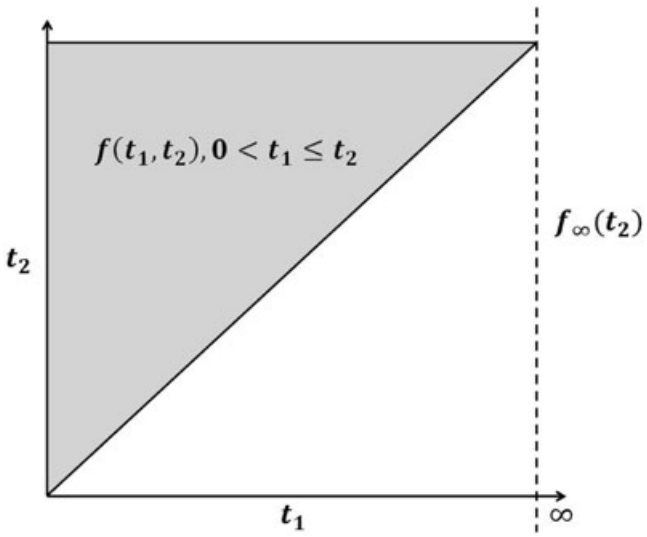

Figure 1. Joint density function of T_1 and T_2.

The probability model for (T_1, T_2) is taken to be absolutely continuous in the upper wedge $t_2 \geqslant t_1$ with joint density $f(t_1, t_2),\ 0 \leqslant t_1 \leqslant t_2$ as shown in Figure 1. Because

$$\int_0^\infty \int_{t_1}^\infty f(t_1, t_2)\, dt_2\, dt_1 = P(T_1 < \infty) \leqslant 1.$$

The balance of the probability is distributed along the line at $t_1 = \infty$ with continuous density $f_\infty(t_2), t_2 > 0$. Note that this specification of the model specifies no probability content in the lower wedge $t_2 < t_1 < \infty$, which is a true reflection of the physical situation. We suggest that this method of modeling is generally preferable to assuming a latent distribution of T_1, T_2 over the region $t_1 > t_2$ as is often done in this case (e.g., Fine et al., 2001). This latter approach is similar to the approach of latent failure times in competing risks, which should generally be avoided in modeling. See Prentice et al. (1978) or Kalbfleisch and Prentice (2002) for additional discussion of this point.

2.1 *The Illness–Death Model*

One approach to modeling data of the type considered here is to utilize a compartment type model with transition functions defining the probabilistic properties. Figure 2 illustrates this situation with three compartments labeled *on study* for the state occupied at time $t = 0$, *recurrence* for the state entered when the intermediate event occurs, and *death* for the absorbing state. The hazard or transition rates are defined as follows:

$$\lambda_1(t_1) = \lim_{\Delta \to 0} P[T_1 \in [t_1, t_1 + \Delta) \,|\, T_1 \geqslant t_1, T_2 \geqslant t_1]/\Delta, \quad (1)$$

$$\lambda_2(t_2) = \lim_{\Delta \to 0} P[T_2 \in [t_2, t_2 + \Delta) \,|\, T_1 \geqslant t_2, T_2 \geqslant t_2]/\Delta, \quad (2)$$

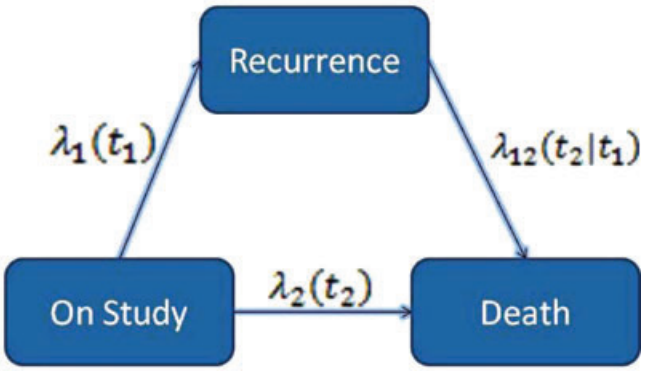

Figure 2. Compartment for semicompeting risks data. This figure appears in color in the electronic version of this article.

$$\lambda_{12}(t_2 \,|\, t_1) = \lim_{\Delta \to 0} P[T_2 \in [t_2, t_2 + \Delta) \,|\, T_1 = t_1, T_2 \geqslant t_2]/\Delta, \quad (3)$$

where $0 < t_1 < t_2$. Note that equations (1) and (2) are the usual cause specific or crude hazard functions for the competing risks part of the model in which either the terminal or nonterminal event occurs first. The hazard (3) defines the rate of the terminal event following the occurrence of the nonterminal event at time $T_1 = t_1$. In general, $\lambda_{12}(t_2 \,|\, t_1)$ can depend on both t_1 and t_2. In a Markov model, however, $\lambda_{12}(t_2 \,|\, t_1) = \lambda_{12}(t_2)$ depends only on t_2. Alternatively, a semi-Markov process specifies that $\lambda_{12}(t_2 \,|\, t_1) = \lambda_{12}(t_2 - t_1)$, with the transition rate depending only on the sojourn time.

With a Markov model, $\lambda_{12}(t)$ and $\lambda_2(t)$ are of particular interest because they correspond to the hazard of T_2 given, respectively, that T_1 has or has not previously occurred. The ratio, $\lambda_{12}(t)/\lambda_2(t)$, is the explanatory hazard ratio, which is one characterization of the dependence between T_1 and T_2. When this explanatory hazard ratio is constant and equal to 1, the occurrence of T_1 has no effect on the hazard of T_2. As in Nielsen et al. (1992), one way to incorporate a dependent structure between T_1 and T_2 is through the use of a shared frailty (or random effects) model. Denoting the frailty by a random variable $\gamma > 0$ with $E(\gamma) = 1$, we define conditional transition functions analogous to equations (1)–(3) as follows:

$$\lambda_1(t_1 \,|\, \gamma) = \gamma \lambda_{01}(t_1), t_1 > 0, \quad (4)$$

$$\lambda_2(t_2 \,|\, \gamma) = \gamma \lambda_{02}(t_2), t_2 > 0, \quad (5)$$

$$\lambda_{12}(t_2 \,|\, \gamma, t_1) = \gamma \lambda_{03}(t_2), 0 < t_1 < t_2. \quad (6)$$

Conditional on γ, this is a standard Markov illness–death model and when $\gamma = 1$ is a constant, the model reduces to the usual Markov illness–death model. With λ_{03} and λ_{02} left arbitrary, we describe this frailty model as "the general model" and, in this case, the dependence of T_2 on T_1 is described both by the conditional (given γ) explanatory hazard ratio $\lambda_{03}(t_2)/\lambda_{02}(t_2)$ as well as by the common frailty γ.

The assumption, $\lambda_{03}(t_2) = \lambda_{02}(t_2)$ is of particular interest because of its relationship to some semicompeting risks models considered in the literature, and also because, in this "restricted model," the dependence of T_1 and T_2 is fully captured by γ. This simple dependence has often been postulated in bivariate survival models with frailties. Therefore, we consider a restricted model in which equation (6) becomes

$$\lambda_{12}(t_2 \,|\, \gamma, t_1) = \gamma \lambda_{02}(t_2), 0 < t_1 < t_2. \quad (7)$$

An alternative condition to equation (7) is given by

$$\begin{aligned} &\lambda_{21}(t_1 \,|\, \gamma, t_2) \\ &= \lim_{\Delta \to 0} P[T_1 \in [t_1, t_1 + \Delta) \,|\, T_1 \geqslant t_1, T_2 = t_2, \gamma]/\Delta \\ &= \gamma \lambda_{01}(t_1), 0 < t_1 < t_2. \end{aligned} \quad (8)$$

In fact, it is easy to show that assumptions (4), (5), and (7) are equivalent to assumptions (4), (5), and (8). Both equations (7) and (8) specify a type of conditional independence between the two failure types. Equation (8) is unusual because it specifies the recurrence rate conditional on a future event and so is not useful as a generative description of the process. Nonetheless, it provides a simple consequence and characterization of the model. A similar formulation conditional on future events is used in Section 3 to obtain a simple partial likelihood type estimator of a parameter in the frailty distribution for γ.

In this article, we assume that the frailty γ has a gamma distribution with mean 1, variance θ, and density $\theta^{-\frac{1}{\theta}} \gamma^{\frac{1}{\theta}-1} \exp(-\frac{\gamma}{\theta})/\Gamma(\frac{1}{\theta})$. The conditional hazards (given γ) in equations (4), (5), and (7) together with the gamma frailty distribution can be seen to correspond to a marginal model with hazards as defined in equations (1)–(3) given by

$$\lambda_1(t_1) = [1 + \theta w(t_1, t_1)]^{-1} \lambda_{01}(t_1), t_1 > 0, \quad (9)$$

$$\lambda_2(t_2) = [1 + \theta w(t_2, t_2)]^{-1} \lambda_{02}(t_2), t_2 > 0, \quad (10)$$

and

$$\lambda_{12}(t_2 \,|\, t_1) = (1 + \theta)[1 + \theta w(t_1, t_2)]^{-1} \lambda_{02}(t_2), 0 < t_1 < t_2, \quad (11)$$

where $\Lambda_{01}(t) = \int_0^t \lambda_{01}(s) ds$, $\Lambda_{02}(t) = \int_0^t \lambda_{02}(s) ds$, and

$$w(t_1, t_2) = \Lambda_{01}(t_1) + \Lambda_{02}(t_2).$$

From equations (9)–(11), it is clear that the marginal restricted model is non-Markovian unless γ is constant ($\theta = 0$).

3. Semicompeting Risk Models—The Restricted Model

Much recent literature has taken a different approach to this data structure in which a joint survivor function is assumed to exist for T_1, T_2 over the positive quadrant, but observation is restricted to the upper wedge in which $T_1 < T_2$. We adopt a similar, though distinct representation for the joint distribution of T_1, T_2.

Based on the model in equations (9)–(11), the joint survivor function of (T_1, T_2) in the upper wedge is

$$S(t_1, t_2) = (1 + \theta \Lambda_{01}(t_1) + \theta \Lambda_{02}(t_2))^{-1/\theta}, 0 < t_1 \leqslant t_2. \quad (12)$$

This survivor function can be seen to correspond to a density function of

$$f(t_1,t_2)=(\theta+1)\lambda_{01}(t_1)\lambda_{02}(t_2)[1+\theta\Lambda_{01}(t_1)+\theta\Lambda_{02}(t_2)]^{-1/\theta-2}, \tag{13}$$

in the upper wedge $0<t_1\leqslant t_2$ and a density along the line $t_1=\infty$ given by

$$f_\infty(t_2)=\lambda_{02}(t_2)(1+\theta\Lambda_{01}(t_2)+\theta\Lambda_{02}(t_2))^{-1/\theta-1}, 0<t_2. \tag{14}$$

By writing $S_1(t_1)=(1+\theta\Lambda_{01}(t_1))^{-1/\theta}$ and $S_2(t_2)=(1+\theta\Lambda_{02}(t_2))^{-1/\theta}$, we find that $S(t_1,t_2)=[S_1(t_1)^{-\theta}+S_2(t_2)^{-\theta}-1]^{-1/\theta}$ which is the same form as in Fine et al. (2001). In fact, Fine et al. (2001) interpret $S(t_1,t_2)$ as a survivor function over the whole positive quadrant and consider $S_1(t_1)$ as the marginal survivor function of T_1. They note, however, that the joint survivor function in the lower wedge cannot be estimated on the data available. In fact, there is no probability content in the region $T_2<T_1<\infty$, and this interpretation of S_1 as a marginal survivor function of T_1 is not justified. Note that in expression (12), however, Λ_{01} has a natural interpretation as an integrated hazard in the related frailty model.

We consider three distinct methods for estimation of θ. In Section 3.1, we describe methods for estimating θ due to Day et al. (1997) and Fine et al. (2001). In Section 3.2, we describe a maximum likelihood approach for the joint estimation of θ,Λ_{01} and Λ_{02}.

3.1 *Methods for Direct Estimation of θ*

Day et al.'s method. Let $N_{i2}(t)$ be the counting process for the observed terminal event for the ith individual,

$$N_{i2}(t)=I(T_{i2}\leqslant t,\delta_{i2}=1).$$

Let $Z_{i2}(t)=I(Y_{i2}\geqslant t)$ be the at risk process for the failure time event. Similarly, let $N_{i1}(t)=I(T_{i1}\leqslant t,\delta_{i1}=1)$ and $Z_{i1}(t)$ be the counting and at risk processes for the nonterminal event.

It can be verified that the model specified by equations (8)–(10) implies that

$$\lambda_{21}(t_1\,|\,T_2=t_2)=(1+\theta)\lambda_{21}(t_1\,|\,T_2\geqslant t_2),$$

where $\lambda_{21}(t_1\,|\,A)=\lim_{\Delta\to0}P(T_1\in[t_1,t_1+\Delta)\,|\,T_1\geqslant t_1,A)/\Delta$. Consider a time t_2, where a terminal event is observed to occur (i.e., a time t_2, where $dN_{\cdot2}(t_2)=1$, the subscript $\cdot$ denoting summation) and the set $R_2(t_2)=\{j:Z_{j2}(t_2)=1\}$. We can construct a partial likelihood for $(1+\theta)$ by considering $R_2(t_2)$ as comprising two samples indexed by $dN_{j2}(t_2)$, which indicates whether or not item j fails at time $t_2, j\in R_2(t_2)$. This gives rise to an estimating equation obtained by taking the partial likelihood score at t_2. Summing these equations over all t_2 gives rise to the estimating equation

$$\sum_{j=1}^n\int_0^\infty\sum_{i=1}^n\int_0^u Z_{i2}(u)\,[N_{i2}(u)-e_\theta(s;u)]\,dN_{i1}(s)dN_{j2}(u)=0,$$

where

$$e_\theta(s;u)=\frac{\sum_{\ell=1}^n N_{\ell2}(u)Z_{\ell1}(s)Z_{\ell2}(u)(1+\theta)^{N_{\ell2}(u)}}{\sum_{\ell=1}^n Z_{\ell1}(s)Z_{\ell2}(u)(1+\theta)^{N_{\ell2}(u)}}.$$

This estimating equation was also obtained by Day et al. (1997) from a different perspective and, by its construction, it can be seen to be unbiased and convex. In practice, standard software for the Cox model can be used to obtain the unique estimator $\hat\theta$ of θ. There is currently available no simple variance estimator for the asymptotic distribution $\hat\theta$, but the variance can be numerically approximated using the jackknife method.

Fine et al.'s method. Consider two independent pairs, (T_{i1},T_{i2}) and $(T_{j1},T_{j2}), i<j$, and define the concordance indicator

$$\Delta_{ij}=I\{(T_{i1}-T_{j1})(T_{i2}-T_{j2})>0\}.$$

Fine et al. (2001) noted that Δ_{ij} is only observed when $\tilde T_{ij1}<\tilde T_{ij2}<\tilde C_{ij}$, where $\tilde T_{ij1}=T_{i1}\wedge T_{j1},\tilde T_{ij2}=T_{i2}\wedge T_{j2},\tilde C_{ij}=C_i\wedge C_j$. Define $\tilde S_{ij}=\min(\tilde T_{ij1},\tilde T_{ij2},\tilde C_{ij}),\tilde R_{ij}=\min(\tilde T_{ij2},\tilde C_{ij})$. Because the predictive hazard ratio equals $(1+\theta)$ in the upper wedge and the expectation of Δ_{ij} conditional on $\tilde T_{ij1}<\tilde T_{ij2}$ is $\frac{1+\theta}{2+\theta}$. Let $D_{ij}=I(\tilde T_{ij1}<\tilde T_{ij2}<\tilde C_{ij})$, the estimation equation based on the concordance indicator can be constructed as

$$U(\theta)=\sum_{i<j}W(\tilde S_{ij},\tilde R_{ij})D_{ij}\left\{\Delta_{ij}-\frac{1+\theta}{2+\theta}\right\},$$

where the weight function $W(u,v)$ is a random function satisfying $\sup_{u,v}\,|\,W(u,v)-\tilde W(u,v)\,|\,\to0$ in probability and $\tilde W$ is a deterministic and bounded function in the support of $(\tilde S_{ij},\tilde R_{ij})$. A variance estimator can be obtained by applying the theory of U-statistics.

3.2 *Maximum Likelihood Joint Estimation*

The likelihood based on the observed data $O_i=(Y_{i2},\delta_{i2},Y_{i1},\delta_{i1})$ and given γ_i, denoted by $L(O_i,\gamma_i)$, is

$$\gamma_i^{\delta_{i1}+\delta_{i2}}\lambda_{01}(Y_{i1})^{\delta_{i1}}\lambda_{02}(Y_{i2})^{\delta_{i2}}\exp\{-\gamma_i[\Lambda_{01}(Y_{i1})+\Lambda_{02}(Y_{i2})]\}.$$

Averaging over the distribution of γ_i and taking a product over i gives the likelihood function L as

$$\begin{aligned}&\prod_{i=1}^n\lambda_{01}(Y_{i1})^{\delta_{i1}}\lambda_{02}(Y_{i2})^{\delta_{i2}}(1+\theta)^{\delta_{i1}\delta_{i2}}\\&\{1+\theta[\Lambda_{01}(Y_{i1})+\Lambda_{02}(Y_{i2})]\}^{-1/\theta-\delta_{i1}-\delta_{i2}},\end{aligned} \tag{15}$$

as can also be obtained directly using the expressions in (12)–(14).

Let $t_{r1},t_{r2},\ldots,t_{rm}$ be the ordered distinct relapse times and $t_{d1},t_{d2},\ldots,t_{df}$ be the ordered distinct failure times. We let $\lambda_{0R}=(\lambda_{011},\lambda_{012},\ldots,\lambda_{01m})^T$ and $\lambda_{0D}=(\lambda_{021},\lambda_{022},\ldots,\lambda_{02f})^T$, where $\lambda_{01j}=d\Lambda_{01}(t_{rj}), j=1,\ldots,m$ and $\lambda_{02j}=d\Lambda_{02}(t_{dj}), j=1,\ldots,f$. Let $\eta=(\theta,\lambda_{0R},\lambda_{0D})$ be the $m+f+1$ dimensional vector of parameters. The maximum likelihood estimates are obtained by taking the derivative of the log likelihood with respect to η and solving the corresponding equations.

Table 1
Simulation results for estimation of $(\theta, \Lambda_{01}, \Lambda_{02})$ *in the semicompeting risks model without covariates*

θ	Par		Day et al. (1997)	Fine et al. (2001)	NPMLE	Par		Day et al.	Fine et al.	NPMLE
				$n = 200$					$n = 400$	
1	θ	Bias	0.051	0.049	−0.014	θ	Bias	0.018	0.021	−0.010
		SD	0.303	0.316	0.281		SD	0.209	0.215	0.203
		ESE	0.304	0.318	0.282		ESE	0.210	0.217	0.201
		CP	0.964	0.970	0.968		CP	0.960	0.976	0.946
	$\Lambda_{01}(1)$	Bias	0.005	0.004	0.001	$\Lambda_{01}(1)$	Bias	0.004	0.003	0.001
		SD	0.068	0.069	0.050		SD	0.056	0.058	0.038
		ESE	0.069	0.070	0.049		ESE	0.057	0.059	0.039
		CP	0.960	0.964	0.944		CP	0.958	0.966	0.950
	$\Lambda_{02}(1)$	Bias	0.008	0.006	0.002	$\Lambda_{02}(1)$	Bias	0.005	0.004	0.002
		SD	0.090	0.092	0.070		SD	0.064	0.062	0.039
		ESE	0.091	0.093	0.071		ESE	0.063	0.061	0.038
		CP	0.954	0.958	0.954		CP	0.948	0.946	0.946
0.5	θ	Bias	0.016	0.015	−0.019	θ	Bias	0.004	0.003	−0.006
		SD	0.209	0.223	0.195		SD	0.135	0.150	0.095
		ESE	0.210	0.225	0.196		ESE	0.136	0.152	0.094
		CP	0.952	0.970	0.956		CP	0.954	0.978	0.948
	$\Lambda_{01}(1)$	Bias	0.003	0.003	0.001	$\Lambda_{01}(1)$	Bias	0.002	0.002	0.001
		SD	0.034	0.040	0.036		SD	0.028	0.030	0.024
		ESE	0.033	0.039	0.037		ESE	0.029	0.031	0.026
		CP	0.948	0.946	0.952		CP	0.952	0.954	0.956
	$\Lambda_{02}(1)$	Bias	0.005	0.006	0.003	$\Lambda_{02}(1)$	Bias	0.003	0.003	0.002
		SD	0.046	0.054	0.050		SD	0.036	0.048	0.032
		ESE	0.047	0.053	0.051		ESE	0.037	0.049	0.031
		CP	0.952	0.946	0.950		CP	0.956	0.958	0.948
2	θ	Bias	0.053	0.045	−0.025	θ	Bias	0.026	0.021	−0.019
		SD	0.513	0.511	0.473		SD	0.355	0.360	0.335
		ESE	0.512	0.513	0.475		ESE	0.354	0.362	0.337
		CP	0.947	0.976	0.964		CP	0.947	0.970	0.962
	$\Lambda_{01}(1)$	Bias	0.008	0.005	0.002	$\Lambda_{01}(1)$	Bias	0.006	0.003	0.002
		SD	0.056	0.054	0.050		SD	0.036	0.034	0.028
		ESE	0.057	0.053	0.049		ESE	0.037	0.035	0.027
		CP	0.952	0.946	0.946		CP	0.952	0.954	0.948
	$\Lambda_{02}(1)$	Bias	0.012	0.009	0.005	$\Lambda_{02}(1)$	Bias	0.007	0.004	0.003
		SD	0.118	0.116	0.106		SD	0.065	0.060	0.048
		ESE	0.119	0.117	0.107		ESE	0.066	0.061	0.050
		CP	0.954	0.952	0.954		CP	0.954	0.956	0.960

A simple algorithm is available to compute the maximum likelihood estimations of θ, Λ_{01}, and Λ_{02}. Further, these estimators, under relatively mild conditions, are asymptotically normal with a covariance matrix that is easily estimated. These results are described in Section 4 in a much more general setting.

3.3 *Comparison of Estimates of* θ

We report on a set of simulations to compare the performance of three methods described above for the estimation of θ in the restricted model. We generated n observations from the models (4), (5), and (7), where γ was simulated from a gamma distribution with mean 1 and shape parameter $1/\theta$ and both $\lambda_{01}(t)$ and $\lambda_{02}(t)$ were set to be constant 1. The censoring times were generated from a uniform distribution on (1, 3). We report results from 500 replications for $\theta = 0.5,\ 1.0,\ 2.0$ and for $n = 200,\ 400$. Table 1 reports average bias (Bias), the empirical standard deviation (SD), the average value of the estimated standard errors (ESEs), and the coverage probability (CP) of the nominal 95% confidence intervals for estimates of $\theta, \Lambda_{01}(1)$, and $\Lambda_{02}(1)$. In comparison with the other two methods, the NPMLE performs very favorably both in terms of bias and mean squared error. Furthermore, the ESEs agree well with the sample standard deviations and the coverage probabilities are accurate.

4. More General Models and Covariance Analysis

We now return to the more general model described in equations (4)–(6) and consider the incorporation of covariates x in the model.

Various regression models could be considered for the conditional (on γ) or marginal transition intensity functions defined earlier. For example, Cox type models could be specified for either and maximum likelihood methods extended for joint estimation of the regression parameters or as $\theta, \Lambda_{01}, \Lambda_{02}$, and Λ_{03}.

Incorporating covariates x in the conditional hazards, analogous to equations (4), (5), and (7), we have

$$\lambda_1(t_1 \mid \gamma, x) = \gamma\lambda_{01}(t_1)\exp\{\beta_1 x\}, t_1 > 0, \tag{16}$$

$$\lambda_2(t_2 \mid \gamma, x) = \gamma\lambda_{02}(t_2)\exp\{\beta_2 x\}, t_2 > 0, \tag{17}$$

and

$$\lambda_{12}(t_2 \mid \gamma, t_1, x) = \gamma\lambda_{03}(t_2)\exp\{\beta_3 x\}, 0 < t_1 < t_2. \tag{18}$$

Again we have assumed that γ has a gamma distribution and is distributed independently of x.

An alternative approach to regression analysis would incorporate the covariates in the marginal hazards, analogous to equations (9)–(11). We have

$$\lambda_1(t_1 \mid x) = [1 + \theta w(t_1, t_1)]^{-1}\lambda_{01}(t_1)\exp\{\beta_1 x\}, t_1 > 0, \tag{19}$$

$$\lambda_2(t_2 \mid x) = [1 + \theta w(t_2, t_2)]^{-1}\lambda_{02}(t_2)\exp\{\beta_2 x\}, t_2 > 0, \tag{20}$$

and

$$\lambda_{12}(t_2 \mid x, t_1) = (1+\theta)[1 + \theta w(t_1, t_2)]^{-1}\lambda_{03}(t_2)\exp\{\beta_3 x\}, \tag{21}$$

where $0 < t_1 < t_2$ and $w(s,t)$ has a more general form than before given by

$$w(s,t) = \Lambda_{01}(s) + \Lambda_{02}(s) + [\Lambda_{03}(t) - \Lambda_{03}(s)], 0 < s \leqslant t.$$

These correspond to two different modeling strategies and it is important to note that the regression parameters in equations (16)–(18) and (19)–(21) have different interpretations, corresponding to regression effects on conditional intensities and marginal intensities, respectively.

In what follows, we focus on the conditional approach. Under the models (16)–(18), the likelihood based on the data $O_i = (Y_{i1}, \delta_{i1}, Y_{i2}, \delta_{i2}, x_i)$ is

$$\begin{aligned} L = \prod_{i=1}^{n} & \lambda_{01}(Y_{i1})^{\delta_{i1}}\lambda_{02}(Y_{i2})^{\delta_{i2}(1-\delta_{i1})}\lambda_{03}(Y_{i2})^{\delta_{i1}\delta_{i2}} \\ & \times \exp\left[\delta_{i1}\beta_1^T x_i + \delta_{i2}(1-\delta_{i1})\beta_2^T x_i + \delta_{i1}\delta_{i2}\beta_3^T x_i\right] \\ & \times (1+\theta)^{\delta_{i1}\delta_{i2}}\Big\{1 + \theta[\Lambda_{01}(Y_{i1})e^{\beta_1^T x_i} + \Lambda_{02}(Y_{i1})e^{\beta_2^T x_i} \\ & \qquad + \Lambda_{03}(Y_{1i}, Y_{2i})e^{\beta_3^T x_i}]\Big\}^{-1/\theta - \delta_{i1} - \delta_{i2}}, \end{aligned}$$

where $\Lambda_{03}(s,t) = \Lambda_{03}(t) - \Lambda_{03}(s)$. Let $\eta = (\theta, \beta_1, \beta_2, \beta_3, d\Lambda_{01}, d\Lambda_{02}, d\Lambda_{03})$. Further let $t_{r1}, t_{r2}, \ldots, t_{rm}$ be the ordered distinct relapse times and $t_{d1}, t_{d2}, \ldots, t_{df}$ be the ordered distinct failure times without relapse, and $t_{rd1}, \ldots, t_{rdg}$ be the ordered distinct failure times following relapse. We let $\lambda_{0R} = (\lambda_{011}, \lambda_{012}, \ldots, \lambda_{01m})^T$, $\lambda_{0D} = (\lambda_{021}, \lambda_{022}, \ldots, \lambda_{02f})^T$, and $\lambda_{0RD} = (\lambda_{031}, \lambda_{032}, \ldots, \lambda_{03g})^T$, where $\lambda_{01j} = d\Lambda_{01}(t_{rj}), j = 1, \ldots, m$, $\lambda_{02j} = d\Lambda_{02}(t_{dj}), j = 1, \ldots, f$, and $\lambda_{03j} = d\Lambda_{03}(t_{rdj}), j = 1, \ldots, g$. Define $d_{t_{rj}}, d_{t_{dj}}$, and d_{rdj} as the number of relapse times at t_{rj}, the number of relapse free deaths at t_{dj}, and the number of deaths with relapse at t_{rdj}, respectively. The score vector is denoted by $U(\eta) = (U_1, U_2^T, U_3^T, U_4^T, U_5^T, U_6^T, U_7^T)^T$, where the respective components of U correspond to partial derivatives with respect to $\theta, \beta_1, \beta_2, \beta_3, \lambda_{0R}, \lambda_{0D}, \lambda_{0RD}$. Define $A_{i1} = \Lambda_{01}(Y_{i1})\exp(\beta_1^T x_i)$, $A_{i2} = \Lambda_{02}(Y_{i1})\exp(\beta_2^T x_i)$, $A_{i3} = \Lambda_{03}(Y_{1i}, Y_{2i})e^{\beta_3^T x_i}$, $A_i = A_{i1} + A_{i2} + A_{i3}$, $B_i = 1/\theta + \delta_{i1} + \delta_{i2}$. Following an approach similar to that of Andersen et al. (1997), we have

$$\begin{aligned} U_1 &= \sum_{i=1}^{n}\left\{\frac{\delta_{i1}\delta_{i2}}{1+\theta} + \frac{1}{\theta^2}\log(1+\theta A_i) - \frac{B_i A_i}{1+\theta A_i}\right\} \\ U_2 &= \sum_{i=1}^{n}\left\{\delta_{i1}x_i - B_i\frac{x_i\theta A_{i1}}{1+\theta A_i}\right\} \\ U_3 &= \sum_{i=1}^{n}\left\{(1-\delta_{i1})\delta_{i2}x_i - B_i\frac{x_i\theta A_{i2}}{1+\theta A_i}\right\} \\ U_4 &= \left\{\sum_{i=1}^{n}\delta_{i1}\delta_{i2}x_i - B_i\frac{x_i\theta A_{i3}}{1+\theta A_i}\right\}. \end{aligned}$$

Finally, the jth elements of U_5, U_6, and U_7 are

$$\begin{aligned} U_{5j} &= d_{t_{rj}}/\lambda_{01j} - \sum_{i=1}^{n}\frac{I(Y_{i1} \geqslant t_{rj})B_i\theta\exp\left(\beta_1^T x_i\right)}{1+\theta A_i}, \\ & j = 1, \ldots, m, \\ U_{6j} &= d_{t_{dj}}/\lambda_{02j} - \sum_{i=1}^{n}\frac{I(Y_{i1} \geqslant t_{dj})B_i\theta\exp\left(\beta_2^T x_i\right)}{1+\theta A_i}, \\ & j = 1, \ldots, f. \\ U_{7j} &= d_{t_{rdj}}/\lambda_{03j} - \sum_{i=1}^{n}\frac{I(Y_{i2} \geqslant t_{rdj} > Y_{i1})B_i\theta\exp\left(\beta_3^T x_i\right)}{1+\theta A_i}, \\ & j = 1, \ldots, g. \end{aligned}$$

Our numerical approach to solve $U(\eta) = 0$ converges quickly and can be summarized as follows:

(1) Let $\theta^{(0)}, \beta_j^{(0)}, \Lambda_{0j}^{(0)}(u), j = 1, 2, 3$ be initial estimates. Typically, we can set $\theta^{(0)} = 1, \beta_j^{(0)} = 0$, and let $\Lambda_{0j}^{(0)}(u)$, be the Nelson–Aalen type estimate of the respective cumulative hazard functions.
(2) Given $\Lambda_{0j}^{(0)}(u), j = 1, 2, 3$, update estimates of θ, β_j to $\theta^{(1)}, \beta_j^{(1)}, j = 1, 2, 3$ from U_1, U_2, U_3, U_4.
(3) Given $\theta^{(1)}, \beta_j^{(1)}, j = 1, 2, 3$, solve the equations $U_5 = 0, U_6 = 0, U_7 = 0$ for updated estimates $\Lambda_{0j}^{(1)}(u), j = 1, 2, 3$.
(4) Update the initial estimates in (1) and repeat steps (2)–(4) until the estimates converge.

To establish the asymptotic properties, we specify the following conditions.

(C1) The variance parameter θ_0 lies in a known interval $[0, K], K < \infty$. The regression parameter $(\beta_{01}, \beta_{02}, \beta_{03})$ belongs to the interior of a compact set. The cumulative baseline hazards $\Lambda_{01}, \Lambda_{02}$, and Λ_{03} are strictly increasing and continuously differentiable on $[0, \tau]$ for given $\tau < \infty$.
(C2) The observations $(Y_{i2}, \delta_{i2}, Y_{i1}, \delta_{i1}, x_i), i = 1, \ldots, n$ are identically and independently distributed. Conditional on $\mathbf{x}$, the right censoring C is independent of T_2 and T_1.
(C3) The covariate $\mathbf{x}$ is bounded with probability one, and if b is a constant vector such that $b^T x = 0$ with probability one, then $b = 0$.

Table 2
Simulation results for covariance analysis in the general model

	$n = 250$					$n = 400$				
θ	Par	Bias	SD	ESE	CP	Par	Bias	SD	ESE	CP
1	β_1	−0.043	0.869	0.854	0.942		0.032	0.684	0.690	0.954
	β_2	0.006	0.852	0.840	0.944		0.005	0.690	0.711	0.962
	β_3	0.038	0.978	0.956	0.940		0.007	0.837	0.845	0.958
	$\Lambda_{01}(1)$	−0.024	0.258	0.251	0.944		−0.014	0.195	0.187	0.942
	$\Lambda_{02}(1)$	−0.021	0.232	0.220	0.942		−0.013	0.194	0.190	0.946
	$\Lambda_{03}(1)$	−0.049	0.624	0.618	0.942		−0.038	0.508	0.495	0.944
	θ	0.019	0.238	0.241	0.952		−0.014	0.197	0.202	0.954
0.5	β_1	0.028	0.735	0.739	0.948		0.026	0.614	0.616	0.950
	β_2	0.083	0.772	0.766	0.946		−0.025	0.582	0.588	0.954
	β_3	−0.054	0.867	0.870	0.952		−0.007	0.693	0.701	0.956
	$\Lambda_{01}(1)$	−0.019	0.226	0.217	0.946		0.010	−0.160	0.156	0.944
	$\Lambda_{02}(1)$	−0.017	0.226	0.218	0.940		0.014	−0.171	0.176	0.952
	$\Lambda_{03}(1)$	−0.042	0.533	0.521	0.938		0.035	−0.435	0.430	0.946
	θ	−0.015	0.199	0.204	0.952		0.009	0.149	0.157	0.956
2	β_1	0.056	1.065	1.072	0.956		−0.010	0.858	0.863	0.954
	β_2	0.126	1.059	1.066	0.958		−0.032	0.855	0.869	0.960
	β_3	0.021	1.310	1.314	0.952		0.010	0.957	0.970	0.962
	$\Lambda_{01}(1)$	−0.022	0.329	0.310	0.938		−0.018	0.268	0.256	0.946
	$\Lambda_{02}(1)$	−0.023	0.312	0.302	0.942		−0.019	0.261	0.254	0.944
	$\Lambda_{03}(1)$	−0.065	0.775	0.767	0.946		−0.050	0.578	0.569	0.942
	θ	0.021	0.243	0.238	0.948		0.013	0.173	0.182	0.958

These regularity conditions are similar to those of Murphy (1995) and Parner (1998). As in Murphy (1995), we consider discrete versions of the baseline hazard functions with jumps only at the distinct relapse times and death times as described above. Let $A(\eta) = -n^{-1}\partial U(\eta)/\partial\eta^T$. The second partial derivatives are listed in Web Appendix A. Let $\hat{\eta}$ be the estimate of η.

Let $\alpha^T = (\theta, \beta_1^T, \beta_2^T, \beta_3^T)$ and let $\Lambda_0 = (\Lambda_{01}, \Lambda_{02}, \Lambda_{03})$. The following theorem states our main results.

THEOREM 1. *Under conditions (C1)–(C3),*

(i) The parameters α and Λ_0 are identifiable.

(ii) With probability one,

$$|\,\hat{\alpha} - \alpha_0\,| + \sup_{t\in[0,\tau)} |\,\hat{\Lambda}_0 - \Lambda_0\,| \to 0.$$

That is $\hat{\alpha}$ and $\hat{\Lambda}_0$ are strongly consistent.

(iii) $\sqrt{n}(\hat{\alpha} - \alpha_0, \hat{\Lambda}_0 - \Lambda_0)$ converges weakly to a zero-mean Gaussian process in the metric space $R^{3p+1} \times \{\ell^\infty[0,\tau]\}^3$, where $\ell^\infty[0,\tau]$ is the linear space consisting of all bounded functions in $[0,\tau]$ and is equipped with the supremum norm. Furthermore, $\hat{\alpha}$ is asymptotically efficient; equivalently, its asymptotic variance attains the semiparametric efficiency bound for α.

The proof of Theorem 1 follows in outline that in Murphy (1995) and is given in Web Appendix B. Let $\hat{\mathcal{I}}$ be the observed information matrix for η. For moderate sample sizes, the estimate of the asymptotic variance of $(\hat{\theta}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ can be estimated using the relevant submatrix of $\hat{\mathcal{I}}^{-1}$. Similarly, the asymptotic covariance function of $\hat{\Lambda}_{01}, \hat{\Lambda}_{02}$, and $\hat{\Lambda}_{03}$ is obtained from the submatrices of $\hat{\mathcal{I}}^{-1}$ associated with $\lambda_{01}, \lambda_{02}$, and λ_{03}, respectively. For large sample sizes, a profile likelihood approach could provide an alternative approach for estimating asymptotic variances (e.g., Murphy and van der Vaart, 2000).

5. Simulation Study

We report here a second set of simulations designed to evaluate the proposed method for the covariance analysis in the general model.

From the model (16)–(18), n observations in which γ had the same values as before were generated. Let $\lambda_{01}(t) = \lambda_{02}(t) = 2\exp(-t)I(0 \leqslant t \leqslant 3) + 2\exp(-3)I(t > 3)$ and $\lambda_{03}(t) = 2\lambda_{01}(t)$. We report results with one covariate, X, having a uniform distribution between 0 and 0.5. We consider $\beta_1 = 1, \beta_2 = 1$, $\beta_3 = 0.5$, and $n = 250$ and 400. The censoring time was simulated from a mixture distribution with probability 0.5 from a uniform distribution on (1.5,3) and probability 0.5 from a point mass at 3. The results obtained from 500 replications are summarized in Table 2. We can see that the NPMLE method performs well for the regression parameters $\beta_1, \beta_2, \beta_3$; the baseline cumulative hazards $\Lambda_{01}, \Lambda_{02}, \Lambda_{03}$; and the frailty parameter θ. The biases are small, the ESEs agree well with the sample SDs, and the coverage probabilities are close to the nominal level. These and similar simulations suggest that the NPMLE method provides an efficient and feasible way of analyzing these illness–death (or semicompeting risks) models with practical sample sizes.

6. Application to Nasopharyngeal Cancer Data

The data come from a randomized clinical trial on nasopharyngeal cancer conducted between September 1997 and May 2003 by the National Cancer Center, Singapore (Wee et al., 2005). Among 221 patients eligible for entry into the trial, 110 patients were randomly assigned to receive radiotherapy

Table 3
Analysis of nasopharyngeal cancer data—the restricted model

Covariates	Time to relapse			Time to death		
	Estimate	SE	*P*-value	Estimate	SE	*P*-value
Treatment (x_1)	−1.064	0.2620	<0.001	−0.980	0.248	<0.001
tsize 4 vs tsize 1 (x_2)	1.811	0.352	<0.001	1.592	0.328	<0.001
tsize 3 vs tsize 1 (x_3)	0.631	0.354	0.074	−0.027	0.301	0.930
tsize 2 vs tsize 1 (x_4)	0.484	0.376	0.198	−0.165	0.345	0.632
nstage 3 vs nstage 0 (x_5)	3.220	0.316	<0.001	2.608	0.223	<0.001
nstage 2 vs nstage 0 (x_6)	1.341	0.365	<0.001	0.959	0.330	0.004
nstage 1 vs nstage 0 (x_7)	0.647	0.470	0.169	−0.296	0.456	0.516
θ	7.037	2.252	0.002			

Table 4
Analysis of nasopharyngeal cancer data—the Markov model

Covariates	Time to relapse			Time to death without relapse			Time to death after relapse		
	Estimate	SE	*P*-value	Estimate	SE	*P*-value	Estimate	SE	*P*-value
x_1	−0.714	0.247	0.004	0.267	0.481	0.579	−0.663	0.302	0.028
x_2	1.071	0.585	0.067	2.489	1.089	0.022	−0.463	0.523	0.377
x_3	0.465	0.547	0.395	1.475	1.093	0.177	−0.858	0.499	0.086
x_4	0.336	0.522	0.520	1.257	1.109	0.257	−0.639	0.486	0.188
x_5	1.303	0.639	0.041	1.899	0.814	0.020	−0.425	0.536	0.428
x_6	0.706	0.611	0.248	0.821	0.829	0.322	−0.376	0.478	0.432
x_7	0.577	0.608	0.343	0.163	0.931	0.861	−0.576	0.540	0.286

(RT) alone, whereas the other 111 received concurrent cisplatin with RT, followed by adjuvant chemotherapy comprising a combination of cisplatin plus fluorouracil. After treatment, patients may experience a cancer recurrence or death or a cancer recurrence followed by death. By the end of the study, 21 patients had cancer recurrence and then died, 15 patients died without cancer recurrence, and 8 patients had cancer recurrence but did not die in the CRT group; in the RT group, 47 patients had cancer recurrence and then died, 8 patients died without cancer recurrence, and 3 patients had cancer recurrence but did not die.

We first fit the restricted model to compare the CRT group and the RT group. Other covariates considered were tumor size and nodal status. The results are summarized in Table 3. After adjusting for the other covariates, the patients in the CRT group tend to have a lower rate of cancer relapse and death than those in the RT group. The results confirm the significant survival benefit of the CRT treatment over the RT treatment. As expected, it is also found that higher rates of cancer relapse and death are associated both with increased tumor size and nodal status. The frailty parameter θ is estimated to be 7.037 with an ESE 2.252($P < 0.002$). According to this estimate, a patient who is known to experience cancer relapse at time t is expected to have about eight times the risk of dying at subsequent times as compared to a patient who has not experienced a relapse of cancer by time t. This indicates that relapse is strongly related to death, which is consistent with the large percentages of deaths among patients with relapse in both the CRT group and the RT group (21/29 and 47/50, respectively).

We also fit the marginal Markov model and the general frailty model. The Markov model is analyzed using partial likelihood methods (Kalbfleisch and Prentice, 2002, p. 270). The general model is fitted by maximum likelihood as discussed earlier, and standard errors are estimated by inverting the observed information matrix. The results are summarized in Tables 4 and 5, respectively. In both the Markov model and the general frailty model, treatment effect (x_1) is significant on time to relapse and time to death after relapse, but not time to death without relapse. For time to death without relapse, the regression coefficient is positive and in the opposite direction of the other two. Actually, this is consistent with the fact that more patients died without cancer recurrence in the CRT (treatment) group (15) than the RT (control) group (8). Thus, both the Markov model and the general frailty model provide a more realistic picture of data. In terms of other covariates, for time to relapse, all three models give similar results except for x_6 but for time to death without relapse and time to death after relapse, the Markov model and the general frailty model give similar results that are more realistic than those of the restrictive model. The general model yields a much smaller estimate for the variance of the frailty 2.02, indicating that some of the dependence of T_2 on T_1 is captured by the unequal baseline hazards. The likelihood ratio statistic for independence ($\theta = 0$) is 5.06 for the general model and 9.82 for the restricted model, indicating fairly strong evidence of a correlation between T_1 and T_2 in both cases. To check the assumptions made in the restricted model (as compared to the general model), we wish to test for $\beta_2 = \beta_3$ and $\lambda_{02} = \lambda_{03}$. For this purpose, define

$$\psi^T = \left(\beta_2^T - \beta_3^T, \int_0^\tau w(t)[\Lambda_{02}(t) - \Lambda_{03}(t)]dt\right),$$

Table 5
Analysis of nasopharyngeal cancer data—the general model

	Time to relapse			Time to death without relapse			Time to death after relapse		
Covariates	Estimate	SE	*P*-value	Estimate	SE	*P*-value	Estimate	SE	*P*-value
x_1	−0.822	0.337	0.015	0.144	0.531	0.786	−0.736	0.348	0.034
x_2	1.567	0.555	0.005	3.142	0.723	<0.001	0.165	0.518	0.750
x_3	0.545	0.535	0.308	1.718	0.831	0.039	−0.827	0.583	0.156
x_4	0.452	0.550	0.411	1.465	0.871	0.093	−0.559	0.621	0.368
x_5	2.217	0.665	0.001	3.011	0.780	<0.001	0.283	0.497	0.570
x_6	0.966	0.543	0.075	1.167	0.781	0.135	−0.040	0.530	0.940
x_7	0.701	0.644	0.276	0.237	1.014	0.815	−0.659	0.714	0.356
θ	2.021	0.933	0.030						

where $w(.)$ is a nonnegative weight function. We chose $w \equiv 1$. A consistent estimator of ψ is

$$\hat{\psi}^T = \left(\hat{\beta}_2^T - \hat{\beta}_3^T, \int_0^\tau w(t)[\hat{\Lambda}_{02}(t) - \hat{\Lambda}_{03}(t)]dt\right),$$

and an estimate of its covariance matrix, $\hat{V}_\psi$, is obtained by the delta method. When the restricted model holds, $\hat{\psi}^T \hat{V}_\psi^{-1} \hat{\psi}$ follows a χ^2_{p+1} distribution, where p is the number of covariates. In our example, the test statistic is 36.71 and yields a *P*-value of $<0.001(\chi^2_8)$, indicating strong evidence against the restricted model, at least as measured by the components of ψ. An alternative approach would consider models that relate the two baseline hazards, such as $\lambda_{03}(t) = \exp(\xi_0)\lambda_{02}(t)$ or $\lambda_{03}(t) = \exp(\xi_0 + \xi_1 t)\lambda_{02}(t)$ and the equality of two baseline hazards corresponds to $\xi_0 = 0$ or $\xi_0 = \xi_1 = 0$, which can be tested using the usual likelihood ratio statistic. In fact, models of this sort, which provide descriptions of the hazard ratios are good compromises between the general and restricted models.

7. Discussion

The illness–death model with a gamma frailty for semicompeting risks data concisely describes the data structure and easily incorporates covariance analysis of semicompeting risks data, which has rarely been addressed in the literature. Within the proposed model, the maximum likelihood estimator can be obtained by a simple and fast numerical algorithm, its asymptotic properties are established, and it admits simple estimates of asymptotic variances and standard errors. Simulation studies indicate good small sample performance.

The gamma frailty is used for its mathematical convenience and the model can also be generalized to allow different effects of frailty on the two events as in Liu, Wolfe, and Huang (2004). Note that other distributions could also be used for the frailty such as the log-normal or positive stable distribution. It would also be of interest to consider different frailty distributions and model checking procedures (Glidden, 1997).

In this article, we have adopted a conditional (on the frailty γ) regression model for covariance analysis. A marginal regression model can be developed and estimated in a similar way. Our main analysis focused on a situation in which the conditional intensities satisfy $\lambda_{02} = \lambda_{03}$ because this assumption leads to a marginal model for semicompeting risks that has been studied in the literature. As discussed in the general model and also illustrated in the application, however, the methods we propose do allow different baseline hazards (i.e., $\lambda_{02} \neq \lambda_{03}$).

8. Supplementary Materials

Web Appendices A and B referenced in Section 4 are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

Acknowledgements

We are grateful to the editor, associate editor, and a referee for their helpful comments, which have greatly improved the manuscript. We also acknowledge Dr Joseph Wee at National Cancer Center of Singapore for providing data from the nasopharyngeal cancer study. This research is partially supported by the grants from National Medical Research Council of Singapore (NMRC/1140/2007) and National University of Singapore (R-155-000-075-112).

References

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

Andersen, P. K., Klein, J. P., Knudsen, K. M., and Palacios, R. T. (1997). Estimation of variance in Cox's regression model with shared gamma frailties. *Biometrics* **53,** 1475–1484.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application to epidemiological studies of familiar tendency in chronic disease epidemiology. *Biometrika* **65,** 141–151.

Day R., Bryant, J., and Lefkopoulou, M. (1997). Adaptation of bivariate frailty models for prediction, with application to biological markers as prognostic factors. *Biometrika* **84,** 45–56.

Fine, J. P., Jiang, H., and Chappell, R. (2001). On semi-competing risks data. *Biometrika* **88,** 907–939.

Fix, E. and Neyman, J. (1951). A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology* **23,** 205–241.

Glidden, D. V. (1997). Pairwise dependence diagnostics for clustered failure-time data. *Biometrika* **94,** 371–385.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Hoboken, New Jersey: Wiley.

Liu, L., Wolfe, R. A., and Huang, X. (2004). Shared frailty model for recurrent events and a terminal event. *Biometrics* **60,** 747–756.

Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Annals of Statistics* **23,** 182–198.

Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association* **95,** 449–465.

Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sørensen, T. I. A. (1992). A counting process approach to maximum likelihood

estimation in frailty models. *Scandinavian Journal of Statistics* **19,** 25–43.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B* **44,** 414–422.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Annals of Statistics* **26,** 183–214.

Peng, L. and Fine, J. P. (2007). Regression modeling of semi-competing risks data. *Biometrics* **63,** 96–108.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34,** 541–554.

Sverdrup, E. (1965). Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health. *Skand. Aktuarietietidskr* **48,** 184–211.

Wang, W. (2003). Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society, Series B* **65,** 257–273.

Wee, J., Tan, E. H., Tai, B. C., Wong, H. B., Leong, S. S., Tan, T., Chua, E. T., Yang, E., Lee, K. M., Fong, K. W., Khoo, H. S., Tan, H. S. K., Lee, K. S., Loong, S., Sethi, V., Chua, E. J., and Machin, D. (2005). Randomized trial of radiotherapy versus concurrent chemoradiotherapy followed by adjuvant chemotherapy in patients with American Joint Committee on Cancer/International Union Against Cancer Stage III and IV nasopharyngeal cancer of the endemic variety. *Journal of Clinical Oncology* **23,** 6730–6738.

Received December 2008. *Revised August* 2009.
Accepted August 2009.