MODELLING MULTIVARIATE FAILURE TIME DATA USING ADDITIVE RISK FRAILTY MODEL

CHONG YAN-CI, ELIZABETH

(B.Sc.(Hons.), NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

OF SCIENCE

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

2012

Acknowledgements

I would first like to thank my supervisors, Dr Xu Jinfeng and A/Prof Tai Bee Choo, and also our research advisor, Prof Jack Kalbfleisch. Your invaluable advice and support in the last five years have encouraged and guided me immeasurably. Thank you for inspiring me as a researcher and all the life skills I have picked up along the way will stay with me. Thank you for the opportunity to participate in ISCB 2011.

Thanks also to the rest of the research team, Gek Hsiang and Zhaojin. The many fortnightly meetings we have had have been interesting and the sharing of research areas has been an eye-opener. Thank you for all the help provided. To Dr Hu Tao, I am grateful for your valuable and timely assistance in the programming aspect of my research and pointing me in a feasible direction.

A deepfelt thanks and gratitude to all my friends and family for supporting me in this "marathon". To my parents, for their constant concern, prayers and physical support in pushing me to keep going. To my brother, sister and friends, for all your friendship and encouragement.

Special thanks to my husband, Randolph, for all the love and support, not just through this PhD, but for the last 10 years. Thank you for all the "tough love", while hard to swallow at times, has enabled me to come to this point of

completion. Your sacrifices, especially the ones made in the last few months, are deeply appreciated.

Last, but definitely not least, I wish to thank God, for empowering and strengthening me. Because of Him, all things have been made possible.

Contents

Summary

Multivariate failure time data arise when two or more distinct failures are recorded on an individual. We consider competing and semi-competing risks data, involving failures of different types. The latter occurs when a terminal event censors a non-terminal event but not vice-versa. The proportional hazards model is commonly used to examine relative risk. As a viable alternative to the proportional hazards model, the additive risks model examines excess risk and provides a flexible tool of modeling multivariate failure time data. We propose a class of additive risk models for the analysis of competing risks and semi-competing risks data. In all cases, we investigate the theoretical and numerical properties of the estimators. Simulations were conducted to assess the performance of the proposed models.

First, we consider the additive risk approach for competing risks data by modeling both the cause-specific and subdistribution hazards. Simulation results show that estimation is fairly accurate with little bias. We also apply our method to a real dataset on prostate cancer and analyse treatment effects of high-dose versus low-dose diethylstilbestrol (DES) on the outcome of interest (cancer death) and competing risks endpoints (cardiovascular death and other causes of death), while accounting for other covariates. Results indicate increased survival chances from cancer death for patients receiving high-dose DES in both cause-specific hazard and subdistribution hazard models.

Secondly, we suggest an additive risk frailty model for semi-competing risks data. Frailties are used to model the dependence between the terminal and non-terminal events and covariate effects are examined by excess risk given the frailty. Splines are used to model the conditional baseline hazard nonparametrically. Simulations indicate that estimates have about 10% bias for moderate sample sizes. Application to a randomized clinical trial on nasopharyngeal cancer shows the practical utility of the model. The incorporation of the dependence structure reveals that patients in the chemotherapy group have increased chances of disease-free survival as compared to the radiotherapy group. Our results show that the chemotherapy group actually has increased risk of death without relapse and a reduced risk of death after relapse.

Finally, the extension to the more general additive-multiplicative frailty risk model for semi-competing risks data is discussed, with a similar splines approximation method for the baseline hazards. Simulations indicate estimation has little bias for the multiplicative component, while the estimates of the additive components had biases of at most 0.1. We re-examine the nasopharyngeal cancer dataset using this additive-multiplicative model under the reduced compartment model, with the treatment variable as a multiplicative effect and adjusting for nodal status and TNM staging as additive effects. Results show the significance of all three variables. Patients in the chemoradiotherapy group have a lower risk of both relapse and death as compared to patients in the radiotherapy group, with the difference in the two treatment groups being even larger in the death arm.

List of Tables

List of Figures

Chapter 1

Introduction

Survival analysis is the analysis of data where the response is the time until an event of interest occurs. In the case of univariate failure time data, research on modelling its underlying distribution and its dependence on explanatory variables are now well-established. However, additional problems arise when we deal with multivariate failure times and types. Multivariate failure time data arise when two or more distinct failures are recorded on an individual. These failures could be *recurrent failures* or *distinct failures of different types* (Kalbfleisch and Prentice, 2002).

Recurrent failures are observed in diverse settings, for instance, repeated episodes of infection, a sequence of asthmatic attacks, or epileptic seizures. Distinct failures of multiple types occur when the failures are of an entirely different nature, such as local or distant recurrences in cancer studies. In the context of competing risks, the failures are usually of different types, and the subject may fail from one of these distinct causes. (Tai *et al.*, 2008). For example, in a randomized clinical trial of patients with Stage III and IV nasopharyngeal cancer, the competing failures of interest were distant metastasis, local relapse, and neck relapses (Wee *et al.*, 2005). Similarly, in a clinical trial investigating

whether antiretroviral treatment delays the development of individual AIDS events, the different events of interest included oesophageal candidiasis, Kaposi sarcoma, *pneumocystis carinii* pneumonia, disseminated *Mycobacterium avium intercellulare*, cytomegalovirus, cryptosporidiosis, and cerebral toxoplasmosis (Delta Coordinating Committee, 1996).

This dissertation deals with multivariate failure time data of the latter type. It is biologically plausible that the failure times of these distinct failure types may be strongly correlated when observed in the same individual. Such failure time data can be considered clustered. In univariate failure time analysis, clustering may also arise when subjects are grouped based on common dependencies within groups. For example, in clustered randomized clinical trials of a general practice (primary care provider), all the patients in a general practice will be allocated to the same intervention, with the general practice forming a cluster. Similarly, in cohort studies on family members in genetic epidemiology, the family unit forms the cluster. In either case, members of a cluster will be more like one another than they are like members of other clusters. We need to take this into account in the analysis and design of the study. Ignoring clustering may result in misleading conclusions.

A particular case of multivariate failure time data is that of *bivariate survival data*. Here, there are two non-negative survival times, T_1 and T_2, that are correlated and have a particular joint survival function that expresses the dependence between the two times. For example, in the Diabetic Retinopathy Study (Huster *et al.*, 1989) the outcome of interest was the time to blindness in each eye of 197 patients with diabetic retinopathy. For each patient, one eye was randomly selected for treatment and the other eye was observed without treatment.

A special case of bivariate failure time data is the so-called *semi-competing risks data* (Fine *et al.*, 2001), where each subject may experience either a terminal event or a non-terminal event. The terminal event censors the non-terminal event, but not vice versa. In a randomised clinical trial conducted in Singapore comparing two treatments on nasopharyngeal cancer (Wee *et al.*, 2005), patients may experience the terminal event - death, or the non-terminal event - relapse of cancer at local or distant sites. It is plausible that the times to death and relapse are highly correlated. The dependence between the times to terminal event and non-terminal event, as well as the asymmetric structure of semi-competing risks data, have created challenges for statistical analysis of such data, especially for covariance analysis.

1.1 Frailty Models

One way of accounting for the dependence in multivariate failure time data is the use of frailties. Frailty models attempt to characterize the association between failure times through the use of a common unobserved random variable, known as the *frailty*. The frailty model has been extensively used for univariate failure time data, especially for clustered data where subjects experience a common dependence within a particular group. Under the structure of a frailty model for bivariate survival data, conditional on the frailty, T_1 and T_2 are considered independent.

Covariate analysis is often implemented through the use of the Cox proportional hazards model. Conditional on the frailty terms, the marginal hazard functions for T_1 and T_2 follow independent proportional hazards models. Let t_{ij}, $(i = 1, \ldots, n$ and $j = 1, 2)$, represent the failure time of the i-th individual for for

the j-th type. Thus, the conditional hazard function is given as

$$\lambda(t_{ij}|\gamma_i, Z_{ij}) = \gamma_i \lambda_0(t_{ij}) e^{\beta^T \mathbf{Z}_{ij}},$$

where $\mathbf{Z}_{ij}$ is the vector of covariates, β is the vector of regression coefficients, $\lambda_0(\cdot)$ is the unspecified baseline hazard function and $\{\gamma_1, \ldots, \gamma_n\}$ are the independent and identically distributed frailties.

Just as there are different kinds of copulas that one can use, there are also various distributions that the frailty variable can follow. A convenient and popular choice is the gamma distribution (Clayton and Cuzick, 1985; Nielsen *et al.*, 1992). Hougaard (1984, 1986a,b) considers the inverse Gaussian and positive stable distributions, and a three-parameter family of distributions, while Yau (2001) suggests that the frailties follow a lognormal distribution. Hougaard (1986b) makes a strong case for the positive stable distribution. Firstly, he points out a shortcoming of the gamma frailty distribution in that the dependence parameter and regression parameters are confounded and the joint distribution can be identified from the marginal distribution. This problem is present for any distribution with a finite mean. Secondly, the positive stable distribution has an added advantage that it preserves the proportionality of the hazards to the marginal distribution.

EM algorithms and maximum likelihood estimation (MLE) are often suggested as the method to estimate the frailty parameter, as well as regression coefficients. Since the frailty term is a latent variable, it makes sense to estimate these terms in the E-step, then use these estimators in the maximization of the likelihood in the M-step. Clayton and Cuzick (1985) use an EM-type algorithm with pseudo-observations of marginal distribution rank score orders to estimate the

regression and association parameters. Nielsen *et al.* (1992) also use the EM algorithm to estimate the regression and association parameters, as well as the unspecified baseline hazard for the proportional hazards model with a gamma frailty. Lam and Kuk (1997) propose the use of the marginal likelihood to estimate the parameters and suggest that this approach works for any frailty distribution with explicit Laplace transform. Gorfine *et al.* (2006) develop a new inference technique that can handle any parametric frailty distribution with finite moments. The method proposed is a pseudo-likelihood method that uses a plug-in estimator for the cumulative hazard function and avoids complicating the iterative optimization process.

1.2 Additive Risk Models and Clustered Data

While the Cox proportional hazards model has been widely discussed and extended, there is another formulation that describes a different aspect of the association between covariates and the failure time — the additive risk model. The additive risk model is adopted when the absolute effects, instead of relative effects, of predictors on the hazard function are of interest. In this way, we can analyse excess risk, instead of relative risk.

The intuitive idea for the additive risk approach is that the background disease incidence rate (or hazard rate) is due to the presence of general factors that are common to all subjects. The exposure to a particular treatment or agent under investigation causes the difference in an individual's overall hazard rate and is unrelated to the general factors. The differences in exposure are represented as excess risk (Breslow and Day, 1980). In some cases, the analysis of the estimated parameters results in the preference of excess risk measure over the relative risk measure.

Aalen (1980) first introduced the nonparametric version of the additive risk model. The estimators obtained by Aalen (1980) were a generalisation of the Nelson-Aalen (or natural) estimator and were based on least-squares type methods. Huffer and McKeague (1991) then extended the estimation to include weighted least-squares estimators for Aalen's additive risk model.

As mentioned earlier, the additive risk can be considered as an additive analogue to the Cox proportional hazards model. Here, the hazard function for the i-th individual is given as

$$\lambda(t|Z_i) = \lambda_0(t) + \beta^T \mathbf{Z}_i(t)$$

where $\mathbf{Z}_i$ is the vector of covariates that are allowed to vary with time, β is the vector of regression coefficients, and $\lambda_0(\cdot)$ is the unspecified baseline hazard function. Lin and Ying (1994) consider this model with time-dependent covariates and develop a semiparametric estimating function for the regression coefficient vector β. They first estimate the cumulative baseline hazard with a natural estimator and use this estimator in the estimating function. The resulting function mimics the martingale feature of the partial likelihood score function under a proportional hazards model. Explicit forms for the estimates were obtained. Under this method, the estimators converge weakly to a multivariate normal distribution with mean 0 and a covariance matrix that can be consistently estimated.

Note that a limitation of the additive risk model is that it is complicated by the constraint that the hazard function must be nonnegative (Huffer and McKeague, 1991; Lin and Ying, 1994). Thus, Lin and Ying (1994) suggest a substitution of $e^{\beta^T \mathbf{Z}_i(t)}$ for $\beta^T Z_i(t)$. However, there is now no explicit solution to the estimating equation and the Newton-Raphson algorithm is required. Analysis is also more

complicated numerically and theoretically.

The estimators of Lin and Ying (1994) are used by Pipper and Martinussen (2004) in applying the additive risk model to the clustered failure time setting with clusters. A marginal additive hazards model like that of Lin and Ying (1994) is suggested and the estimating function follows accordingly. Bearing in mind that we can no longer assume independence between failure times in clusters, working independence estimators that parallel those suggested by Lin and Ying (1994) are obtained. It is also shown that the working independence estimators of the regression coefficients are consistent and converge in distribution to a normal vector with zero mean and the estimator of the baseline cumulative hazard converges weakly to a Gaussian process.

Since it is of interest to estimate measures of dependence between failure times in a cluster, Pipper and Martinussen (2004) assume an additive marginal hazard, a parametric frailty and independence between the frailties and covariates in the respective clusters. Frailties, indicated as γ_k for $k = 1, \ldots, K$, are assumed to be independent and identically distributed positive random variables with Laplace transform $\phi_\theta(u) = E_\theta\{e^{-u\gamma_1}\}$, where θ is parameter associated with the frailty distribution. Through the use of the innovation theorem (Andersen et al., 1993), observed intensities are obtained up to time t and again, estimating equations are found that follow those of Lin and Ying (1994). Proper estimating equations are obtained by first inserting natural estimators for the baseline hazard and cumulative baseline hazard. The incorporation of the dependence through the frailty variable results in a more efficient estimation of the regression parameters as compared to the working independence estimators. While the choice of frailty distribution is not specified, conditions are placed on the frailty distribution to ensure that the Laplace transform $\phi_\theta(u)$ behaves nicely at the boundary. Thus,

this excludes distributions such as the positive stable distribution.

We now restrict our focus to semi-competing risks data and look at various methods of modelling such data. It is noted that the methods discussed are applicable to the more general cases of multivariate and clustered survival data.

1.3 Nonparametric Estimation in Semi-Competing Risks

Semi-competing risks occur when there is a terminal failure time, T_2 , and a non-terminal failure time, T_1 (Fine *et al.*, 2001). The random variable T_2 may censor T_1, but not vice-versa. This results in dependent censoring and there is possible correlation between T_1 and T_2. Such data are often encountered in medical studies. In cancer trials, when the goal is to estimate disease-free survival, the relapse of the disease is the non-terminal event of interest. However, terminal death from other causes censors the relapse.

Because of the asymmetric data structure, modelling can only be defined on the upper wedge, $\mathcal{U} = \{(t_1, t_2) : 0 < t_1 \leq t_2 < \infty\}$. Fine *et al.* (2001) introduced the term "semi-competing risk" and developed a method to model the dependence of the two event times. They also developed an estimator for the marginal distribution of the non-terminal event, as it is often of scientific interest to model its marginal distribution.

To model the dependence structure, they posited the use of the Clayton (1978) copula. A copula is a parametric method of transforming marginal distributions and expressing the transformed variables as a multivariate joint distribution.

The Clayton copula is specified as

$$S(t_1, t_2) = \left\{ S_1(t_1)^{1-\theta} + S_2(t_2)^{1-\theta} - 1 \right\}^{1/(1-\theta)},$$

where $S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$ is the joint survival function and $S_1(t_1) = P(T_1 > t_1)$ and $S_2(t_2) = P(T_2 > t_2)$ are the marginal survival functions of T_1 and T_2 respectively. Clayton's copula is an Archimedean copula with $\phi_\theta(t) = (t^{1-\theta} - 1)/(\theta - 1)$. It is useful to note that because of the structure of semi-competing risks data, the distribution of the terminal event, T_2, can be easily estimated using existing methods for univariate failure time data. Under the copula model, θ is known as the association parameter. Hence, it is of interest to estimate θ and S_1.

An estimator for the association parameter is obtained from a concordance estimating function and is determined as the ratio of concordant to disconcordant pairs. This is based on the idea that the cross-ratio function is equal to the association parameter for the Clayton copula (Oakes, 1989). The cross-ratio function is defined as the ratio of the hazard function of the conditional distribution of T_2, given $T_1 = t_1$, to that of T_2, given $T_1 > t_1$ and can be written in notation as

$$\frac{\lambda(t_2|T_1 = t_1)}{\lambda(t_2|T_1 > t_1)}. \tag{1.1}$$

With regard to the estimator of S_1, it is obtained through a rearrangement of the copula distribution given earlier. Some algebra gives

$$S_1(t) = \{S_2(t)^{1-\theta} - S_m(t)^{1-\theta} + 1\}^{1/(1-\theta)},$$

where $S_m(t) = P(T_1 > t, T_2 > t) = S(t, t)$ is the copula model defined at t. Using this definition, Fine *et al.* (2001) suggest that the estimator of S_1 can be

found by

$$\hat{S}_1(t) = \{\hat{S}_2(t)^{1-\hat{\theta}} - \hat{S}_m(t)^{1-\hat{\theta}} + 1\}^{1/(1-\hat{\theta})},$$

where $\hat{S}_2$ and $\hat{S}_m$ are the Kaplan-Meier nonparametric estimators for S_2 and S_m and $\hat{\theta}$ is the estimate of θ found earlier.

Jiang *et al.* (2005) provides another estimator for the survival function of T_1. The estimator is based on self-consistent estimating equations and is a step function that jumps at observed times. In contrast, Fine's estimator jumps not only at the observed times, but also at times outside the observed range. Thus, in this aspect, Jiang's estimator is an improvement and has better properties than the one proposed by Fine *et al.* (2001).

Wang (2003) extends the above model to a more general class of Archimedean copulas to model the dependency. These copulas can be written as

$$C_\theta(u, v) = \phi_\theta^{-1}\{\phi_\theta(u) + \phi_\theta(v)\}, \qquad 1 \geq v, u \geq 0,$$

where ϕ_θ is a non-increasing convex function defined on (0,1] with $\phi_\theta(1) = 0$. Wang (2003) considers two general dependence structures defined on the upper wedge $\mathcal{U}$ — one based on the cross-ratio function defined earlier and the other based on the Archimedean copula to model the joint distribution. From these two dependence structures, Wang (2003) suggests several estimating functions for the association parameter, θ. The variance of the estimator is complicated and a resampling method, such as the jackknife approach, is used to obtain an estimate of the variance.

Lakhal *et al.* (2008) provides a unified framework that generalizes the estimation of the association parameter for the family of Archimedean copulas. They

also use the cross-ratio function, defined in Equation (1.1), to construct an estimating function and show that the estimating functions provided by Fine *et al.* (2001) and Wang (2003) are special cases of their function. In addition, they present a method of estimating the survival function of T_1. The copulagraphic estimator they used was first introduced by Zheng and Klein (1995). In the latter, they assumed that the joint distribution of the failure and censoring times follow a known copula and derived estimating functions for the marginal survival function. Rivest and Wells (2001) found a closed-form expression for the survival function when the Archimedean copula is employed and this is the estimator that Lakhal *et al.* (2008) used. Under the method proposed, limiting distributions for the estimated copula parameter and survival function of T_1 are found and this is an improvement over the previous estimator.

The link between bivariate distributions generated by frailty models and Archimedean copulas is established by Oakes (1989). He presents a criterion based on the cross-ratio function and demonstrates that any bivariate frailty model leads to an Archimedean survival function, though the converse does not hold.

1.4 Regression Modelling in Semi-competing Risks

The methods mentioned so far deal with nonparametric estimation of the survival function of T_1. In addition, the copula model does not include any covariates. Thus, Peng and Fine (2007) focused on regression modelling, employing time-varying effects for the marginal survival function of T_1, given by

$$S_1(t|Z) = g\{\theta_0(t)^T \mathbf{Z}\},$$

where $g(\cdot)$ is a known monotone function and $\mathbf{Z}$ and $\theta_0(t)$ are $(p+1)\times 1$ vectors of covariates and time-dependent coefficients respectively. Since in the semi-competing risks setting, T_2 censors T_1, a model for the dependence structure is required in order to estimate $\theta_0(t)$. Hence, a time-independent copula function $C(u, v, w)$, is used and the joint survival function is given as

$$S(t_1, t_2|\mathbf{Z}) = C\{S_1(t_1|\mathbf{Z}), S_2(t_2|\mathbf{Z}), \alpha_0(t_1, t_2)\}, \qquad \text{for } 0 \leq t_1 \leq t_2.$$

Similar to earlier definitions, $S_1(t_1|\mathbf{Z}) = P(T_1 > t_1|\mathbf{Z})$ and $S_2(t_2|\mathbf{Z}) = P(T_2 > t_2|\mathbf{Z})$ are the marginal survival functions of T_1 and T_2 for a given covariate vector $\mathbf{Z}$, while $\alpha_0(t_1, t_2)$ is the time-dependent association parameter. Peng and Fine (2007) also specified that the marginal survival function of T_2 to be of the same form as $S_1(t_1|\mathbf{Z})$, though the link function and coefficient vector need not be the same. $S_2(\cdot|\mathbf{Z})$ is specified as

$$S_2(t|\mathbf{Z}) = h\{\eta_0(t)^T\mathbf{Z}\},$$

and the estimator for the coefficient vector, denoted as $\widehat{\eta}_0$, can be obtained using existing methods.

The simultaneous estimation for (α_0, θ_0), where $\alpha_0(t) = \alpha_0(t, t)$ is done via nonlinear estimating functions, which are obtained from a nonlinear binary regression model of the covariates Z on $I(\min(T_1, T_2) > t)$, given that $T_2 > t$, where $I(\cdot)$ is the indicator function. These functions jointly estimate α_0 and θ_0, separately at each t, adopting the "working independence" assumption across time (Liang and Zeger, 1986). Thus, the estimators obtained, $\widehat{\alpha}_0(t)$ and $\widehat{\theta}_0(t)$, are step functions which jump only at observed failure and censoring times. It is also suggested that a sensitivity analysis could be carried out based on

the proposed estimation procedure. That is, at each t, we can vary $\alpha_0(t)$ and estimate $\theta_0(t)$ at each value of $\alpha_0(t)$. This results in bounds on the covariate effects at t and gives us a rough idea of how sensitive the covariate effects are to changes in the dependence structure.

Under this structure of Peng and Fine (2007), variance estimators can be obtained through the delta method. In addition, nonparametric test statistics are constructed to test the null hypothesis of r linear combinations of $\alpha_0(t)$ and $\theta_0(t)$. One of these statistics is motivated by the Wald test, while another is a supremum-norm test. A graphical method of model checking is suggested in order to test the goodness-of-fit. This involves using the idea of a P-P plot to graph the fitted joint survival function against its nonparametric estimate, for a given covariate value. However, formal goodness-of-fit tests are not introduced here.

Hsieh and Wang (2008) propose a method of regression analysis for semi-competing risks data involving discrete covariates only. Again, their methodology assumes the family of Archimedean copulas for the dependence structure. However, separate copula models with different association parameters are assumed for different covariate groups. They suggest a two-stage inference procedure, where their main focus is on the estimation of the regression covariates. In the first stage, a modification of Wang's (2003) approach is used to estimate the association parameters, while the marginal distributions are estimated using the approach suggested by Fine *et al.* (2001). These estimators are then plugged into the second-stage estimating equation for the regression parameters.

While the copula model has been used widely in the area of semi-competing risks data, there are various disadvantages to the use of this type of model. Firstly,

the copula function requires the assumption of a marginal distribution for T_1, which presumes the existence of T_1 as a latent time. This is a rather hypothetical and unnatural concept and can be considered controversial. In addition, there is little literature on the modelling of regression covariates, which is often of interest in medical studies. The estimation methods proposed are also rather complicated. In contrast, our proposed method uses the additive risk frailty model as discussed below in Equation (1.2). It simplifies the estimation to nonlinear least-squares and the incorporation of time-dependent covariates is relatively simple.

So far in the literature, the modelling of semi-competing risks data employs the use of a parametric copula model for the dependence structure. In this thesis, we propose an alternative way of examining the possible correlation between T_1 and T_2 via the frailty models. The frailty approach for semi-competing risks data has been analysed by Xu *et al.* (2010) and Lim (2010). Both employ the use of proportional hazards conditional on the frailty. However, Xu *et al.* (2010) used a nonparametric method to describe the baseline hazards, while Lim (2010) assumes a parametric Weibull form. In both works, the frailty was assumed to follow the Gamma distribution with mean 1 and variance θ.

1.5 Layout of Thesis

Before embarking on our approach for modelling semi-competing risks data, we first examine a simpler situation. In Chapter 2, we look at the competing risks scenario, where an individual faces possible failure from multiple causes and the failure from one cause censors the failure from the others. Currently, there are two ways to model competing risks, either through the cause-specific hazard or the subdistribution hazard (Fine and Gray, 1999) and the proportional hazards

model is used. We apply the additive risk model to both the cause-specific hazard and the subdistribution hazard. Simulations conducted to examine the performance of the proposed model show that the estimation works well in both approaches. As an application, we also analyse a real dataset on prostate cancer (Green and Byar, 1980) and examine the treatment effect on the competing risks endpoints of cancer death, cardiovascular death and other causes of death. We apply both the cause-specific hazards and subdistribution hazards model to the dataset since both can provide complementary information about the data.

Next, as an alternative to the proportional hazards frailty models proposed by Xu *et al.* (2010) and Lim (2010), which were described at the end of the previous section, we propose an additive risk frailty approach for the modelling of semi-competing risks data. The random effect, or *frailty*, is used to model the dependence and the *additive risk* model is used to incorporate covariate effects. The nature of additive risk frailty modelling enables us to develop a class of estimation equations which can be numerically and conveniently solved by standard iterative least squares, or nonlinear least squares estimation. Theoretical properties of the estimator for both the dependence parameter and the regression coefficients can be rigorously established with their variance formula explicitly derived and consistently estimated by the sandwich formula and plug-in methods.

Under the additive risk frailty model, conditional on the frailty, the hazards model is an additive one. We have 3 hazard functions for each individual — one for time to relapse, one for time to death without relapse and one for time to death after relapse. The set of hazard functions facing the i-th individual is

given as

$$\begin{aligned}
\lambda_1(t|T_{i2} \geq t, \gamma_i, \mathbf{Z}_i) &= \gamma_i(\lambda_{01}(t) + \beta_1^T \mathbf{Z}_i(t)), \\
\lambda_2(t|T_{i1} \geq t, \gamma_i, \mathbf{Z}_i) &= \gamma_i(\lambda_{02}(t) + \beta_2^T \mathbf{Z}_i(t)), \\
\lambda_3(t|T_{i1} < t, \gamma_i, \mathbf{Z}_i) &= \gamma_i(\lambda_{03}(t) + \beta_3^T \mathbf{Z}_i(t)).
\end{aligned} \tag{1.2}$$

These hazards can be represented in a compartment model, shown in Figure 1.1.

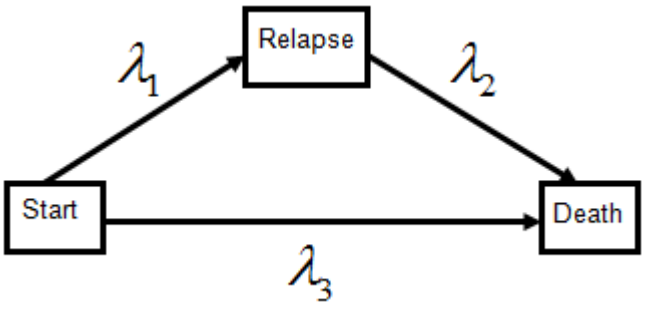

Figure 1.1: Compartment model for semi-competing risks. Hazard functions as shown in Equation (1.2)

In Chapter 3, we look at the context of semi-competing risks and the model discussed at the beginning of this section. In this thesis, for the fitting of the baseline hazards, we use the method of B-splines (de Boor, 1978). This reduces the estimation of parameters to a finite number. Simulations indicate that the method works well for moderate sample sizes. In addition, consistency and asymptotic normality can be established for the estimators of the parameters of interest — the frailty variance and the regression coefficients. We apply our method to a randomized clinical trial in nasopharyngeal cancer (Wee *et al.*, 2005) and analyse treatment effect, adjusting for nodal status and TNM staging. Results show that the model fitting treatment and accounting for nodal status and TNM staging gives similar results to the model fitting treatment only.

In Chapter 4, we extend the method explored in Chapter 3 and generalise it to the additive-multiplicative model. Under such a model, we allow some of the covariates to have an excess risk effect and others to have a relative risk effect. Simulations on semi-competing risks data show that estimation works well for the reduced model (relapse-death). The nasopharyngeal cancer dataset analysed in Chapter 3 is re-examined here with the treatment covariate having a multiplicative effect and nodal status and TNM staging as additive effects. With this new model, all covariates now have significant effects.

1.6 Contributions to the Medical Literature

The work arising from Chapter 2 of this thesis has been presented at a seminar talk in the first NUS Department of Statistics and Applied Probability PhD Students' Conference held in 2010. Parts of Chapter 3 have been presented at the 32nd Annual Conference of the International Society for Clinical Biostatisticians (Ottawa, Canada), as well as in a poster presentation at the Second Singapore Conference on Statistical Science (NUS, 19–20 September 2011).

Chapter 2

Additive Risk Models for Competing Risks Data

2.1 Introduction

Semi-competing risks data can be analysed in the competing risks setting, if we consider time to first event. Competing risks is a special case of multivariate failure time data and is the situation where an individual can potentially experience failure from one of several distinct causes. Under the classical competing risks framework, the causes of failure can be terminal and absorbing and there can be more than two causes of failure. In this chapter, we explore the modelling of such data, using the additive model for two different approaches. This model will be extended to model semi-competing risks in Chapter 3.

Competing risks are commonly observed in medical research, where subjects can experience failure from disease processes and/or non-disease-related causes. For instance, a multicentre randomized clinical trial conducted on bone marrow transplant patients records competing risks endpoints including recovery, relapse, chronic graft versus host disease and death (Couban *et al.*, 2002). An-

other example is data from a randomised clinical trial comparing treatment for patients with prostate cancer, where competing risks endpoints observed were cancer, cardiovascular and other causes of death (Green and Byar, 1980; Kay, 1986). The occurrence of one event either precludes the occurrence of another event under investigation or alters the probability of occurrence of other events (Gooley *et al.*, 1999). It is easy to see that there is dependence between the time to an event and the censoring mechanism.

Existing literature models competing risks data using two methods — the *cause-specific hazard* and *the subdistribution hazard*. We next look at these two approaches in Sections 2.1.1 and 2.1.2 respectively.

2.1.1 Cause-Specific Hazard

The *cause-specific hazard* is observed when we consider competing risks as latent (unobserved) failure times. We define the multivariate survival function as

$$S(t_1, t_2, \ldots, t_K|\mathbf{Z}) = P(T_1 > t_1, T_2 > t_2, \ldots, T_K > t_K|\mathbf{Z}),$$

where $T_1, \ldots, T_K$ are potential, unobserved event times for each of K event types and $\mathbf{Z}$ is the covariate vector. Under the competing risks scenario, only one event is observed, since the occurrence of this event will preclude the occurrence of other events, that is, $T = \min\{T_1, T_2, \ldots, T_K\}$. The event variable, ϵ, then takes on values $0, 1, 2, \ldots, K$, where 0 means the observation is censored and a non-zero values which of the K events has occurred.

From the multivariate survival function $S(t_1, t_2, \ldots, t_K|\mathbf{Z})$, we can obtain the

marginal survival function for event type k as

$$S_k(t|\mathbf{Z}) = S(t_1 = 0, t_2 = 0, \ldots, t_k = t, \ldots, t_K = 0|\mathbf{Z}).$$

The cause-specific hazard is defined as

$$h_k(t|\mathbf{Z}) = \left(- \frac{\partial \log(S(t_1, t_2, \ldots, t_K|\mathbf{Z}))}{\partial t_k} \right)_{(t_1=t_2=\ldots=t_K=t)}, \tag{2.1}$$

It can also be written as

$$\begin{aligned} h_k(t|\mathbf{Z}) &= \lim_{\delta t \to 0} \left\{ \frac{P(t < T \leq t + \delta t, \epsilon = k|T > t, \mathbf{Z})}{\delta t} \right\} \\ &= \frac{f_k(t|\mathbf{Z})}{S(t|\mathbf{Z})}, \end{aligned}$$

where

$$f_k(t) = \left(- \frac{\partial S(t_1, t_2, \ldots, t_K|\mathbf{Z})}{\partial t_k} \right)_{(t_1=t_2=\ldots=t_K=t)},$$

and $S(t|\mathbf{Z})$ is the overall survival function and is defined as

$$S(t|\mathbf{Z}) = \exp \left\{ - \int_0^t \sum_{k=1}^{K} h_k(t|\mathbf{Z}) dt \right\}. \tag{2.2}$$

One disadvantage of this approach is the non-identifiability of the joint distribution. In the special case of two competing risks, the two marginal distributions can result in more than one joint distribution. When only the first event is observed, the possible dependence between the competing events cannot be modelled. In such situations, only the marginal distribution function and the cause-specific hazard defined in Equation (2.1) can be modelled.

2.1.2 Subdistribution Hazard

The other approach based on the *subdistribution hazard* is a result of observing competing risks as a bivariate random variable. Such data can be presented as a pair (T, ϵ), where T is observed time and ϵ is as defined earlier. If $\epsilon = 0$, then T is the censored time. If $\epsilon = k$ $(k = 1, 2, \ldots, K)$, then T is the observed time that event of type k occurred. We then have the cumulative incidence function (CIF), or subdistribution, for the event of type k $(k = 1, 2, \ldots, K)$ as

$$F_k(t|\mathbf{Z}) = P(T \leq t, \epsilon = k|\mathbf{Z}).$$

As can be seen, the CIF is the joint probability that an event of type k occurs at or before time t.

The hazard of the subdistribution (Gray, 1988) is defined as

$$\lambda_k(t|\,\mathbf{Z}) = \lim_{\delta t \to 0} \left\{ \frac{P(t < T \leq t + \delta t, \epsilon = k \mid T > t \text{ or } (T \leq t \text{ and } \epsilon \neq k), \mathbf{Z})}{\delta t} \right\} \tag{2.3}$$

and this is the function that is usually modelled. The cumulative incidence function can then be modelled through $F_k(t|\,\mathbf{Z}) = 1 - \exp\left[\int_0^t \lambda_k(u)du\right]$. The subdistribution hazard λ_k can be considered as the hazard function for the improper random variable $T^* = I(\epsilon = k) \times T + \{1 - I(\epsilon = k)\} \times \infty$.

2.1.3 Existing Methodology for Modelling Competing Risks

To model competing risks data, the standard approach is to model the cause-specific hazards for different failure types. An important question in statistical analyses is whether one group of patients fare better than another group. For

example, in considering a clinical trial on prostate cancer where treatment is the main exposure of interest (Green and Byar 1980), it may sometimes also be important to consider the effects of other covariates such as age and disease history and account for these in the regression model when assessing the effect of treatment. Thus, covariate models for competing risks have been applied to cause-specific hazards to account for covariates like treatment. In particular, the proportional hazards model has been widely used in competing risks situations, just as in the usual survival model (Prentice *et. al*, 1978; Larson, 1984). The additive risk model was also considered as under the competing risks scenario, it seemed biologically more plausible and tends to give a more intuitive interpretation for relative survival than the multiplicative risk model (Shen and Cheng, 1999). This is because multiplicative models postulate that the hazards due to the event of interest are related to the hazards of the competing events. As such, estimation under the multiplicative model can result in illogical factors for mortality rates (Buckley, 1984).

Cause-specific hazards modelling reduces to univariate modelling since we only consider failure times of our cause of interest, ie. T_i where $\epsilon_i = k$, and all other failure times T_i with $\epsilon_i \neq k$ (failure times not of our cause of interest) are considered censored observations. Hence, there is only a single outcome being recorded, with a single censoring indicator. When the cause-specific hazard is modelled, the cumulative incidence function is often used to summarize the cause-specific failure time data through the relationship

$$F_k(t|\,\mathbf{Z}) = \int_0^t S(u|\,\mathbf{Z}) d\Lambda_k(u|\,\mathbf{Z}), \tag{2.4}$$

where $S(t|\,\mathbf{Z})$ is the all-cause survival function defined in Equation (2.2) and $\Lambda_k(t|\,\mathbf{Z})$ is the cumulative cause-specific hazard function for the k-th event. This

is an indirect way of modelling the subdistribution. It has been noted in other works that the effect of a covariate on the cause-specific hazards of a specific failure type can be very different from the effect of the same covariate on the corresponding cumulative incidence function (Gray 1988, Pepe 1991). Thus, the focus of this chapter will be on the direct modelling of the subdistribution through the associated subdistribution hazard.

In previous work in the literature, the cumulative incidence function has been modelled nonparametrically, as well as with discrete covariates. Gray (1988) considered K-sample tests to compare the cumulative incidence of a particular failure type among different groups. Fine and Gray (1999) introduced a proportional hazards model for the subdistribution. The proportional hazards was applied to the hazard of the subdistribution given in Equation (2.3). Under Fine and Gray's (1999) formulation, the risk set for censoring complete data (where the potential censoring time is always observed) at time t for failure type k was defined as

$$R(t) = \{i : (C_i \wedge T_i \geq t) \cup (\epsilon_i \neq k \cap T_i \leq t \cap C_i \geq t)\}$$

and the subdistribution hazard for failure type k was specified as

$$\lambda_k(t|\,\mathbf{Z}) = \lambda_{k0}(t)\exp(\beta_0^T\mathbf{Z}(t))$$

where λ_{k0} is the unspecified baseline hazard for failure type k, β_0 is the unknown p-vector regression coefficients for the possibly time-varying covariates $\mathbf{Z}$, so that the cumulative incidence function is now

$$F_k(t|\,\mathbf{Z}) = 1 - \exp\left[\int_0^t \lambda_{k0}(u)\exp(\beta_0^T\mathbf{Z}(u))du\right].$$

Three different scenarios were considered — complete data (without censoring), censoring complete data (failure times and potential censoring times all known) and incomplete data (when usual right censoring is present). The last scenario involved the use of inverse probability of censoring weighting (IPCW) techniques.

Sun *et al.* (2006) proposed a more general additive-multiplicative model for the subdistribution hazard of the form

$$\lambda_k(t|\,\mathbf{X},\mathbf{Z}) = \alpha(t)^T\mathbf{X} + \lambda_{k0}(t)\exp(\beta_0^T\mathbf{Z}),$$

where $\alpha(t)$ is an unknown q-vector of time-varying coefficients representing the additive effects of covariates $\mathbf{X}$ on λ_k, β_0 is a p-vector of unknown regression coefficients denoting the multiplicative effects of covariates $\mathbf{Z}$ on λ_k and λ_{k0} is as defined earlier. This model was first introduced by Martinussen and Scheike (2002) in the non-competing risk situation. Inference on the model was accomplished through the use of IPCW techniques to obtain score functions.

2.2 Proposed Additive Hazards Models

Let T and C be the failure and censoring times, $\epsilon \in \{1,\ldots,K\}$ be the cause of failure (where the K causes are assumed to be observable) and $\mathbf{Z}$ be a $p \times 1$ bounded vector of covariates. For the usual right-censored data, we observe $X = \min(T,C)$, $\delta = I(T \leq C)$ and $\mathbf{Z}$. Assume that $\{X_i, \delta_i, \delta_i\epsilon_i, Z_i\}$ are independent and identically distributed for $i = 1,\ldots,n$. For simplicity, we assume that C is independent of T, given $\mathbf{Z}$. Here, we take failure type 1 to be our event of interest. Let $N_i(t)$ be the counting process for the i-th individual, given by $N_i(t) = I(T_i \leq t, \epsilon_i = 1, \delta_i = 1)$.

Here, we introduce the additive risk model for the cause-specific hazard with fixed covariates. The model is given as

$$h_1(t|Z) = h_{01}(t) + \beta_1^T \mathbf{Z}, \tag{2.5}$$

where h_1 is the cause-specific hazard for event 1, $h_{01}(t)$ is the unknown baseline hazard, β_1 and $\mathbf{Z}$ are p-vector regression coefficients and covariates respectively.

In contrast to Fine and Gray's (1999) proportional hazards model for the subdistribution, we propose an additive hazards model for the subdistribution. The subdistribution hazard then takes the form

$$\lambda_1(t|Z) = \lambda_{01}(t) + \alpha(t)^T \mathbf{Z}, \tag{2.6}$$

where $\alpha(t)$ are time-varying coefficients to be estimated. This is analogous to Aalen's additive model. In Sun *et. al* (2006), $\alpha(t)$ is allowed to vary nonparametrically. In contrast, we fix the time-varying coefficient and as an example, assume it to be $\alpha(t) = \beta_1 e^{-t}$.

2.3 Model Fitting

Both the cause-specific hazards model and subdistribution hazards model can be fitted using the method provided by Lin and Ying (1994). The cause-specific hazards model is straightforward and is a direct application of the method, with the at-risk indicator as $Y_i(t) = I(C_i \wedge T_i \geq t)$ for the i-th individual. Under the method by Lin and Ying (1994), the estimating equation to estimate β_1 can be

defined as

$$U(\beta_1) = \sum_{i=1}^{n} \int_0^{\infty} \left\{ \mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t) \right\} \left\{ dN_i(t) - Y_i(t)\beta_1^T \mathbf{Z}_i(t)dt \right\}, \tag{2.7}$$

where the covariate $\mathbf{Z}$ is allowed to vary over time and

$$\bar{\mathbf{Z}}(t) = \sum_{j=1}^{n} Y_j(t)\mathbf{Z}_j(t) \Big/ \sum_{j=1}^{n} Y_j(t).$$

The resulting estimator can be explicitly expressed as

$$\hat{\beta}_1 = \left[\sum_{i=1}^{n} \int_0^{\infty} Y_i(t) \left\{ \mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t) \right\}^{\otimes 2} dt \right]^{-1} \left[\sum_{i=1}^{n} \int_0^{\infty} \left\{ \mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t) \right\} dN_i(t) \right], \tag{2.8}$$

where $a^{\otimes 2} = aa^T$, and a is a column vector.

In our additive subdistribution hazards model, we set the coefficients of the covariates to be time-varying, where $\alpha(t) = \beta_1 e^{-t}$, and the covariates are fixed, as shown in Equation (2.6). In order to use Lin and Ying's method, we set β_1 as the coefficient vector and $\mathbf{Z}e^{-t}$ as the time-varying covariate vector $\mathbf{Z}(t)$.

Under the subdistribution hazards model with censoring complete data (that is, censoring is only from administrative loss-to-follow up and the potential censoring time is always observed; Fine and Gray, 1999), the risk indicator at time t for the i-th individual is defined as

$$Y_i(t) = I(\{C_i \wedge T_i \geq t\} \cup \{\epsilon_i \neq 1 \cap T_i \leq t \cap C_i \geq t\}).$$

The estimator given in (2.8) can then be applied, where

$$\mathbf{Z}_i(t) = \mathbf{Z}_i e^{-t}, \qquad \bar{\mathbf{Z}}(t) = \sum_{j=1}^{n} Y_j(t)\mathbf{Z}_j(t) \Big/ \sum_{j=1}^{n} Y_j(t).$$

2.4 Theoretical Properties

In this section, we assume, without loss of generality, that T is bounded on $[0,1]$. Also, let β_{10} be the true value of β_1 and $\beta_{10} \in \mathcal{B}$, where $\mathcal{B}$ is a compact set of $\mathbb{R}^p$ with nonempty interior. We first list some of the conditions needed.

1. (Finite interval). $\int_0^1 \lambda_0(t)dt < \infty$.

2. (Asymptotic stability). There exists a matrix function z_1 and vector function z_2 defined on $[0,1]$ such that

$$\sup_{t\in[0,1]} \left\| n^{-1} \sum_{i=1}^{n} Y_i(t)\{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}^{\otimes 2} - z_1(t) \right\| \xrightarrow{P} 0,$$

$$\sup_{t\in[0,1]} \left\| n^{-1} \sum_{i=1}^{n} Y_i(t)\{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}^{\otimes 2}\mathbf{Z}_i(t) - z_2(t) \right\| \xrightarrow{P} 0.$$

3. (Lindeberg condition). There exists $\delta > 0$ such that

$$n^{-1/2} \sup_{1\le i\le n, t\in[0,1]} \left\{ |\mathbf{Z}_i(t)| Y_i(t) I(\beta_{10}^T \mathbf{Z}_i(t) > -\delta |\mathbf{Z}_i(t)|) \right\} \xrightarrow{P} 0.$$

4. (Asymptotic regularity conditions). $\mathbf{Z}$ has bounded support on $\mathbb{R}^p$ where p is the dimension of $\mathbf{Z}$. Also, z_1 and z_2 obtained in Condition 2 are bounded and the matrix

$$\Sigma = \int_0^1 \left[z_1(t)\lambda_0(t) + \beta_{10}^T z_2(t) \right] dt$$

is positive definite.

Theorem 2.4.1. (Consistency of $\hat{\beta}_1$) *Suppose that Conditions 1–4 listed above hold, then* $\hat{\beta}_1 \xrightarrow{P} \beta_{10}$.

Proof. To prove the consistency of $\hat{\beta}_1$, we first write down the log-likelihood

function under the additive risk model, $C(\beta_1, t)$, which is given as

$$C(\beta_1, t) = \sum_{i=1}^{n} \Big[\int_0^t \log(\lambda_0(u) + \beta_1^T \mathbf{Z}_i(u)) dN_i(u) - \int_0^t Y_i(u)(\lambda_0(u) + \beta_1^T \mathbf{Z}_i(u)) du \Big].$$

We now consider the process

$$\begin{aligned} X(\beta_1, t) &= n^{-1}(C(\beta_1, t) - C(\beta_{10}, t)) \\ &= n^{-1} \textstyle\sum_{i=1}^{n} \Big[\int_0^t \log \Big(\frac{\lambda_0(u) + \beta_1^T \mathbf{Z}_i(u)}{\lambda_0(u) + \beta_{10}^T \mathbf{Z}_i(u)} \Big) dN_i(u) - \int_0^t Y_i(u)(\beta_1 - \beta_{10})^T \mathbf{Z}_i(u) du \Big] \end{aligned}$$

and

$$\begin{aligned} A(\beta_1, t) \;=\; n^{-1} \sum_{i=1}^{n} \Big[\int_0^t \log \Big(\frac{\lambda_0(u) + \beta_1^T \mathbf{Z}_i(u)}{\lambda_0(u) + \beta_{10}^T \mathbf{Z}_i(u)} \Big) Y_i(u)(\lambda_0(u) + \beta_1^T \mathbf{Z}_i(u)) du - \\ \int_0^t Y_i(u)(\beta_1 - \beta_{10})^T \mathbf{Z}_i(u) du \Big] \end{aligned}$$

For each β_1, $X(\beta_1, t) - A(\beta_1, t)$ is a local square integrable martingale with

$$\langle X(\beta_1, t) - A(\beta_1, t), X(\beta_1, t) - A(\beta_1, t) \rangle = B(\beta_1, t),$$

where $\langle W, W \rangle$ is the predictable covariation process of W (Andersen *et al.*, 1993) and

$$B(\beta_1, t) \;=\; n^{-2} \sum_{i=1}^{n} \int_0^t \Big[\log \Big(\frac{\lambda_0(u) + \beta_1^T \mathbf{Z}_i(u)}{\lambda_0(u) + \beta_{10}^T \mathbf{Z}_i(u)} \Big) \Big]^2 Y_i(u)(\lambda_0(u) + \beta_{10}^T \mathbf{Z}_i(u)) du.$$

By Conditions 1, 2 and 4, we can see that $B(\beta_1) = B(\beta_1, 1)$ tends to 0 in probability. Therefore, by the inequality of Lenglart (I.2) in the appendix of Andersen and Gill (1982), we see that $X(\beta_1) = X(\beta_1, 1)$ converges to the same limit as $A(\beta_1) = A(\beta_1, 1)$ for each $\beta_1 \in \mathcal{B}$, where $\mathcal{B}$ was defined at the start of this section. By the boundedness conditions in Condition 4, we can obtain first and second derivatives which also have limits according to the limiting function

of $A(\beta_1)$. The first and second derivatives of $A(\beta_1)$ are then given by

$$n^{-1}\sum_{i=1}^{n}\Big[\int_0^1 \frac{Y_i(u)(\lambda_0(u)+\beta_{10}^T Z_i(u))Z_i(u)}{\lambda_0(u)+\beta_1^T Z_i(u)} - Y_i(u)Z_i(u)du\Big]$$

and

$$n^{-1}\sum_{i=1}^{n}\Big[\int_0^1 -\frac{Y_i(u)(\lambda_0(u)+\beta_{10}^T Z_i(u))[Z_i(u)]^{\otimes 2}}{(\lambda_0(u)+\beta_1^T Z_i(u))^2}du\Big]$$

respectively.

We can see that at $\beta_1 = \beta_{10}$, the first derivative is zero and the second derivative is the negative of a positive definite matrix. Thus for each $\beta_1 \in B$, $X(\beta_1)$ converges in probability to a concave function of β_1 with a unique maximum at $\beta_1 = \beta_{10}$. Since $\hat{\beta}_1$ maximises the random concave function $X(\beta_1)$, it follows by convex analysis (Andersen and Gill, 1982) that $\hat{\beta}_1 \xrightarrow{P} \beta_{10}$. □

Theorem 2.4.2. (Asymptotic normality of $\hat{\beta}_1$) *The function $n^{1/2}(\hat{\beta}_1 - \beta_{10})$ is asymptotically normal with mean 0 and covariance matrix consistently estimated by $V_n^{-1}\hat{\Sigma}_n V_n^{-1}$, where*

$$V_n = n^{-1}\sum_{i=1}^{n}\int_0^1 Y_i(t)\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}^{\otimes 2}dt, \tag{2.9}$$

and $\hat{\Sigma}_n$ is the estimator of the variance-covariance matrix of $n^{-1/2}U(\beta_{10})$, where $U(\beta_{10})$ is the estimating function given in Equation (2.7) evaluated at the true value.

Proof. One of the benefits of an additive risk model for competing risks is that the model provided by Lin and Ying (1994) has a closed form, given in Equation (2.8). The estimating function $U(\beta_1)$ to obtain this estimator is defined in

Equation (2.7). At the true value β_{10}, the estimating function becomes

$$\begin{aligned} U(\beta_{10}) &= \sum_{i=1}^{n} \int_0^1 \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\} \Big\{ dN_i(t) - Y_i(t)[d\Lambda_0(t) + \beta_{10}^T \mathbf{Z}_i(t)dt] \\ &\qquad + Y_i(t)[d\Lambda_0(t) + \beta_{10}^T \mathbf{Z}_i(t)dt] - Y_i(t)\beta_{10}^T \mathbf{Z}_i(t)dt \Big\} \\ &= \sum_{i=1}^{n} \int_0^1 \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}\{dM_i(t) + Y_i(t)d\Lambda_0(t)\} \\ &= \sum_{i=1}^{n} \int_0^1 \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}\{dM_i(t)\}, \end{aligned}$$

where $M_i(\cdot)$ is a local square integrable martingale and is defined by $N_i(t) = M_i(t) + \int_0^t Y_i(u)\{d\Lambda_0(u) + \beta_{10}^T \mathbf{Z}_i(u)du\}$ for every i and t.
To prove the asymptotic normality of $\hat{\beta}_1$, we first note that it reduces to proving the normality of $n^{-1/2} \sum_{i=1}^{n} \int_0^1 \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\} dN_i(t)$. Since

$$dN_i(t) = dM_i(t) + Y_i(t)(\lambda_0(t) + \beta_{10}^T \mathbf{Z}_i(t))dt,$$

the required normality proof reduces to proving the normality of

$$n^{-1/2} \sum_{i=1}^{n} \int_0^1 \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\} dM_i(t).$$

Hence, it remains to prove $n^{-1/2} U(\beta_{10})$ converges weakly to a p-variate normal with mean 0.

Firstly, we can see that $U(\beta_{10})$ is a local square integrable martingale. Applying Rebolledo's Central Limit Theorem for local square integrable martingales (Andersen and Gill, 1982), we have

$$H_{il}(t) = n^{-1/2}\{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}_l$$

Then

$$\begin{aligned}
&\int_0^t \sum_{i=1}^n \{H_{ij}(s)H_{il}(s)\lambda_i(s)ds\} \\
= &\int_0^t \sum_{i=1}^n \{n^{-1}\{\mathbf{Z}_i(s) - \bar{\mathbf{Z}}(s)\}_j\{\mathbf{Z}_i(s) - \bar{\mathbf{Z}}(s)\}_l Y_i(s)\{\lambda_0(s) + \beta_{10}^T\mathbf{Z}_i(s)\}ds\} \\
= &\Big[\int_0^t n^{-1}\sum_{i=1}^n Y_i(s)\{\mathbf{Z}_i(s) - \bar{\mathbf{Z}}(s)\}^{\otimes 2}\{\lambda_0(s) + \beta_{10}^T\mathbf{Z}_i(s)\}ds\Big]_{jl} \\
= &\Big[\int_0^t n^{-1}\sum_{i=1}^n Y_i(s)\{\mathbf{Z}_i(s) - \bar{\mathbf{Z}}(s)\}^{\otimes 2}d\Lambda_{1i}(s)\Big]_{jl} \quad \text{for } j,l = 1,\ldots,p
\end{aligned}$$

where $\Lambda_{1i}(t) = \int_0^t \Big\{\lambda_0(s) + \beta_{10}^T\mathbf{Z}_i(s)\Big\}ds$. By finite interval, stability and regularity conditions, the above tends to Σ defined in Condition 4.

To verify (I.4) in the Appendix of Andersen and Gill (1982), we have, for $\xi > 0$,

$$\begin{aligned}
&\int_0^1 \sum_{i=1}^n H_{il}(t)^2\lambda_i(t)I\{|H_{il}(t)| > \xi\}dt \\
= &\int_0^1 \sum_{i=1}^n n^{-1}\Big\{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}_l^2 Y_i(t)\{\lambda_0(t) + \beta_{10}^T\mathbf{Z}_i(t)\} \\
&\qquad\qquad I\{|n^{-1/2}\{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}_l| > \xi\}\Big\}dt \\
\leq &\int_0^1 \sum_{i=1}^n 4n^{-1}|\mathbf{Z}_i(t)|_l^2 I\{n^{-1/2}|\mathbf{Z}_i(t)|_l > \xi\}Y_i(t)\{\lambda_0(t) + \beta_{10}^T\mathbf{Z}_i(t)\}dt \\
&+ \int_0^1 \sum_{i=1}^n 4n^{-1}|\bar{\mathbf{Z}}(t)|_l^2 I\{n^{-1/2}|\bar{\mathbf{Z}}(t)||_l > \xi\}Y_i(t)\{\lambda_0(t) + \beta_{10}^T\mathbf{Z}_i(t)\}dt.
\end{aligned}$$

Thus it is sufficient to verify that

$$\begin{aligned}
n^{-1}\int_0^1 \sum_{i=1}^n \Big\{&|\mathbf{Z}_i(t)|^2 I\{n^{-1/2}|\mathbf{Z}_i(t)| > \xi,\ \beta_{10}^T\mathbf{Z}_i(t) > -\delta|\mathbf{Z}_i(t)|\} \\
&Y_i(t)\{\lambda_0(t) + \beta_{10}^T\mathbf{Z}_i(t)\}\Big\}dt \xrightarrow{P} 0, \qquad (2.10)
\end{aligned}$$

$$n^{-1}\int_0^1 \sum_{i=1}^{n} \Big\{ |\mathbf{Z}_i(t)|^2 I\{n^{-1/2}|\mathbf{Z}_i(t)| > \xi,\ \beta_{10}^T \mathbf{Z}_i(t) \le -\delta|\mathbf{Z}_i(t)|\} \\ Y_i(t)\{\lambda_0(t) + \beta_{10}^T \mathbf{Z}_i(t)\}\Big\} dt \xrightarrow{P} 0, \tag{2.11}$$

$$n^{-1}\int_0^1 \sum_{i=1}^{n} |\bar{\mathbf{Z}}(t)|_l^2 I\{n^{-1/2}|\bar{\mathbf{Z}}(t)||_l > \xi\} Y_i(t)\{\lambda_0(t) + \beta_{10}^T \mathbf{Z}_i(t)\} dt \xrightarrow{P} 0. \tag{2.12}$$

Using Condition 3, we have (2.10) using

$$\mathbf{P}\Big[\exists i, t : n^{-1/2}|Z_i(t)| > \xi,\ \beta_{10}^T Z_i(t) > -\delta|Z_i(t)|,\ Y_i(t) = 1\Big] \longrightarrow 0.$$

For (2.11), note that the left hand side of (2.11) is bounded by

$$n^{-1}\int_0^1 \sum_{i=1}^{n} |\mathbf{Z}_i(t)|^2 \{\lambda_0(t) - \delta|\mathbf{Z}_i(t)|\} I\{|\mathbf{Z}_i(t)| > n^{1/2}\xi\} dt.$$

When $\beta_{10}^T \mathbf{Z}_i(t) \le -\delta|\mathbf{Z}_i(t)|$, we have $\lambda_0(t) - \delta|\mathbf{Z}_i(t)| \ge 0$. Hence by Condition 4, the quantity on the left hand side is bounded by some positive finite quantity. Equation (2.12) is easily verified by the boundedness and regularity conditions in 1 and 4.

Thus, $n^{-1/2}U(\beta_{10})$ converges weakly to a certain continuous Gaussian function. With the process evaluated at $t = 1$, the covariance matrix is Σ. A consistent estimator of Σ is

$$\hat{\Sigma}_n = n^{-1}\sum_{i=1}^{n} \int_0^1 \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}^{\otimes 2} dN_i(t), \tag{2.13}$$

where the proof is as follows.

Let

$$\begin{aligned} W(t) &= n^{-1}\sum_{i=1}^{n}\int_0^1 \{\mathbf{Z}_i(t)-\bar{\mathbf{Z}}(t)\}^{\otimes 2}\{dN_i(t)-d\Lambda_i(t)\} \\ &= n^{-1}\sum_{i=1}^{n}\int_0^1 \{\mathbf{Z}_i(t)-\bar{\mathbf{Z}}(t)\}^{\otimes 2} dM_i(t), \end{aligned}$$

where $d\Lambda_i(t) = d\Lambda_0(t)+\beta_{10}^T\int_0^t \mathbf{Z}_i(u)du$. It can be seen that $W(t)$ is a local square integrable martingale with mean 0. We now consider $W = W(\infty)$, where we have

$$\langle W, W\rangle = n^{-2}\sum_{i=1}^{n}\int_0^1 \left[\{\mathbf{Z}_i(t)-\bar{\mathbf{Z}}(t)\}^{\otimes 2}\right]^2 d\Lambda_i(t).$$

Using Lenglart's inequality (I.2) in the appendix of Andersen and Gill (1982), we have for all $\delta,\ \eta > 0$,

$$\mathbf{P}\{\sup_{t\in[0,1)}|W(t)| > \eta\} \le \frac{\delta}{\eta^2} + \mathbf{P}\{\langle W, W\rangle > \delta\}.$$

Since $\langle W, W\rangle$ is bounded and finite in probability, we can show that the right hand side of the above inequality disappears and hence,

$$\hat{\Sigma}_n \xrightarrow{P} \Sigma.$$

With this consistent estimator of the covariance matrix and the asymptotic normality of $n^{-1/2}U(\beta_{10})$, we thus obtain the asymptotic normality of $n^{-1/2}(\hat{\beta}_1 - \beta_{10})$ with covariance matrix consistently estimated by $V_n^{-1}\hat{\Sigma}_n V_n^{-1}$, where V_n is defined in Equation (2.9) and $\hat{\Sigma}_n$ is the estimator of the variance-covariance matrix of $n^{-1/2}U(\beta_{10})$ defined in Equation (2.13).

2.5 Simulation Studies Based on Additive Hazards Model

The methods of data generation for the cause-specific hazards and subdistribution hazards are introduced in Section 2.5.1. Simulation results for the additive models on both types of hazard functions are presented in Section 2.5.1.

2.5.1 Data Generation for Competing Risks

Data generation using cause-specific hazards

Simulation studies for competing risks data often utilise the latent failure time model, which has been criticized. Since the cause-specific hazards can determine the competing risk process, Beyersmann *et al.* (2009) introduce a simulation design that depends only on the cause-specific hazards and not on unobservable quantities. Using the simple case of a main event of interest (Event 1) and a single competing risk (Event 2), we briefly introduce the algorithm here:

1. Specify the cause-specific hazards, $h_1(t)$ and $h_2(t)$. Here, we specify them to be the additive hazards indicated in Equation (2.5).

2. Generate survival times T using the all-cause hazard, $h_1(t) + h_2(t)$.

3. For each time T, assign event indicator ϵ the value 1 with probability $h_1(t)/(h_1(t) + h_2(t))$. Since there are only 2 events in the simple case, it reduces to a binomial experiment.

4. Generate external censoring times C from the Uniform$[0, a]$ distribution for some $a > 0$.

Data generation using subdistribution hazards

Following Fine and Gray (1999), we propose a similar algorithm to generate data from an additive subdistribution hazard (SH) of a competing risk. Denote the SH as $\lambda_1(t|Z) = P\left[T \in dt, \epsilon = 1 | (T \geq t) \cup (\epsilon \neq 1 \cap T \leq t)\right]$. For the cumulative incidence function (CIF) of interest, we use the additive hazards model on the SH as denoted in Equation (2.6). The CIF can then be written as

$$\begin{aligned} F_1(t|Z) &= 1 - \exp\left[-\int_0^t \left\{\lambda_{01}(t) + \alpha^T(t)\mathbf{Z}\right\}\right] \\ &= 1 - \exp\left[-\left\{\Lambda_{01}(t) + \int_0^t \alpha(u)^T\mathbf{Z}du\right\}\right] \\ &= 1 - \exp\left[-\Lambda_{01}(t)\right]\exp\left[-\int_0^t \alpha(u)^T\mathbf{Z}du\right] \\ &= 1 - \left[1 - F_1(t|0)\right]\exp\left[-\int_0^t \alpha(u)^T\mathbf{Z}du\right] \end{aligned}$$

and

$$\begin{aligned} 1 - F_1(t|0) &= 1 - P(T \leq t, \epsilon = 1|0) \\ &= P(\epsilon = 1|0) + P(\epsilon = 2|0) - P(T \leq t, \epsilon = 1|0) \\ &= P(\epsilon = 2|0) + P(T \geq t, \epsilon = 1|0) \\ &= P(\epsilon = 2|0) + P(T \geq t|\epsilon = 1, 0) \cdot P(\epsilon = 1|0) \end{aligned}$$

Here, we use 3 conditions:

1. $P(\epsilon = 1|0) + P(\epsilon = 2|0) = 1$,
2. $P(\epsilon = 1|0) = p$ (independent of Z), and
3. $T|\epsilon = 1, \mathbf{Z} = 0$ follows the exponential(1) distribution.

We then obtain

$$\begin{aligned} 1 - F_1(t|0) &= (1-p) + p \cdot \exp(-t) \\ &= 1 - p\left[1 - \exp(-t)\right] \end{aligned}$$

Thus, we have

$$F_1(t|\mathbf{Z}) = 1 - \left[1 - p\left\{1 - e^{-t}\right\}\right] \exp\left[-\int_0^t \alpha(u)^T \mathbf{Z} du\right]$$

If we specify $\alpha(t) = \beta_1$ (ie. a constant coefficient vector), we can see that when $t \to \infty$ and $\mathbf{Z} = 0$, $F_1(t|0) \to P(\epsilon = 1|0) = p$.
However, when $Z \neq 0$, we get $F_1(t|Z) \to P(\epsilon = 1|\mathbf{Z}) = 1$ since $\exp(-\beta_1^T \mathbf{Z} t) \to 0$.

Thus, in this model, $\alpha(t)$ is constrained to be time-varying and here, we specify the form as $\alpha(t) = \beta_1 e^{-t}$. With this form of $\alpha(t)$, we have

$$\begin{aligned} \lambda_1(t|\mathbf{Z}) &= \lambda_{01}(t) + \beta_1^T e^{-t} \mathbf{Z}, \text{ and} \\ F_1(t|\mathbf{Z}) &= 1 - \left[1 - p\left\{1 - e^{-t}\right\}\right] \exp(-\beta_1^T \mathbf{Z}(1 - e^{-t})). \end{aligned}$$

Under this specification, the covariate effect is a monotone one that decreases towards zero over time. The time-varying effect can also be generalised to account for different functional types.

Let T and C be the failure and censoring times and ϵ be the cause of failure. We assume two causes of failure — the event of interest (denoted as event 1) and the competing risk (event 2). Let $\mathbf{Z}$ be a $p \times 1$ bounded time-independent

covariate vector. The algorithm for generating data from this model is based on Fine and Gray (1999) and is as follows:

1. Generate $\mathbf{Z}$, the covariate variable.

2. Generate the cause of failure, ϵ.

   (a) Calculate $F_1(\infty|\mathbf{Z}) = P(\epsilon = 1|\mathbf{Z}) =: p^*$.

   (b) Generate U_1 from Uniform[0,1].

   (c) If $U_1 < p^*$, then set $\epsilon = 1$ and generate failure times from event 1. Else, set $\epsilon = 2$.

3. Generate T, conditional on ϵ.

   (a) If $\epsilon = 1$, generate T from $G_1(t) = P(T \leq t|\epsilon = 1, \mathbf{Z})$ using numerical methods, where

$$\begin{aligned} G_1(t) &= \frac{P(T \leq t, \epsilon = 1|Z)}{P(\epsilon = 1|Z)} \\ &= \frac{1 - [1 - p\{1 - e^{-t}\}] \exp(-\beta_1^T \mathbf{Z}(1 - e^{-t}))}{1 - [1 - p] \exp(-\beta_1^T Z)} \end{aligned}$$

      i. Generate U_2 from Uniform[0,1].

      ii. Solve for t such that $G_1(t) = U_2$, i.e., $t = G^{-1}(U_2)$, using **R** function **optim**.

   (b) If $\epsilon = 2$, generate T from $P(T \leq t|\epsilon = 2, \mathbf{Z})$, which follows the exponential distribution with rate $(1 + \beta_2^T \mathbf{Z})$

4. Generate external censoring times C from the Uniform$[0, a]$ distribution for some $a > 0$.

2.5.2 Simulation Results for Competing Risks

Cause-specific hazards simulations

In this section, we first present the simulation results for the cause-specific hazards. We applied the method of Lin and Ying (1994) to data generated using the method described in Section 2.5.1. For the purposes of simulation, we generated a main event of interest (Event 1) and a competing event (Event 2). Failures from cause 2 were taken as censored observations. A single discrete covariate Z was generated, first from the Bernoulli(0.5) distribution, and the true parameter values for the regression coefficients of the respective cause-specific additive hazards models were assumed to be $\beta_1 = \beta_2 = 1$. Continuous covariates were also considered in a separate simulation with a single Normal$(0, 0.25^2)$ covariate generated and parameter values set at $\beta_1 = \beta_2 = 1$.

Simulations of sample size 200 were conducted 1000 times for three degrees of censoring (10%, 30%, 60%). Table 2.1 gives the following estimators under all the different simulation scenarios: (i) $AVE(\hat{\beta}_1)$, estimated with the average of $\hat{\beta}_1$ from 1000 samples and standard errors of the average estimator given in parentheses; (ii) $SD(\hat{\beta}_1)$, estimated with the empirical standard deviation of $\hat{\beta}_1$ from the 1000 samples; (iii) $AVE(\hat{SD})$, the average of the 1000 standard error estimators. It can be seen from Table 2.1 that the method works well for both continuous and discrete cases, with little bias. In the discrete case, the biases range between 0.002 and 0.028 with standard errors of the mean estimates (in parentheses) varying about 0.012, while the biases for the continuous case are higher and range between 0.005 and 0.037 with standard errors of the mean estimates varying about 0.015. The biases and standard errors tend to increase with the level of censoring. In addition, the $SD(\hat{\beta}_1)$ values are close to those of $AVE(\hat{SD})$, although the differences are slightly larger in the continuous case.

Estimated standard errors of β_1 increased with the level of censoring in both the discrete and continuous case.

Table 2.1: Estimating Equation Estimators for main event of interest (Event 1) based on a Cause-Specific Hazards Model with single $\mathbf{Z}$, assuming $\beta_1 = \beta_2 = 1$, varying censoring from 10% to 60%.

		Z from **Bernoulli(0.5)**		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 3.5]$	10	1.010(0.010)	0.342	0.326
$[0, 1]$	30	1.002(0.012)	0.362	0.366
$[0, 0.38]$	60	1.028(0.015)	0.490	0.481

		Z from **Normal**$(0, 0.25^2)$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 5.2]$	10	1.021(0.013)	0.430	0.420
$[0, 1.64]$	30	0.995(0.015)	0.490	0.476
$[0, 0.57]$	60	1.037(0.021)	0.679	0.639

We also check to see that the method works well when fitting the additive risk model for the competing risk. In our simulations for the cause-specific model, the competing risk was also generated from the additive risk model, as mentioned in Section 2.5.1. Table 2.2 shows the results for a single scenario with a single Bernoulli(0.5) covariate, where $(\beta_1, \beta_2) = (-0.5, 1)$ at 10%, 30% and 60% censoring. From Table 2.2, the biases are relatively small, with higher biases at higher levels of censoring. The observations for $SD(\hat{\beta}_1)$ and $AVE(\hat{SD})$ were similar to the ones made for Event 1, with differences being small and the estimations increasing with the level of censoring.

Table 2.2: Estimating Equation Estimators of β_2 for competing event (Event 2) based on a Cause-Specific Hazards Model with single Z from Bernoulli(0.5), assuming $\beta_1 = -0.5, \beta_2 = 1$, varying censoring from 10% to 60%.

		$\beta_1 = -0.5, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_2$)*	*SD($\hat{\beta}_2$)*	*AVE($\hat{SD}$)*
$[0, 4.7]$	10	1.012(0.009)	0.285	0.282
$[0, 1.49]$	30	0.879(0.009)	0.251	0.280
$[0, 0.51]$	60	0.965(0.013)	0.385	0.397

Since we are interested to evaluate the treatment effect in most clinical trials, we next look at the discrete Bernoulli(0.5) covariate, with differing values of β_1 and β_2. We fix the value of β_2 to be 1 and β_1 takes on values 0.5 (adverse treatment effect), 0 (no treatment effect) and -0.5 (beneficial treatment effect). Table 2.3 shows the results of these simulations at the same censoring proportions as before. All estimates have very little bias (absolute bias less than 0.01) and standard deviation estimates were similar for $SD(\hat{\beta}_1)$ and $AVE(\hat{SD})$. Higher censoring proportions result in higher standard errors and standard deviation estimates, as well as lower biases per unit of standard error. Censoring of 10% and 60% gave biases per unit of standard error around 0.033 and 0.015 respectively.

Table 2.3: Estimating Equation Estimators for main event of interest (Event 1) based on a Cause-Specific Hazards Model with single Z from Bernoulli(0.5), assuming 10% to 60% censoring and varying β_1 and β_2.

		$\beta_1 = 0.5, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 3.75]$	10	0.491(0.009)	0.286	0.279
$[0, 1.19]$	30	0.503(0.010)	0.314	0.315
$[0, 0.41]$	60	0.495(0.013)	0.415	0.418
		$\beta_1 = 0, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 4.15]$	10	0.008(0.007)	0.239	0.234
$[0, 1.3]$	30	0.009(0.009)	0.272	0.268
$[0, 0.46]$	60	0.006(0.011)	0.352	0.352
		$\beta_1 = -0.5, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 4.7]$	10	$-0.489(0.006)$	0.195	0.195
$[0, 1.49]$	30	$-0.499(0.007)$	0.216	0.221
$[0, 0.51]$	60	$-0.505(0.009)$	0.296	0.291

Subdistribution hazards simulations

Following the data generation method outlined in Section 2.5.1, we conduct simulations for the additive subdistribution hazards model with censoring-complete data and report the results below. We first assume a single covariate, Z, from the Bernoulli(0.5) distribution and that the true parameter values are $(p, \beta_1, \beta_2) = (0.3, 1, 1)$. We used three different degrees of censoring (ie. 10%, 30% and 60%), with 1000 samples, each of size 200, generated for each case. We also consider the case of a single continuous covariate from Normal$(0.5, 0.25^2)$ distribution, with parameter values set to be the same as with the discrete case. Table 2.4 shows the results for these simulations. The estimators here are similar to those reported previously. The biases for the discrete case (range of 0.001 – 0.025) are smaller than those for the continuous case (range of 0.01 – 0.031). It is interesting to note that for the continuous case, lower censoring proportions resulted in higher biases, although the biases are still small. Standard deviation estimates were similar for both methods, regardless of censoring proportions, and increased with the level of censoring.

We also examined simulations with varying values of p, as well as looked at covariate effect by differing the values of β_1. Tables 2.5, 2.6 and 2.7 show the results for adverse, no and beneficial treatment effect for 3 different values of p — 0.3, 0.6 and 0.9. Each covariate effect and level of p is examined at the same levels of censoring as before. In all the simulations, the estimating equations performed well, with little bias, with most of the biases ranging between 0.004 and 0.01. The largest bias observed was 0.023. The standard errors for the average values obtained were about 0.009 and varied little regardless of degree of censoring or the values of p or (β_1, β_2). The empirical and model variances were similar. There were larger variance estimators at higher degrees of censoring.

Table 2.4: Censoring Complete Estimating Equation Estimators for main event of interest (Event 1) based on a Subdistribution Hazards Model with single Z, assuming $\beta_1 = \beta_2 = 1$, varying censoring from 10% to 60%.

		Z from **Bernoulli(0.5)**		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 8]$	10	0.999(0.010)	0.323	0.319
$[0, 2.5]$	30	1.004(0.011)	0.348	0.336
$[0, 0.85]$	60	1.025(0.012)	0.365	0.386

		Z from **Normal**$(0.5, 0.25^2)$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 7]$	10	1.031(0.020)	0.649	0.637
$[0, 2.4]$	30	1.024(0.021)	0.669	0.661
$[0, 0.82]$	60	1.010(0.025)	0.760	0.771

However, given the degree of censoring and combination of (β_1, β_2), the variance estimators were similar for varying p. This seems to indicate a robustness in our estimation method in relation to p, since p is not a parameter that is estimated in our proposed model.

Table 2.5: Censoring Complete Estimating Equation Estimators for main event of interest (Event 1) based on a Subdistribution Hazards Model with single Z from Bernoulli(0.5) and $p = 0.3$, assuming 10% to 60% censoring and varying β_1 and β_2.

		$\beta_1 = 0.5, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 8]$	10	0.506(0.009)	0.288	0.285
$[0, 2.65]$	30	0.496(0.009)	0.304	0.297
$[0, 0.9]$	60	0.507(0.011)	0.354	0.343

		$\beta_1 = 0, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 8.2]$	10	0.007(0.008)	0.248	0.251
$[0, 2.58]$	30	−0.004(0.008)	0.266	0.264
$[0, 0.87]$	60	−0.010(0.010)	0.305	0.306

		$\beta_1 = -0.5, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 7.5]$	10	−0.508(0.007)	0.224	0.221
$[0, 2.33]$	30	−0.495(0.007)	0.233	0.233
$[0, 0.78]$	60	−0.511(0.009)	0.269	0.275

Table 2.6: Censoring Complete Estimating Equation Estimators for main event of interest (Event 1) based on a Subdistribution Hazards Model with single Z from Bernoulli(0.5) and $p = 0.6$, assuming 10% to 60% censoring and varying β_1 and β_2.

		$\beta_1 = 0.5, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 8.76]$	10	0.505(0.009)	0.277	0.283
$[0, 2.72]$	30	0.502(0.009)	0.296	0.297
$[0, 0.92]$	60	0.506(0.011)	0.350	0.342

		$\beta_1 = 0, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 9]$	10	−0.002(0.008)	0.250	0.249
$[0, 2.83]$	30	0.008(0.008)	0.267	0.263
$[0, 0.97]$	60	0.011(0.010)	0.302	0.301

		$\beta_1 = -0.5, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 9]$	10	−0.500(0.007)	0.220	0.221
$[0, 2.86]$	30	−0.501(0.007)	0.226	0.230
$[0, 0.98]$	60	−0.495(0.008)	0.270	0.262

Table 2.7: Censoring Complete Estimating Equation Estimators for main event of interest (Event 1) based on a Subdistribution Hazards Model with single Z from Bernoulli(0.5) and $p = 0.9$, assuming 10% to 60% censoring and varying β_1 and β_2.

		$\beta_1 = 0.5, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 8.8]$	10	0.514(0.009)	0.281	0.284
$[0, 2.79]$	30	0.477(0.009)	0.303	0.297
$[0, 0.94]$	60	0.496(0.011)	0.332	0.340

		$\beta_1 = 0, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 9.7]$	10	−0.018(0.008)	0.249	0.252
$[0, 3.1]$	30	−0.007(0.008)	0.264	0.261
$[0, 1.09]$	60	0.014(0.009)	0.296	0.294

		$\beta_1 = -0.5, \beta_2 = 1$		
[0,a]	*Censoring proportion*	*AVE($\hat{\beta}_1$)*	*SD($\hat{\beta}_1$)*	*AVE($\hat{SD}$)*
$[0, 10.8]$	10	−0.501(0.007)	0.223	0.220
$[0, 3.5]$	30	−0.504(0.007)	0.226	0.228
$[0, 1.27]$	60	−0.492(0.008)	0.246	0.250

2.6 Application to Prostate Cancer Dataset

The dataset application in this chapter is a randomised clinical trial comparing treatment for patients with prostate cancer in stages 3 and 4. It was first analysed by Green and Byar (1980). This dataset consisted of 506 patients randomly allocated to one of four different doses of diethylstilbestrol (DES) (placebo, 0.2mg, 1mg and 5mg daily). Treatment effect was examined for the risk of death from different causes. Green and Byar (1980) applied the proportional hazards model with an exponential baseline on the survival times and considered interaction effects between treatment and other covariates, without accounting for the competing risks nature of the data. Kay (1986) re-examined the data using a subset of 483 patients, those with complete information on all the relevant variables, and applied the proportional hazards model on the cause-specific hazards. This allowed a direct measure of the effect of treatment on the different causes of death — cancer, cardiovascular disease (CVD) and other causes. Cheng *et al.* (1998) also analysed this dataset using the same subset of patients but with different classifications of death, that is, death from prostate cancer, CVD and other causes. They predicted the cumulative incidence function (defined in Section 2.1.2) based on proportional hazards on the cause-specific hazards. Ng and McLachlan (2003) used the same subset of data with the same endpoints as Cheng *et al.* (1998) and proposed a semi-parametric mixture model to account for the competing risks nature of the data.

We aim to examine this dataset on prostate cancer using the same failure types as Kay (1986) — death due to cancer, death due to CVD and death due to other causes. There were 483 patients in the subset with complete information. For analysis purposes, the four different doses of DES were classified into two groups — low-dose DES (0 and 0.2mg) and high-dose (1.0mg and 5.0mg). The

Table 2.8: Coding of the covariates in the prostate cancer data (Green and Byar, 1980)

	Value		
Variable	0	1	2
RX	0 (placebo) and 0.2mg	1.0mg and 5.0mg	
AG	<75 years	75 to 79 years	$\geq$ 80 years
WT	$\geq$ 100	80 to 99	<80
PF	Normal	Limited	
HX	No	Yes	
HG	$\geq$ 12 g/100 ml	9–11.9 g/100 ml	<9 g/100 ml
SZ	$< 30cm^2$	$\geq 30cm^2$	
SG	$\leq$ 10	>10	

treatment indicator, RX, assigned value 0 for the low-dose group and value 1 for the high-dose group. Following the previous work of Kay (1986), there were seven other covariates included in the model: age (AG), weight index (WT), performance rating (PF), history of cardiovascular disease (HX), serum hemoglobin (HG), size of primary lesion (SZ) and Gleason stage/grade category (SG). The coding of these covariates are according to Green and Byar (1980) and is given in Table 2.8.

Of the 483 patients, there were 241 patients in the low-dose group and 242 patients in the high-dose group. There were 344 deaths in total, with 149 cancer deaths, 139 CVD deaths and 56 deaths from other causes. Other causes include deaths from respiratory diseases, other specified or unspecified noncancer causes and unknown causes. The remaining 139 patients were censored, giving a censoring proportion of 28.8%. While Kay (1986) and Cheng *et al.* (1998) use

goodness-of-fit tests to suggest that the Cox model on the cause-specific hazards is a reasonable fit to the data, we apply the additive risk model on the three cause-specific hazards to examine excess risk. The estimates of the regression coefficients based on the additive risk model outlined earlier for the three causes are given in Table 2.9. We fit the treatment indicator as well as the seven other covariates, keeping in line with previous analysis work. We also fit the additive risk model for the overall survival. The survival time was measured in months.

Table 2.9: Parameter estimates for overall survival and cause-specific hazards (data from Green and Byar, 1980).

	.	Cause-specific		
Coefficient	Overall Survival	Cancer	CVD	Others
RX	−0.0029	−0.0043 *	0.0030 *	−0.0017
	(0.0022)	(0.0015)	(0.0014)	(0.0009)
AG	0.0066 *	0.0000	0.0035 *	0.0031 *
	(0.0023)	(0.0014)	(0.0015)	(0.0012)
WT	0.0036	0.0014	0.0004	0.0018 *
	(0.0020)	(0.0014)	(0.0012)	(0.0008)
PF	0.0131 *	0.0047	0.0058	0.0025
	(0.0062)	(0.0041)	(0.0039)	(0.0025)
HX	0.0090 *	−0.0004	0.0094 *	−0.0000
	(0.0025)	(0.0015)	(0.0017)	(0.0010)
HG	0.0079 *	0.0063 *	0.0002	0.0014
	(0.0035)	(0.0025)	(0.0019)	(0.0014)
SZ	0.0203 *	0.0183 *	−0.0009	0.0029
	(0.0058)	(0.0048)	(0.0025)	(0.0021)
SG	0.0077 *	0.0101 *	−0.0008	−0.0015
	(0.0023)	(0.0016)	(0.0015)	(0.0009)

Standard errors given in parentheses and * indicates significance at 5% level.

As shown from Table 2.9, the conclusion with regards to the statistical significance of individual covariates are similar to that of Kay (1986). For overall survival, the treatment effect is not significant, with the high-dose group having higher survival probability. The remaining covariates have significant effect, with WT having marginal significance and worse survival for increased WT. As for the cause-specific analysis, it can be seen that the high-dose DES group does better with reduced risk of cancer death (estimated excess risk = -0.00426 per month), but fared worse with increased risk of CVD death (estimated excess risk = 0.00303 per month). Hemoglobin, tumour size and stage/grade significantly affected the risk of cancer death, with higher values associated with higher risk. Older patients had a higher risk of CVD death, as are those with a history of CVD. Older patients with a higher value for the weight index were at higher risk of deaths from other causes (noncancer, nonCVD or unknown), with a marginal beneficial treatment effect.

We now examine the same dataset using the subdistribution hazards model described in Section 2.2, Equation (2.6), with estimators defined in Section 2.3. In the model, we assumed that the regression coefficients are time-varying and take the form $\alpha(t) = \beta_1 e^{-t}$, where α and β_1 are column vectors of length 8, corresponding to the eight covariates fitted. Table 2.10 shows the estimators of β_1 when fitted for the subdistribution hazards for each of the three competing events — cancer death, CVD death and death due to other causes. Figure 2.1 shows the cumulative incidence functions comparing treatment for each of the three causes of death, when adjusted for the average value of the other covariates.

The results indicate a lack of a significant treatment effect for the subdistribution hazards model with time-varying coefficients of the form $\beta_1 e^{-t}$. However,

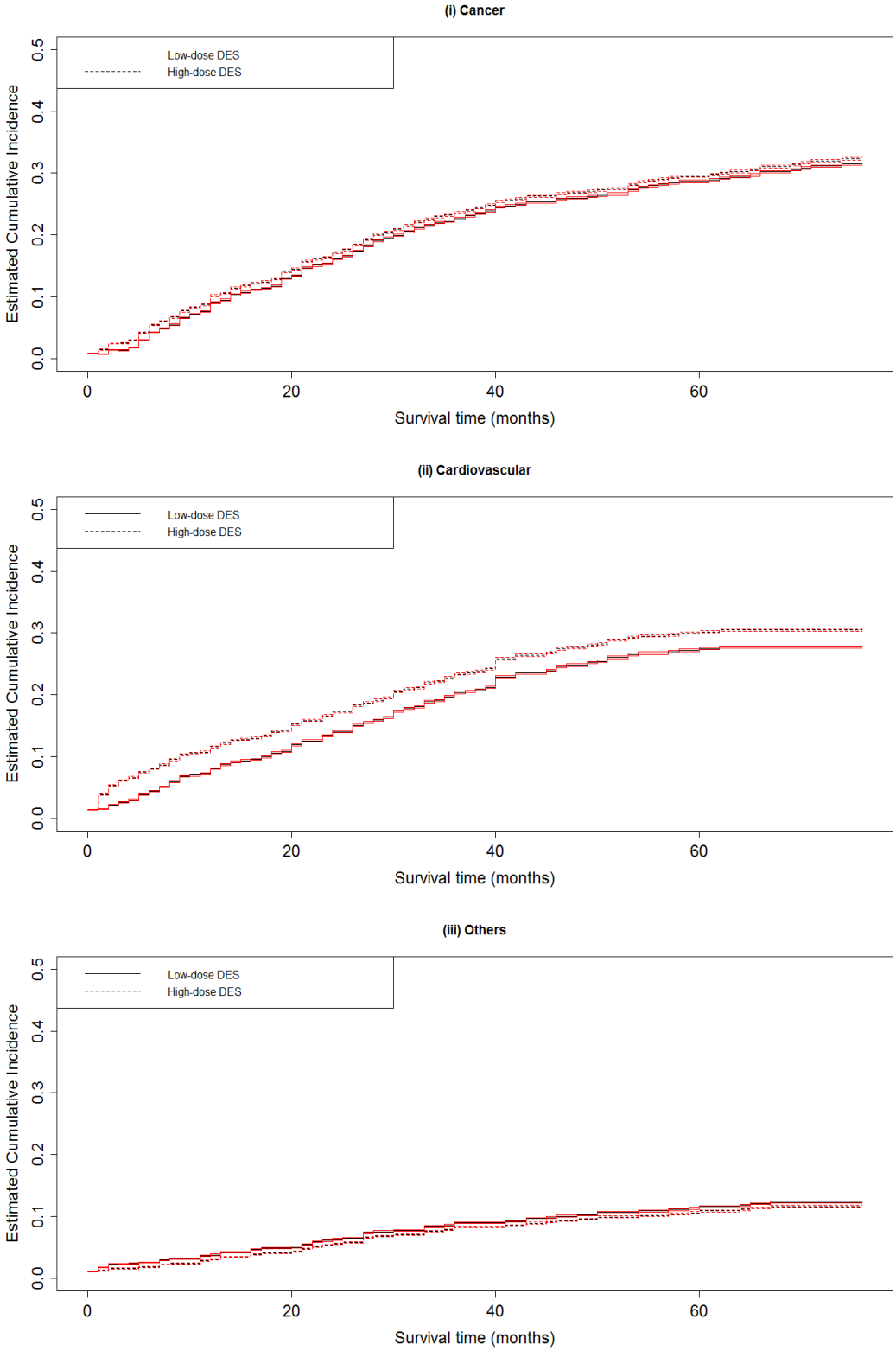

Figure 2.1: Cumulative incidence functions (CIF) comparing low- and high-dose DES patients with average values of other covariates (data from Green and Byar, 1980). CIFs are plotted for: (i) cancer, (ii) cardiovascular, (iii) other causes of death.

Table 2.10: Parameter estimates for subdistribution hazards (data from Green and Byar, 1980).

	Subdistribution hazard		
Coefficient	Cancer	CVD	Others
RX	0.0122	0.0386	−0.0079
	(0.0179)	(0.0255)	(0.0189)
AG	0.0184	0.0102	0.0107
	(0.0133)	(0.0244)	(0.0249)
WT	−0.0130	−0.0004	0.0123
	(0.0149)	(0.0159)	(0.0070)
PF	0.0208	0.0054	0.0797
	(0.0415)	(0.0517)	(0.0637)
HX	−0.0144	0.0778 *	−0.0065
	(0.0191)	(0.0282)	(0.0209)
HG	0.0235	0.0281	0.0613
	(0.0238)	(0.0349)	(0.0453)
SZ	0.0641	0.0014	−0.0006
	(0.0525)	(0.0416)	(0.0396)
SG	0.0229	0.0147	0.0145
	(0.0180)	(0.0247)	(0.0170)

Standard errors given in parentheses and * indicates significance at 5% level.

we can see that under the subdistribution hazards model, the covariate effects differ from those under the cause-specific hazards model, in particular, the treatment effect on cancer deaths. Since the risk varies over time, we will take time at month 1 as an example. Under the cause-specific hazards model, patients receiving high-dose DES had a lower risk of cancer death. However, under the subdistribution hazards model, patients receiving high-dose DES had a higher instantaneous risk of cancer death such that at month 1, high-dose patients had

a $0.0122e^{-1} = 0.00449$ higher risk than low-dose patients.

Using the model from Table 2.10, we can estimate the probability of different death outcomes over time for patients with different characteristics. A patient between 75–79 years of age with a moderate weight index and has a history of CVD has an estimated probability of 0.25 of cancer death within 4 years if he receives high-dose DES. A patient with the same characteristics would die from cardiovascular disease within 4 years with probability 0.3, and from other causes with probability 0.08. Thus, for patients who have a history of CVD, cancer death may not be the primary concern. On the other hand, a patient between 75–79 years of age with a moderate weight index and a primary lesion more than $30cm^2$ has a probability of 0.30 of cancer death within 4 years, whereas the estimated probability of cardiovascular death and other causes is 0.25 and 0.06 respectively.

2.7 Discussion

In this chapter, we applied the additive hazards model to competing risks via two approaches — the cause-specific hazard and the subdistribution hazard, with the focus on the estimation of the regression coefficients. The purpose of using additive hazards models is to act as a complement to the widely-used proportional hazards models which examine relative risk, even in competing risks data. Using counting process theory, we prove the consistency and asymptotic normality of the estimators. Simulations were conducted from new approaches for both the cause-specific and subdistribution hazards. Results from the simulations show that the data generation and estimation procedures work well. Application of the methods to the prostate cancer dataset yields similar conclusions as Kay (1986) under the cause-specific hazards model. We also analysed

the data from the subdistribution hazards approach, to see if it yielded a different result from the cause-specific hazards approach used earlier and in other literature. Fine (1999) examines this dataset using a general class of transformation models for the cumulative incidence function. The proportional hazards and odds model was selected for cancer and CVD outcomes respectively. Our results differ from Fine (1999) as we observed that high-dose DES increased the incidence of death, while Fine (1999) observed otherwise. However, we observe similar results for the modelling CVD cumulative incidence in that high-dose DES increases the incidence of CVD death. While the additive subdistribution hazards model does not seem a reasonable fit for the prostate cancer dataset, one reason for this is that we have specified the form of the time-varying function of the regression coefficient. Further work could look at other time-varying forms that could be used for the subdistribution model.

Chapter 3

A Frailty Model with Conditional Additive Hazards for Semi-Competing Risks Data

3.1 Introduction

In studies involving time-to-event outcome, a subject may experience multiple failures, such as repeated episodes of infection, or distant or local recurrences in cancer studies. In the context of competing risks, the failures are usually of different types. In this chapter, we consider a variation of the competing risks problem, known as semi-competing risks (Fine, Jiang and Chappell, 2001) where a non-terminal event is censored by a terminal event but not vice versa. Under the semi-competing risks framework, each subject is associated with two potential failure times — a non-terminal failure (e.g. relapse) and a terminal failure (e.g. death). It is biologically plausible that the failure times observed for each individual may be strongly correlated.

As an example, consider the randomized clinical trial for nasopharyngeal cancer

(NPC), conducted in Singapore between September 1997 and May 2003 (Wee *et al.*, 2005). This study compared standard radiotherapy treatment to chemoradiotherapy followed by adjuvant chemoradiotherapy on patients with American Joint Committee on Cancer/International Union Against Cancer (1997) Stage 3 and 4 nasopharyngeal cancer of the endemic variety. The end-points of interest were distant metastasis and disease-free survival. In both cases, relapse is the non-terminal event of interest, but could be censored by death. Wee *et al.* (2005) used competing risks methodology to analyse the data, but this does not utilise the additional information provided by the dependent censoring structure of semi-competing risks.

Existing literature uses the method of copulas to model the joint survival function between the two failure times, both without covariates (Fine, Jiang and Chappell, 2001; Jiang *et al.*, 2005; Wang, 2003; Lakhal *et al.*, 2008) and with covariates (Peng and Fine, 2007; Hsieh and Wang, 2008). These methods were discussed in Chapter 1. In this chapter, we introduce an alternative way to fit such data that has a straightforward way of incorporating covariates and measuring the dependent relationship between T_1 and T_2. Instead of modelling the joint survival function, we propose to model the hazards of each branch of the compartment model in Figure 1.1, conditional on a frailty term. In addition, we model the baseline hazards using B-splines (de Boor, 1978).

Splines are piecewise polynomials satisfying continuity constraints at the knots joining the pieces. As the number of knots increases, very flexible families of models are created, so spline methods are a good alternative to exploring the nature of relationships, especially those of a continuous smooth form. The shape of a spline function depends on: (i) the order of the function; (ii) the knot sequence; and (iii) the continuity conditions at each knot. Splines can

be expressed as a linear combination of basis functions — functions that span the space of all piecewise polynomials with a specific order, knot sequence and continuity conditions. Commonly used spline bases include the truncated power basis and the B-spline basis.

Such methods have been considered for modelling in survival analysis. Splines using the truncated power basis were considered by Etezadi-Amoli and Ciampi (1987) in the modelling of the baseline hazard in the extended hazard regression model, which is a generalisation that includes the proportional hazards and accelerated failure time models. Cubic B-splines were used by Sleeper and Harrington (1990) as regression splines (without penalty functions) and by Gray (1992, 1994) through penalized likelihoods to model arbitrary covariate effect as an extension to the Cox model. Rosenberg (1995) also used cubic B-splines for hazard function estimation without covariates. Cubic splines are often sufficient and flexible enough to reflect the changes in the hazard function.

In this thesis, we focus our use of splines to that of cubic B-splines in the fitting of the baseline hazards in our conditional additive risk model for semi-competing risks data. We also apply this method to the NPC dataset and aim to model the hazards of relapse, death without relapse and death after relapse respectively, adjusting for treatment, nodal status and TNM staging.

3.2 Proposed Model and Estimation

3.2.1 Additive Hazards for Semi-Competing Risks

Let T_1 and T_2 be the failure times of the non-terminal and terminal events respectively. There is a censoring time C that is independent of both T_1 and T_2, such as administrative loss to follow-up. Denote the observed failure times

as $Y = \min(T_2, C)$ and $X = \min(T_1, Y)$. We have the corresponding event indicators as $\delta_1 = I(T_1 < Y)$ and $\delta_2 = I(T_2 < C)$ for the occurrence of the non-terminal and terminal events respectively. The observed data for the i-th individual is then $\Omega_i = \{X_i, \delta_{i1}, Y_i, \delta_{i2}, \mathbf{Z}_i\}$, where $\mathbf{Z}_i$ is a $p \times 1$ covariate vector.

Under the additive risk frailty model, conditional on the frailty, the hazards model is an additive one. We have 3 hazard functions for each individual — one for time to relapse, one for time to death without relapse and one for time to death after relapse. Their relationships can be seen in the compartment model shown in Figure 1.1 of Chapter 1. Here we restate the set of of hazard functions for the i-th individual from Chapter 1:

$$\begin{aligned}
\lambda_1(t|T_{i2} \geq t, \gamma_i, \mathbf{Z}_i) &= \gamma_i(\lambda_{01}(t) + \boldsymbol{\beta}_1^T \mathbf{Z}_i(t)), \\
\lambda_2(t|T_{i1} \geq t, \gamma_i, \mathbf{Z}_i) &= \gamma_i(\lambda_{02}(t) + \boldsymbol{\beta}_2^T \mathbf{Z}_i(t)), \\
\lambda_3(t|T_{i1} \leq t, \gamma_i, \mathbf{Z}_i) &= \gamma_i(\lambda_{03}(t) + \boldsymbol{\beta}_3^T \mathbf{Z}_i(t)),
\end{aligned} \tag{3.1}$$

where γ_i $(i = 1, \ldots, n)$ is independent and identically distributed from the Gamma distribution with shape θ^{-1} and scale θ, such that the mean of the frailty parameter is 1 and the variance is θ. We assume the Gamma distribution here for mathematical convenience. Conditional on the frailty, the above hazards are independent for each individual. It is of interest to estimate $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \theta, \Lambda_{01}(t), \Lambda_{02}(t), \Lambda_{03}(t)\}$. In general, $\lambda_3(t_2|t_1, \gamma, \mathbf{Z})$ can depend on both t_1 and t_2. In this thesis, we consider the semi-Markov process, where $\lambda_3(t_2|t_1, \gamma, \mathbf{Z}) = \lambda_3(t_2 - t_1|\gamma, \mathbf{Z})$ which depends only on the time between relapse and death.

A simpler, restricted model that might be of interest is to assume $\beta_2 = \beta_3$ and $\lambda_{02} = \lambda_{03}$. In this restricted model, the dependence between T_1 and T_2 is fully

captured by the frailty parameter, γ.

3.2.2 Estimation of Additive Risk Frailty Model

For the three unknown baseline hazards, we propose to use cubic B-splines with M breakpoints $\{\xi_1, \ldots, \xi_M\}$, making the maximization of the likelihood function over a finite dimensional space. This is in contrast to the nonparametric plug-in estimator provided by Lin and Ying (1994) in the usual survival setting. With splines, the baseline hazard functions can be estimated as

$$g_{0m}(t) = \sum_{j=1}^{J} c_{mj} B_{mj}(t), \qquad m = 1, 2, 3, \tag{3.2}$$

where $\{B_{m1}(t), \ldots, B_{mJ}(t)\}$ are the B-spline basis functions for each of the three hazards facing an individual.

Three important mathematical properties of cubic B-splines (deBoor, 1978) are, for each $m = 1, 2, 3$,

1. $B_{mj}(t) = 0$ if $t \notin (\tau_j, \tau_{j+4})$,
2. $B_{mj}(t) > 0$ if $\tau_j < t < \tau_{j+4}$,
3. $\sum_{j=1}^{J} B_{mj}(t) = 1$,

where the knot sequence τ here is defined to be $\tau = \{a = \tau_1 = \ldots = \tau_4, \tau_5 = \xi_1, \ldots, \tau_{4+M} = \xi_M, \tau_{4+M+1} = \ldots = \tau_{M+8} = b\}$ and a and b are the observed minimum and maximum of the survival times. In this thesis, we consider only splines that do not require estimation of the number (M) and placement of the knots.

We next set up the likelihood under the additive hazards frailty model. Under the assumption of independent hazards given the frailty, the conditional

likelihood for the i-th individual is

$$\begin{aligned} L_i^c &= [\gamma_i\{\lambda_{01}(X_i)+\boldsymbol{\beta}_1^T\mathbf{Z}_i\}]^{\delta_{i1}}[\gamma_i\{\lambda_{02}(Y_i)+\boldsymbol{\beta}_2^T\mathbf{Z}_i\}]^{\delta_{i2}(1-\delta_{i1})} \\ &\quad [\gamma_i\{\lambda_{03}(Y_i)+\boldsymbol{\beta}_3^T\mathbf{Z}_i\}]^{\delta_{i1}\delta_{i2}} exp\Bigg[-\gamma_i\Bigg\{\Lambda_{01}(X_i)+(\boldsymbol{\beta}_1^T\mathbf{Z}_i)X_i+\Lambda_{02}(X_i) \\ &\quad +(\boldsymbol{\beta}_2^T\mathbf{Z}_i)X_i+(\Lambda_{03}(Y_i)-\Lambda_{03}(X_i))+(\boldsymbol{\beta}_3^T\mathbf{Z}_i)(Y_i-X_i)\Bigg\}\Bigg]. \end{aligned}$$

The unconditional observed likelihood for the i-th individual is obtained by integrating out the frailty term, which we have assumed to follow the Gamma distribution with shape θ^{-1} and scale θ. We then obtain the unconditional likelihood

$$\begin{aligned} L_i &= \Big[\lambda_{01}(X_i)+\boldsymbol{\beta}_1^T\mathbf{Z}_i\Big]^{\delta_{i1}}\Big[\lambda_{02}(Y_i)+\boldsymbol{\beta}_2^T\mathbf{Z}\Big]^{\delta_{i2}(1-\delta_{i1})}\Big[\lambda_{03}(Y_i)+\boldsymbol{\beta}_3^T\mathbf{Z}_i\Big]^{\delta_{i1}\delta_{i2}} \\ &\quad (1+\theta)^{\delta_{i1}\delta_{i2}}\Bigg[1+\theta\Bigg\{\Lambda_{01}(X_i)+(\boldsymbol{\beta}_1^T\mathbf{Z}_i)X_i+\Lambda_{02}(X_i) \\ &\quad +(\boldsymbol{\beta}_2^T\mathbf{Z}_i)X_i+(\Lambda_{03}(Y_i)-\Lambda_{03}(X_i))+(\boldsymbol{\beta}_3^T\mathbf{Z}_i)(Y_i-X_i)\Bigg\}\Bigg]^{1/\theta+\delta_{i1}+\delta_{i2}}. \end{aligned}$$

Thus the overall observed loglikelihood for semi-competing risks can be written as

$$\begin{aligned} \ell &= \sum_{i=1}^{n}\Bigg\{\delta_{i1}\log\Big[\lambda_{01}(X_i)+\boldsymbol{\beta}_1^T\mathbf{Z}_i\Big]+\delta_{i2}(1-\delta_{i1})\log\Big[\lambda_{02}(Y_i)+\boldsymbol{\beta}_2^T\mathbf{Z}_i\Big] \\ &\quad +\delta_{i1}\delta_{i2}\log\Big[\lambda_{03}(Y_i)+\boldsymbol{\beta}_3^T\mathbf{Z}_i\Big]+\delta_{i1}\delta_{i2}\log(1+\theta) \\ &\quad -(\frac{1}{\theta}+\delta_{i1}+\delta_{i2})log\Bigg[1+\theta\Bigg\{\Lambda_{01}(X_i)+(\boldsymbol{\beta}_1^T\mathbf{Z}_i)X_i+\Lambda_{02}(X_i) \\ &\quad +\Big(\boldsymbol{\beta}_2^T\mathbf{Z}_i\Big)X_i+(\Lambda_{03}(Y_i)-\Lambda_{03}(X_i))+(\boldsymbol{\beta}_3^T\mathbf{Z}_i)(Y_i-X_i)\Bigg\}\Bigg]\Bigg\}, \quad (3.3) \end{aligned}$$

where $\lambda_{0m}(t)$ is estimated by Equation (3.2) and $\Lambda_{0m}(t)$ $(m=1,2,3)$ is esti-

mated by

$$G_{0m}(t) = \sum_{j=1}^{J} c_{mj} \int_0^t B_{mj}(u)du. \tag{3.4}$$

The loglikelihood equation (3.3) is the function we aim to maximise with respect to the parameters $\boldsymbol{\vartheta}_n = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T, \theta, \varphi_n = (G_{01}, G_{02}, G_{03})^T)$. Score functions and the hessian matrix are accordingly calculated for the information matrix.

3.3 Theoretical Properties

By parameterizing the baseline hazard, we consider the space Θ_n, spanned by the parameters we aim to estimate, in this case, $\boldsymbol{\vartheta}_n$, which are the regression coefficients $\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \boldsymbol{\beta_3}$, the variance of the Gamma frailty distribution, θ, and the spline approximations to the baseline hazard functions, φ_n. This space is known as a *sieve space* (Grenander, 1981) and approximates the infinite-dimensional parameter space Θ, spanned by $\{\beta_1, \beta_2, \beta_3, \theta, \Lambda_{01}(t), \Lambda_{02}(t), \Lambda_{03}(t)\}$. Here, the estimator $\hat{\boldsymbol{\vartheta}}_n$ is obtained by maximizing the empirical criterion function $\ell_n(\boldsymbol{\vartheta}) = (1/n)\sum_{i=1}^n \ell(\boldsymbol{\vartheta}, \Omega_i)$, where $\ell_n(\boldsymbol{\vartheta})$ is the loglikelihood function defined in Equation (3.3) multiplied by a constant (1/n) and $\ell(\boldsymbol{\vartheta}, \Omega_i)$ is the contribution of the i-th individual to the loglikelihood.

In this section, we assume that T_1 and T_2 are bounded on $[0, 1]$. Define H_r as the collection of all functions on $[0, 1]$ whose mth order derivative satisfies the Hölder condition of order v with $r = m + v$. That is, for any $h \in H_r$, there exists a positive constant c such that $|h^{(m)}(s) - h^{(m)}(t)| \leq c|s - t|^v$, for any $0 \leq s, t \leq 1$. Let C_r be the space of all bivariate functions $h(t_1, t_2)$ on $[0, 1]^2$ such that $D^{(u_1,u_2)}h = \partial^{u_1+u_2}h/\partial^{u_1}t_1\partial^{u_2}t_2$ is continuous and Lipschitz of order v: $\left\|D^{(u_1,u_2)}(T) - D^{(u_1,u_2)}(S)\right\| \leq W_0|T - S|^v$ for any $T, S \in [0, 1]^2$ and $u_1 + u_2 < r - v$, where W_0 is a finite constant. Also let P_{ϑ_0} be the probability

measure under the true model parameters and P_n be the empirical probability measure.

Before we examine some of the properties of $\hat{\boldsymbol{\vartheta}}_n$, the MLE of the sieve space, we first give some conditions needed for the following theorems. Denote $\boldsymbol{\vartheta}_0$ as the true values of the parameters to be estimated.

1. T_1 and T_2 have supports on $[0, 1]$ and $\mathbf{Z}$ has bounded support on $\mathbb{R}^p$ where p is the dimension of $\mathbf{Z}$.

2. $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{10}^T, \boldsymbol{\beta}_{20}^T, \boldsymbol{\beta}_{30}^T)^T \in B$ where B is a compact set in $\mathbb{R}^{3p}$ with nonempty interior. Also, $\lambda_{01}, \lambda_{02} \in H_r$ and $\lambda_{03} \in C_r$.

3. $r \geq 2$ where r is the measure of smoothness of λ_j in the definitions of H_r and C_r.

Theorem 3.3.1. (Consistency of estimator $\hat{\boldsymbol{\vartheta}}_n$). *Suppose the conditions (1–3) above hold true, then $\hat{\boldsymbol{\vartheta}}_n$ is a consistent estimator of $\boldsymbol{\vartheta}_0$.*

Proof. The proof follows closely to Xue *et al.* (2004). Let $F_n = \{\ell(\boldsymbol{\vartheta}, \cdot) : \boldsymbol{\vartheta} \in \Theta_n\}$ and $N(\epsilon_n, F_n, P_n)$ be the covering number defined by Pollard (1984). Using the results of the symmetrization lemma and II.31 of Pollard (1984), we have as $n \longrightarrow \infty$

$$P(\sup_{F_n} |P_n\ell - P\ell| > 8\epsilon_n) \longrightarrow 0.$$

Hence

$$\sum_{i=1}^{\infty} P\Big\{ \sup_{F_n} |P_n\ell - P\ell| > 8\epsilon_n \Big\} < +\infty,$$

where $P = P_{\vartheta_0}$ and by Borel-Cantelli lemma, under P_{ϑ_0},

$$\sup_{\boldsymbol{\vartheta} \in \Theta_n} |P_n\ell - P\ell| \longrightarrow 0, \qquad a.s.. \tag{3.5}$$

Now let

$$\begin{aligned} \zeta_{1n} &= \sup_{\boldsymbol{\vartheta}\in\Theta_n} |P_n\ell - P\ell|, \\ \zeta_{2n} &= P_n\ell(\boldsymbol{\vartheta}_0, \cdot) - P\ell(\boldsymbol{\vartheta}_0, \cdot). \end{aligned} \tag{3.6}$$

Denote $K_\epsilon = \{\boldsymbol{\vartheta} : d(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0) \geq \epsilon, \boldsymbol{\vartheta} \in \Theta_n\}$, where we define a distance metric d as the sum of the L_2 distance between each parameter in $\boldsymbol{\vartheta}$, based on the probability measure of (T_1, T_2). Then

$$\begin{aligned} \inf_{K_\epsilon} P\ell(\boldsymbol{\vartheta}, \Omega) &= \inf_{K_\epsilon} \Big\{ P\ell(\boldsymbol{\vartheta}, \Omega) - P_n\ell(\boldsymbol{\vartheta}, \Omega) + P_n\ell(\boldsymbol{\vartheta}, \Omega) \Big\} \\ &\leq \zeta_{1n} + \inf_{K_\epsilon} P_n\ell(\boldsymbol{\vartheta}, \Omega) \end{aligned} \tag{3.7}$$

If $\hat{\boldsymbol{\vartheta}}_n \in K_\epsilon$, we have

$$\begin{aligned} \inf_{K_\epsilon} P_n\ell(\boldsymbol{\vartheta}, \Omega) &= P_n\ell(\hat{\boldsymbol{\vartheta}}_n, \Omega) \\ &\leq P_n\ell(\boldsymbol{\vartheta}_0, \Omega) \\ &= \zeta_{2n} + P\ell(\boldsymbol{\vartheta}_0, \Omega) \end{aligned} \tag{3.8}$$

The choice of the sieve space results in $\inf_{K_\epsilon} P\ell(\boldsymbol{\vartheta}, \Omega) - P\ell(\boldsymbol{\vartheta}_0, \Omega) = \delta_\epsilon$, where $\delta_\epsilon > 0$. Hence, by (3.7) and (3.8),

$$\begin{aligned} \inf_{K_\epsilon} P\ell(\boldsymbol{\vartheta}, \Omega) &\leq \zeta_{1n} + \zeta_{2n} + P\ell(\boldsymbol{\vartheta}_0, \Omega) \\ &= \zeta_n + P\ell(\boldsymbol{\vartheta}_0, \Omega), \end{aligned}$$

where $\zeta_n = \zeta_{1n} + \zeta_{2n}$. Thus we get $\zeta_n \geq \delta_\epsilon$. Furthermore, we have $\{\hat{\boldsymbol{\vartheta}}_n \in K_\epsilon\} \subseteq \{\zeta_n \geq \delta_\epsilon\}$. By (3.5) and the Strong Law of Large Numbers, we have $\zeta_{1n} = o(1)$, $\zeta_{2n} = o(1)$ a.s.. Therefore, by $\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \{\hat{\boldsymbol{\vartheta}}_n \in K_\epsilon\} \subseteq \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \{\zeta_n \geq \delta_\epsilon\}$, we get $d(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}_0) \xrightarrow{P} 0$ and thus prove the consistency of $\hat{\boldsymbol{\vartheta}}_n$. □

Theorem 3.3.2. (Asymptotic normality of $\hat{\vartheta}_n$). *Let* $\eta = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T, \theta)^T$. *If the conditions (1–3) above hold, then* $\sqrt{n}(\hat{\eta} - \eta_0)$ *is asymptotically normal.*

Proof. This proof follows closely to Murphy (1995). To obtain the score functions, one can use the usual method of differentiating the loglikelihood ℓ_n with respect to the parameters given in $\boldsymbol{\vartheta}$. An equivalent method would be to consider one-dimensional submodels through the estimators and differentiate at the estimator. Set $\boldsymbol{\beta}_{jt} = th_{1j} + \boldsymbol{\beta}_j$, $\theta_t = th_2 + \theta$ and $\Lambda_{0jt}(\cdot) = \int_0^{\cdot}[1 + th_{3j}(u)]d\hat{\Lambda}_{0j}(u)$ for $j = 1, 2, 3$, where h_{1j} are vectors of length p, h_2 is a scalar and h_{3j} are functions. Use these models in ℓ_n and differentiate and set $t = 0$ to obtain $S_n(\boldsymbol{\vartheta})$. Note that the estimator that maximises ℓ_n will have $S_n(\hat{\boldsymbol{\vartheta}}) = 0$. S_n can be written as $S_n = S_{n1} + S_{n2} + S_{n3}$, where

$$\begin{aligned} S_{n1}(\boldsymbol{\vartheta})(h_1) &= \sum_{i=1}^{n}\Big\{\frac{\delta_{i1}h_{11}^T\mathbf{Z}_i}{\lambda_{01}(X_i) + \boldsymbol{\beta}_1^T\mathbf{Z}_i} + \frac{\delta_{i2}(1-\delta_{i1})h_{12}^T\mathbf{Z}_i}{\lambda_{02}(X_i) + \boldsymbol{\beta}_2^T\mathbf{Z}_i} + \frac{\delta_{i1}\delta_{i2}h_{13}^T\mathbf{Z}_i}{\lambda_{03}(Y_i) + \boldsymbol{\beta}_3^T\mathbf{Z}_i} \\ &\quad -(\frac{1}{\theta} + \delta_{i1} + \delta_{i2})\frac{\theta[h_{11}^T\mathbf{Z}_iX_i + h_{12}^T\mathbf{Z}_iX_i + h_{13}^T\mathbf{Z}_i(Y_i - X_i)]}{1 + \theta\Lambda}\Big\}, \end{aligned}$$

$$S_{n2}(\boldsymbol{\vartheta})(h_2) = \sum_{i=1}^{n}\Big\{\frac{\delta_{i1}\delta_{i2}h_2}{1+\theta} + \frac{h_2}{\theta^2}\log[1+\theta\Lambda] - (\frac{1}{\theta} + \delta_{i1} + \delta_{i2})\frac{h_2\Lambda}{1+\theta\Lambda}\Big\},$$

and

$$\begin{aligned} S_{n3}(\boldsymbol{\vartheta})(h_3) &= \sum_{i=1}^{n}\Big\{\frac{\delta_{i1}h_{31}(X_i)\lambda_{01}(X_i)}{\lambda_{01}(X_i) + \boldsymbol{\beta}_1^T\mathbf{Z}_i} + \frac{\delta_{i2}(1-\delta_{i1})h_{32}(X_i)\lambda_{02}(X_i)}{\lambda_{02}(X_i) + \boldsymbol{\beta}_2^T\mathbf{Z}_i} \\ &\quad + \frac{\delta_{i1}\delta_{i2}h_{33}(Y_i)\lambda_{03}(Y_i)}{\lambda_{03}(Y_i) + \boldsymbol{\beta}_3^T\mathbf{Z}_i} - (\frac{1}{\theta} + \delta_{i1} + \delta_{i2})\theta\Big[\frac{\int_0^{X_i} h_{31}(u)\lambda_{01}(u)du}{1+\theta\Lambda} \\ &\quad + \frac{\int_0^{X_i} h_{32}(u)\lambda_{02}(u)du + \int_{X_i}^{Y_i} h_{33}(u)\lambda_{03}(u)du}{1+\theta\Lambda}\Big]\Big\}, \end{aligned}$$

where $h_1 = (h_{11}^T, h_{12}^T, h_{13}^T)$, $h_3 = (h_{31}, h_{32}, h_{33})$, $\Lambda = \Lambda_1(X_i) + \Lambda_2(X_i) + (\Lambda_3(Y_i) - \Lambda_3(X_i))$ and $\Lambda_j(t) = \Lambda_{0j}(t) + \boldsymbol{\beta}_j^T\mathbf{Z}_it$ for $j = 1, 2, 3$.

We begin by proving the conditions in Theorem 2 of Murphy (1995). We obtain the condition that $\sqrt{n}(S_n(\boldsymbol{\vartheta}_0) - S(\boldsymbol{\vartheta}_0))$ converges weakly to a normal distribu-

tion with mean 0, by central limit theorem and continuous mapping theorem, where S_n is the empirical score function of the likelihood and S is the asymptotic version of S_n. The approximation condition is a technical proof and is omitted here.

Denote $\dot{S}(\boldsymbol{\vartheta}_0)$ as the linear operator on the set of all linear combinations of $(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)$ for all $\boldsymbol{\vartheta} \in \Theta$. The classical relationship between the asymptotic variance of the score function (the information matrix for $\boldsymbol{\vartheta}$) and the derivative of the score equation, $-\dot{S}(\boldsymbol{\vartheta}_0)$ holds. By writing $S(\boldsymbol{\vartheta})$ as a linear combination of $\hat{\beta} - \beta_0$, $\hat{\theta} - \theta_0$ and $d(\hat{\Lambda}_{0j} - \Lambda_{0j})$ plus error terms, we obtain $-\dot{S}(\boldsymbol{\vartheta}_0)(\boldsymbol{\vartheta}_0)(h)$ as the variance for $\sqrt{n}S_n(\boldsymbol{\vartheta}_0)(h)$. We thus get

$$S(\hat{\boldsymbol{\vartheta}}) - S(\boldsymbol{\vartheta}_0) = \dot{S}(\boldsymbol{\vartheta}_0)(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0).$$

Given the convergence to normality of $\sqrt{n}(S_n(\boldsymbol{\vartheta}_0) - S(\boldsymbol{\vartheta}_0))$ and the fact that with S being the asymptotic vrsion of S_n, $S(\boldsymbol{\vartheta}_0) = 0$ and $S_n(\hat{\boldsymbol{\vartheta}}) = 0$, we can now write

$$\sqrt{n}(S_n(\boldsymbol{\vartheta}_0) - S(\boldsymbol{\vartheta}_0)) = \sqrt{n}(S(\boldsymbol{\vartheta}_0) - S(\hat{\boldsymbol{\vartheta}})) - \sqrt{n}((S_n - S)(\hat{\boldsymbol{\vartheta}}) - (S_n - S)(\boldsymbol{\vartheta}_0)),$$

which in turn can be rewritten as

$$\sqrt{n}(S_n(\boldsymbol{\vartheta}_0) - S(\boldsymbol{\vartheta}_0)) = \sqrt{n}\dot{S}(\boldsymbol{\vartheta}_0)(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_P(1).$$

With the asymptotic normality of $\sqrt{n}(S_n(\boldsymbol{\vartheta}_0) - S(\boldsymbol{\vartheta}_0))$ and the invertibility of $\dot{S}(\boldsymbol{\vartheta}_0)$, we get the convergence in distribution of $\sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0)$ to the normal

distribution. We now write

$$\sqrt{n}g_1^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \sqrt{n}g_2(\hat{theta} - \theta_0) + \sum_{j=1}^{3} \int \sqrt{n}g_{3j}d(\hat{\Lambda}_{0j} - \Lambda_{0j}),$$

where $(g_1^T, g_2, g_{31}, g32, g33)$ are coefficients to form linear combinations of $\sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0)$. Thus, by Cramer-Wold device (van der Vaart, 1998) and setting $g_{3j} = 0$ for $j = 1, 2, 3$, we have the asymptotic normality of $\sqrt{n}(\hat{\eta} - \eta_0)$.

3.4 Simulation Studies Based on Additive Risks Frailty Model

3.4.1 Data Generation for Semi-Competing Risks

Based on the conditional independence of the hazard functions, we first generate covariate vector Z and the gamma frailty term. Given the covariate values and the frailty, we then generate two independent times, T_1 (time to non-terminal event) and T_2 (time to terminal event), from their respective distributions based on λ_1 and λ_2 from Equation (3.1). If $T_1 > T_2$, then we consider the non-terminal event as censored. Conversely, if $T_1 < T_2$, then the non-terminal event has occurred and we generate a third time from the third branch of the compartment model to obtain the time to the terminal event after the non-terminal event has occurred. Censoring times, C, are generated independently of T and from the Uniform$[0, a]$ distribution for some $a > 0$.

3.4.2 Simulation Results for Additive Risk Frailty Model

We consider the simple case of a single discrete covariate Z generated from the Bernoulli(0.5) distribution. Simulations were conducted for a sample size of

400, and at two degrees of censoring, 10% and 30%. We fix the baseline hazard to be from the exponential distribution with mean 1. Two sets of true values are considered for the three regression coefficients of the respective hazards — $(\beta_1, \beta_2, \beta_3) = (1, 1, 1)$ and $(\beta_1, \beta_2, \beta_3) = (0.5, 1, 2)$; and the true value of θ for the Gamma frailty distribution is first set at 0.95.

We generate 1000 simulations for each combination of censoring, regression coefficients and θ. Table 3.1 shows the results and for each parameter combination gives: (i) the Estimate (Est), the average from 1000 estimates; (ii) the Empirical Standard Error (EmpSE), the standard deviation of the 1000 estimators; (iii) the Model Standard Error (ModSE), the average of 1000 standard error estimators, using the information matrix. These notations are similar to those reported in the simulations in Chapter 2, where Est, EmpSE and ModSe are similar to $AVE(\hat{\beta}_1)$, $SD(\hat{\beta}_1)$ and $AVE(\hat{SD})$ in Chapter 2 respectively. We also conducted 1000 simulations for different levels of censoring and regression coefficients, and varying θ. Tables 3.2 and 3.3 shows the results for $\theta = 0.5$ and 1.5 respectively. The tables indicate that the biases are relatively small and are about 10% of the true value. All the simulations indicate that larger values of the parameters result in larger standard errors. Similar observations can be made for higher percentages of censoring. In addition, standard errors increase slightly with censoring. Varying values of θ do not affect the estimates much.

Figure 3.1 shows the plots of the spline estimators of the basline survival function for the time to relapse. Under the simulation, the true survival function is $S_0(t) = e^{-t}$ and is indicated on the diagram with the solid black line. The grey lines are the estimators obtained from each of the 1000 samples. As can be seen, the spline estimators approximate the general shape of the true survival function well.

Table 3.1: Estimators for Additive Risk Frailty Model for Semi-Competing Risks Data with single Z from Bernoulli(0.5) and $\theta = 0.95$, with 10% and 30% censoring and varying β.

			$(\beta_1, \beta_2, \beta_3) = (1, 1, 1)$		
[0,a]	*Cens*		**Est**	**EmpSE**	**ModSE**
$[0, 12]$	10%	β_1	1.094	0.38	0.38
		β_2	1.066	0.36	0.38
		β_3	1.092	0.19	0.23
		θ	1.045	0.17	0.19
$[0, 2.4]$	30%	β_1	1.108	0.40	0.41
		β_2	1.087	0.38	0.41
		β_3	1.090	0.46	0.51
		θ	1.071	0.23	0.24

			$(\beta_1, \beta_2, \beta_3) = (0.5, 1, 2)$		
[0,a]	*Cens*		**Est**	**EmpSE**	**ModSE**
$[0, 12]$	10%	β_1	0.549	0.28	0.31
		β_2	1.066	0.35	0.37
		β_3	2.162	0.64	0.67
		θ	1.046	0.17	0.20
$[0, 2.4]$	30%	β_1	0.584	0.31	0.34
		β_2	1.105	0.39	0.41
		β_3	2.154	0.72	0.77
		θ	1.083	0.22	0.25

Table 3.2: Estimators for Additive Risk Frailty Model for Semi-Competing Risks Data with single Z from Bernoulli(0.5) and $\theta = 0.5$, with 10% and 30% censoring and varying β.

			$(\beta_1, \beta_2, \beta_3) = (1, 1, 1)$		
[0,a]	*Cens*		**Est**	**EmpSE**	**ModSE**
$[0, 6.6]$	10%	β_1	1.127	0.35	0.35
		β_2	1.117	0.35	0.34
		β_3	1.120	0.40	0.41
		θ	0.608	0.16	0.16
$[0, 1.7]$	30%	β_1	1.132	0.38	0.39
		β_2	1.120	0.37	0.39
		β_3	1.105	0.45	0.49
		θ	0.625	0.21	0.21

			$(\beta_1, \beta_2, \beta_3) = (0.5, 1, 2)$		
[0,a]	*Cens*		**Est**	**EmpSE**	**ModSE**
$[0, 6.6]$	10%	β_1	0.562	0.27	0.28
		β_2	1.103	0.34	0.33
		β_3	2.182	0.59	0.61
		θ	0.592	0.16	0.16
$[0, 1.7]$	30%	β_1	0.591	0.30	0.32
		β_2	1.140	0.37	0.38
		β_3	2.194	0.69	0.73
		θ	0.640	0.22	0.22

Table 3.3: Estimators for Additive Risk Frailty Model for Semi-Competing Risks Data with single Z from Bernoulli(0.5) and $\theta = 1.5$, with 10% and 30% censoring and varying β.

			$(\beta_1, \beta_2, \beta_3) = (1, 1, 1)$		
[0,a]	*Cens*		**Est**	**EmpSE**	**ModSE**
$[0, 30]$	10%	β_1	1.085	0.41	0.42
		β_2	1.067	0.41	0.41
		β_3	1.081	0.46	0.47
		θ	1.616	0.23	0.23
$[0, 3.9]$	30%	β_1	1.104	0.43	0.45
		β_2	1.106	0.42	0.45
		β_3	1.103	0.50	0.54
		θ	1.651	0.28	0.28

			$(\beta_1, \beta_2, \beta_3) = (0.5, 1, 2)$		
[0,a]	*Cens*		**Est**	**EmpSE**	**ModSE**
$[0, 30]$	10%	β_1	0.571	0.32	0.34
		β_2	1.099	0.39	0.41
		β_3	2.192	0.69	0.72
		θ	1.640	0.24	0.23
$[0, 3.9]$	30%	β_1	0.580	0.34	0.37
		β_2	1.124	0.43	0.45
		β_3	2.207	0.79	0.82
		θ	1.662	0.29	0.31

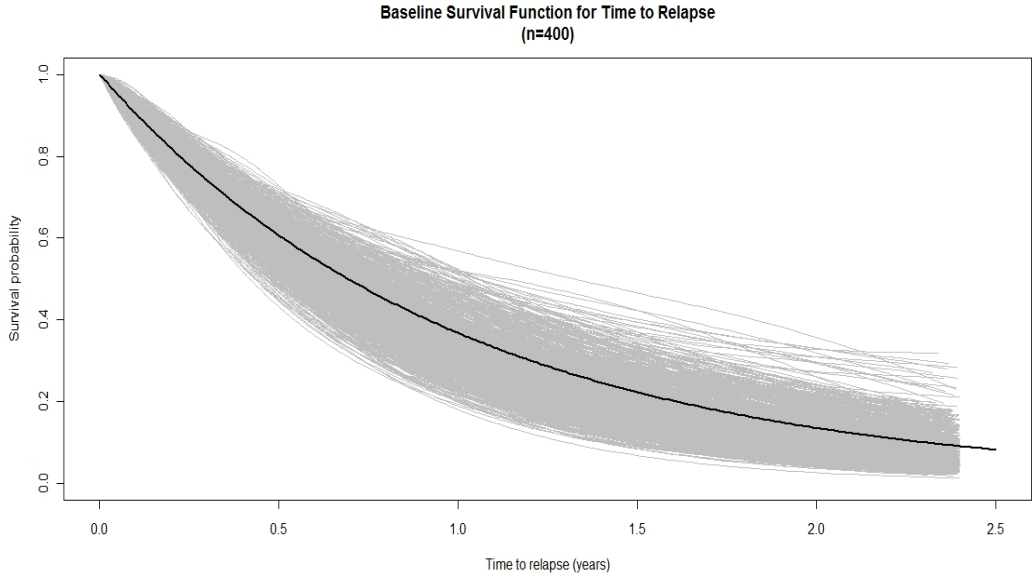

Figure 3.1: Plot of 1000 estimators of the baseline survival functions for time to relapse. Estimators obtained using splines. Bold line indicates true survival function.

3.5 Application to NP01 Clinical Trial

The dataset used in this chapter was briefly mentioned in Chapter 1 and is a randomised clinical trial conducted in Singapore comparing two treatments on nasopharyngeal cancer (Wee *et al.*, 2005). In the trial (NP01) conducted between September 1997 and May 2003, patients may experience the non-terminal event of interest — cancer relapse of any type, and/or the terminal event — death. There were 221 patients who were randomly assigned in total, with 110 receiving radiotherapy (RT) alone and 111 receiving chemoradiotherapy (CRT). Of the 110 patients receiving radiotherapy alone, there were 48 relapses and 44 deaths (of which 39 were disease related); of the 111 patients receiving chemoradiotherapy, there were 27 relapses and 24 deaths (of which 21 were disease related). The median follow-up time was 3.2 years. Both treatment groups were well balanced with respect to most characteristics — gender, race, tumour size, nodal status, TNM staging (disease stage). Full details of the trial are

provided by Wee *et al.* (2005).

The NP01 dataset has been analysed using the competing risks approach to compute the cumulative incidence function. However, this approach does not take into account the additional information due to the unique dependent censoring. The competing risks approach would not utilise the information from the survival times of patients who have suffered relapse, but can still be analysed for time to death (the upper wedge). Here, we aim to use the model from Section 3.2 to incorporate this special nature of the data.

In this trial, two endpoints were of interest — time to relapse and time to mortality. Although death was the primary outcome, relapse was also considered to be an important endpoint because it has been found that a substantial proportion of patients with Stage III or IV endemic NPC relapsed locoregionally and/or systematically with RT alone. Table 3.4 shows the breakdown of relapse and deaths according to the compartment model for each treatment group.

Table 3.4: Number of relapses and deaths in each treatment group (data from Wee *et al.*, 2005).

	No. of patients		
Event	CRT	RT	Total
Relapse without death	11	9	20
Death without relapse	8	5	13
Relapse with death	16	39	55
Total	35	53	88

We fit the additive risk frailty model, using the Gamma distribution with shape θ^{-1} and scale θ. As an initial analysis, we fit only the treatment covariate, with

the patients receiving CRT as the reference group. Table 3.5 shows the result of the model fitting. Treatment was significant in the relapse arm and only marginally significant in the death after relapse arm. The estimate indicates that the addition of chemotherapy to radiotherapy seems to decrease the hazard of relapse of an individual by 0.0746 for an individual receiving CRT as compared to one receiving RT alone. A similar observation can be made for hazard of death after relapse. The instantaneous risk of death after relapse decreases by 0.353 for an individual receiving CRT as compared to one receiving RT alone. The estimate of the variance of the frailty parameter was $\hat{\theta} = 0.00012$. This indicates little relationship between the three arms of the compartment model for the semi-competing risks data of NP01 under this gamma frailty model.

It can also be observed that under the semi-competing risks setting, the treatment effect is different in the two death arms. For death without relapse, the addition of chemotherapy resulted in worser survival rates as compared to patients receiving only RT, although this effect was not significant. In contrast, patients receiving CRT had better chances of survival from death after relapse as compared to patients receiving RT and this effect was marginally significant.

Table 3.5: Estimation of treatment effect based on Additive Risk Frailty Model for Semi-competing Risks (data from Wee *et al.*, 2005).

	Relapse	Death without Relapse	Death after Relapse
Treatment [a]	0.075 *	−0.013	0.353
	(0.033)	(0.021)	(0.205)
Loglikelihood	−329.854		

Standard errors given in parentheses and * indicates significance at 5% level.
[a]CRT as reference group

Figure 3.2 shows the survival curves comparing the treatment effect under the

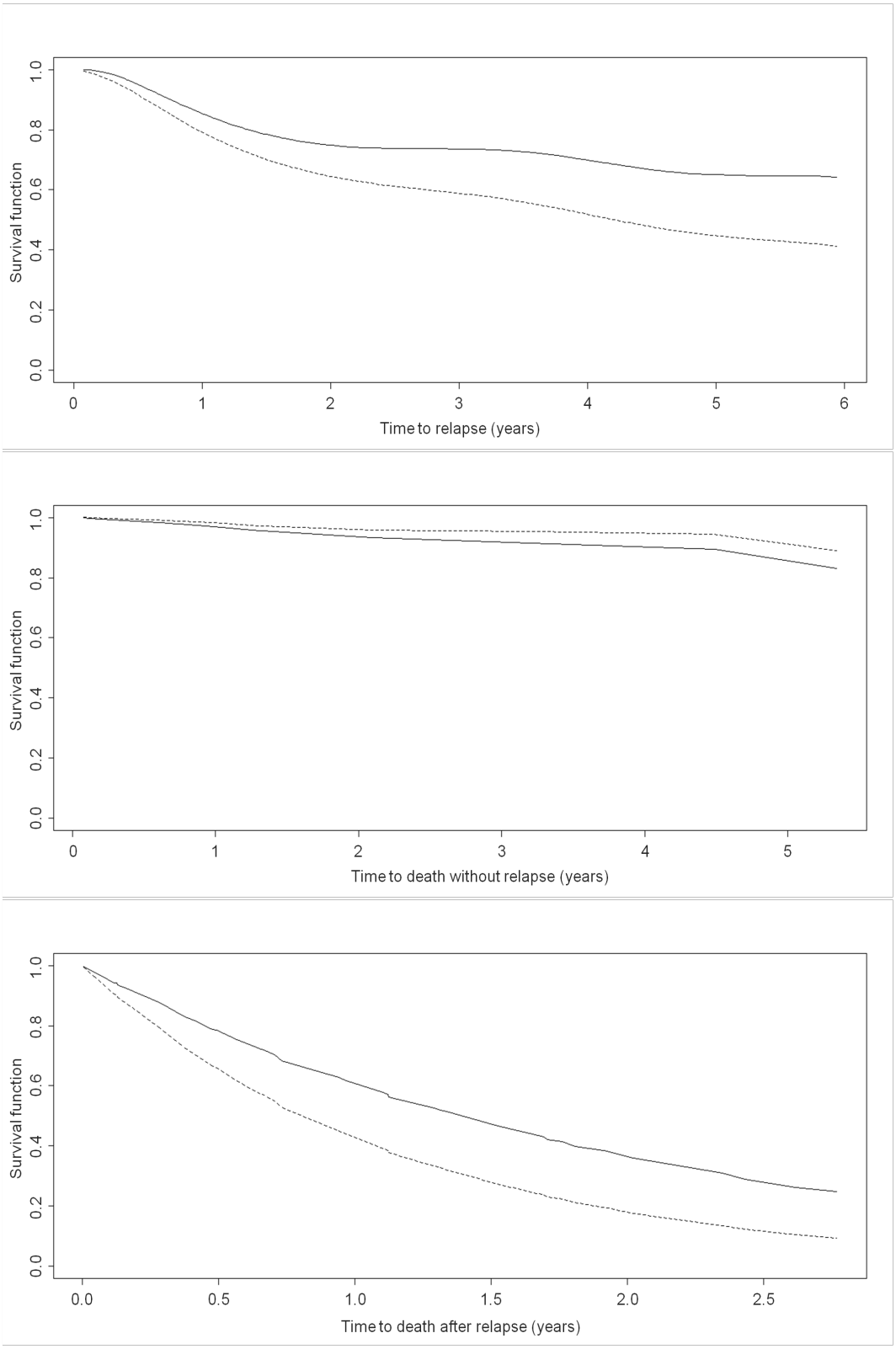

Figure 3.2: Survival functions comparing treatment effect on time to (from top to bottom): (a) relapse; (b) death without relapse; and (c) death after relapse (data from Wee *et al.*, 2005). — gives the survival function for patients receiving CRT; - - - gives the survival function for patients receiving RT.

above model. The cumulative effect of an additive risk model can be seen over time. Based on Figure 3.2, the 2- and 3-year disease-free survival rates were about 75% *v* 65% and 73% *v* 58% for locally advanced NPC patients in the CRT *v* RT group respectively. As for overall survival rate, we can observe the two arms of death without relapse and death after relapse. In the death without relapse arm, there was little treatment effect, but estimation indicates that the addition of chemotherapy to radiotherapy results in a worse survival rate; the 2- and 3-year survival rate was about 92% *v* 96% and 90% *v* 95% for patients in the CRT *v* RT groups respectively. In the death after relapse arm, the survival rates are worse, with the 1- and 2-year survival rate after relapse has occurred at about 62% *v* 44% and 36% *v* 16% for patients in the CRT *v* RT groups respectively. As expected, patients who suffer relapse had worser survival rates as compared to those who died without relapse, in particular, patients who received RT appear to have less than 50% survival chance if they experience relapse as compared to patients in the same treatment group who did not suffer relapse.

We also fit the additive risk frailty model to account not only for treatment, but also for nodal status (stratifying between N0–2 and N3) and tumour size (TNM staging, stratifying between Stage 2–3 and 4). Table 3.6 shows the results of the fitting. The conclusions with regard to statistical significance of individual covariates are the same as before, even after accounting for the strata of nodal status and TNM staging. One reason for this observation could be that the treatment groups were well-balanced with respect to most characteristics and hence, we would not expect the inclusion of these covariates to affect our analysis. The frailty variance was estimated at 0.006 (SE=0.022), which implies an insignificant dependence relationship between the three arms of the relapse-death model. Figure 3.3 shows the survival curves comparing the treatment

effect on relapse under the adjusted model. As can be seen, patients with higher nodal status and TNM staging have lower recurrence-free rates.

Table 3.6: Estimators for Additive Risk Frailty Model for Semi-competing Risks, accounting for treatment, nodal status and TNM staging (data from Wee *et al.*, 2005).

	Relapse	Death without Relapse	Death after Relapse
Treatment [a]	0.071 *	−0.001	0.355
	(0.033)	(0.058)	(0.210)
Nodal status [b]	0.038	−0.009	0.018
	(0.056)	(0.062)	(0.252)
TNM staging [c]	0.038	0.024	0.027
	(0.038)	(0.100)	(0.275)
Loglikelihood	−328.136		

Standard errors given in parentheses and * indicates significance at 5% level.
[a]CRT as reference group
[b]N0–2 as reference group
[c]Stage 3 as reference group

In both models analysed so far, the frailty parameter seems to indicate a lack of association between relapse and death. Thus, we consider the restricted model, which assumes that the two death arms share the same hazard function, conditional on the frailty, that is, $\beta_2 = \beta_3$ and $\lambda_{02} = \lambda_{03}$. Table 3.7 shows the results of the restricted model evaluating treatment effect only.

The estimates obtained show significant protective effect of CRT as compared to RT for both the relapse and death outcomes, although the effect in the relapse arm is only significant at the 10% level. Also, the protective effect increases in the death arm, as can be seen from the larger value of the estimate. In addition, the frailty parameter is now estimated at 1.28 (SE=0.335) and is now

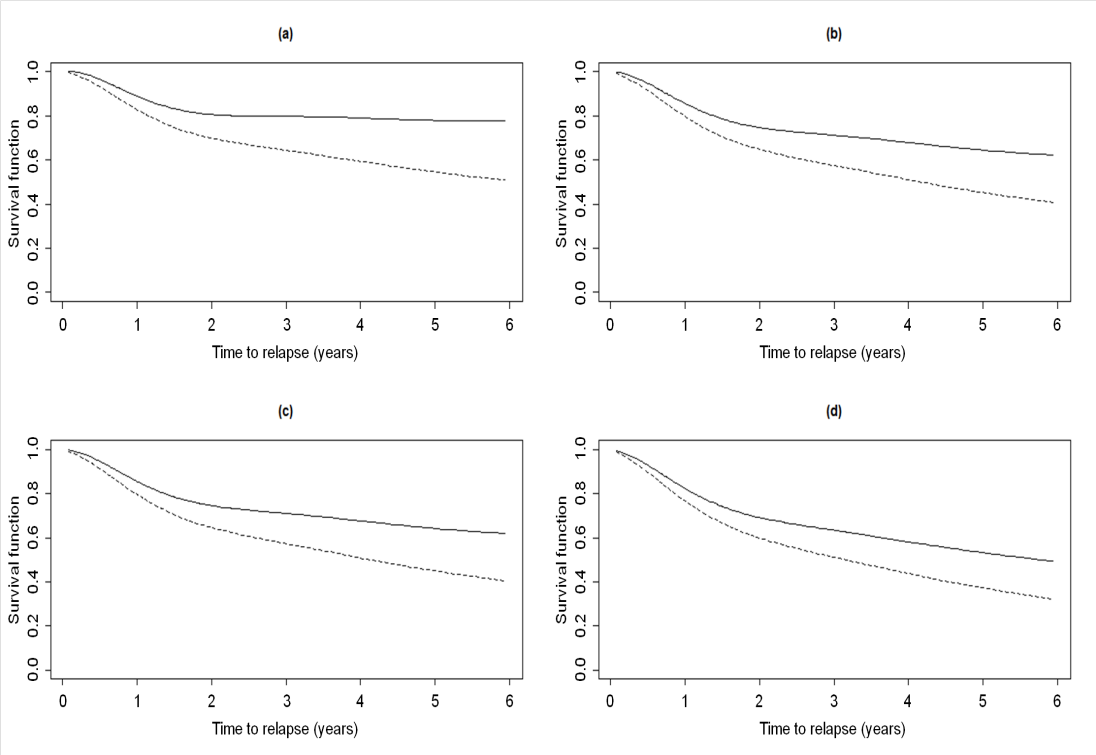

Figure 3.3: Survival functions comparing treatment effect on time to relapse, stratifying for nodal status and TNM staging: (a) Nodal status N0–2, TNM Stage 2–3; (b) Nodal status N3, TNM Stage 2–3; (c) Nodal status N0–2, TNM Stage 4; (d) Nodal status N3; TNM Stage 4 (data from Wee *et al.*, 2005). — gives the survival function for patients receiving CRT; - - - gives the survival function for patients receiving RT.

Table 3.7: Estimation of treatment effect based on Restricted Additive Risk Frailty Model for Semi-Competing Risks (data from Wee *et al.*, 2005).

	Relapse	Death
Treatment [a]	0.088	0.106 *
	(0.052)	(0.035)
Loglikelihood	−364.44	

Standard errors given in parentheses and * indicates significance at 5% level.
[a]CRT as reference group

significant, whereas the estimate under the general (three arms) compartment model was close to 0 and not significant. This new estimate indicates that there is an association between relapse and death. The survival rates for relapse and death can be examined in Figure 3.4. The 2- and 3-year disease-free survival can be seen from the survival functions for relapse and are estimated at 83% *v* 73% and 80% *v* 66% for the CRT *v* RT groups respectively. The 2- and 3-year overall survival rates are estimated from the survival functions for death and are 87% *v* 75% and 81% *v* 65% for the CRT *v* RT groups respectively. These estimates are close to the ones obtained by Wee *et al.* (2005).

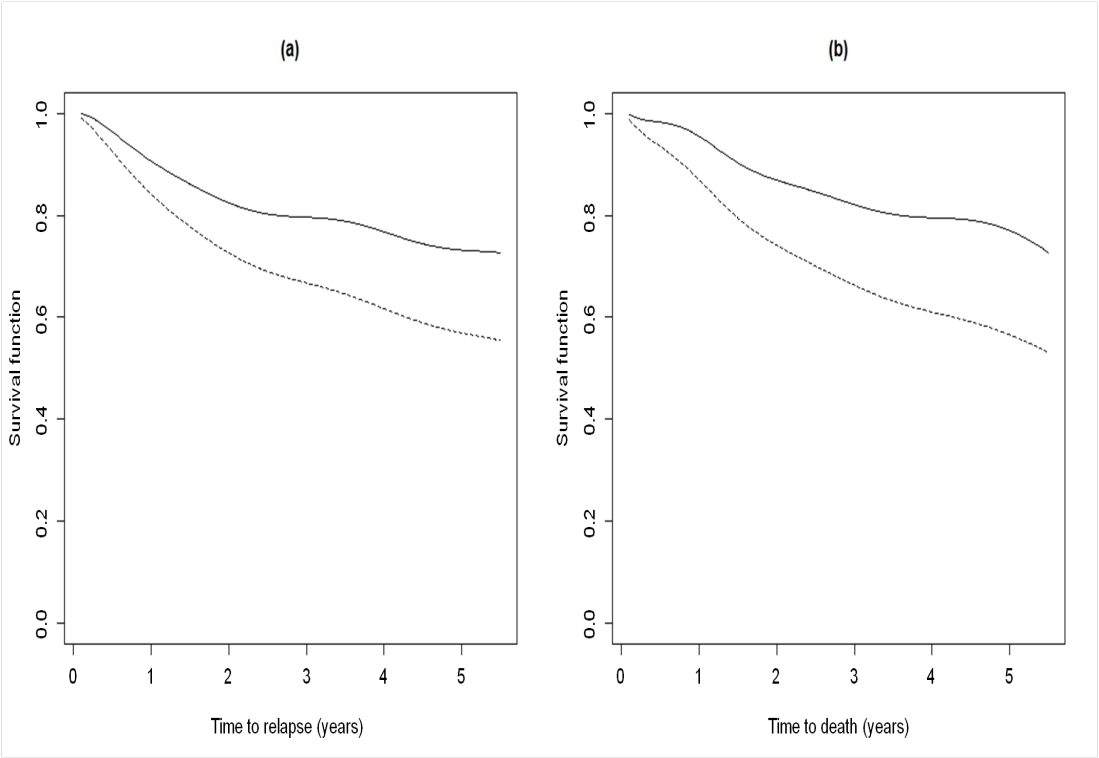

Figure 3.4: Survival functions comparing treatment effect on: (a) time to relapse and (b) time to death, for an individual under restricted additive model (data from Wee *et al.*, 2005). — gives the survival function for patients receiving CRT; - - - gives the survival function for patients receiving RT.

We also conducted a likelihood ratio test to see if the general model was a better fit for the data, or if the restricted model was sufficient. The test statistic was

69.18 and yielded a p-value of <0.001 under the χ^2 distribution with 9 degrees of freedom. This indicates strong evidence against the restricted model, when we consider our formulation of the general model.

In a similar fashion, we also fit the additive restricted frailty model and adjusted for treatment, nodal status and TNM staging. The estimates are given in Table 3.8. Compared to the restricted model with treatment only, the estimates for treatment effect are similar and only the treatment estimates are marginally significant in this model. The estimate of the frailty parameter was 2.01 (SE=0.502) and was highly significant, indicating a strong association betwen relapse and death. A likelihood ratio test was used to compare this adjusted model against the same adjusted general compartment model fitted earlier. The test statistic obtained was 46.8 with a p-value <0.001 under the χ^2 distribution with 11 degrees of freedom, indicating strong evidence against the restricted model.

Table 3.8: Estimators for Restricted Additive Risk Frailty Model for Semi-Competing Risks, accounting for treatment, nodal status and TNM staging (data from Wee *et al.*, 2005).

	Relapse	Death
Treatment [a]	0.134	0.108
	(0.070)	(0.060)
Nodal status [b]	0.136	0.065
	(0.146)	(0.110)
TNM staging [c]	0.070	0.135
	(0.100)	(0.097)
Loglikelihood	−351.534	

Standard errors given in parentheses and * indicates significance at 5% level.
[a]CRT as reference group
[b]N0–2 as reference group
[c]Stage 3 as reference group

3.6 Discussion

In this chapter, we applied additive hazards to the conditional frailty model to account for covariates, where the frailty term measured the potential relationship between relapse and death times for each individual. The additive hazards model can provide a complementary analysis to the usual Cox model where the covariate effect is multiplicative. Simulations indicated that the method works reasonably well. Further discussion on the methodology is found in Chapter 5.

Application of the conditional additive risk frailty model on the NP01 dataset utilises more information than the original literature of Wee *et al.* (2005), which analysed the data as competing risks. In contrast, our proposed model accounts for the unique censoring relationship between relapse and death. Results from the model fitting indicate similar results to Wee *et al.* (2005) in that the addition of nodal status and TNM staging as covariates did not yield different estimates from the model that included treatment only. We also confirmed the findings that chemotherapy improves the relapse control rate in NPC.

Under the semi-competing risks setting, we obtained further insight on overall survival as we now examine death as two outcomes - with and without relapse of NPC. In the trial conducted by Wee *et al.* (2005), overall 2- and 3-year survival rates were found to be 85% v 78% and 80% v 65% for CRT and RT treatment groups respectively. Similar survival rates were also obtained under the restricted additive fraily model. However, when compared against the general model, the restricted model was not sufficient as our likelihood ratio test indicated.

For the general model, we obtained higher survival rates in our analysis for 2- and 3-year survival rates for the two treatment groups in the death without

relapse arm. On the other hand, the death with relapse arm had substantially lower 2- and 3-year survival rates as compared to Wee *et al.* (2005). This is also seen in the estimates obtained. One possible reason for the difference could be the unique relationship between death and relapse and by accounting for this relationship as we have done in our model, it reveals the distinct difference between the two death outcomes rather than examining all the deaths for overall survival, while examining death as a single outcome without accounting for relapse might cause the two effects to cancel each other out.

This dataset was analysed in Lim's (2010) thesis, which considered a parametric proportional hazards model with a Weibull baseline and Gamma shared frailty. Only treatment effect was accounted for and her model did not indicate significant effect of adjuvant chemotherapy, although the observations made from the estimates alone (regardless of significance) showed the same conclusion as the analysis done in this chapter.

Xu *et al.* (2010) also analysed the same dataset, but considered the proportional hazards model with shared frailty and nonparametric baseline hazards. Other differences in our analyses and that of Xu *et al.* include different covariates used in adjustment. This chapter adjusts for nodal status (N0–2 v N3) and TNM staging (Stage 3 v Stage 4), while the latter adjusts for tumour size and nodal status and accounts for them as categorical variables with 4 levels each. The observations on treatment effect are similar in both analyses — patients in the CRT group had significantly increased chances of disease-free survival and overall survival after relapse, but experienced decreased chances of overall survival without relapse, although the last effect was not significant. In addition, the estimate of the frailty parameter was found to be highly significant in their analysis, but was not in this chapter.

The estimated frailty parameter under the general model was found to be a small and non-significant value. A possible reason for this is that the general compartment model allows the hazards in the death arms to differ. The frailty in the restricted model captures the difference of hazards with and without relapse since we assume $\beta_2 = \beta_3$ and $\lambda_{02} = \lambda_{03}$. Thus intuitively, we can expect the frailty variance to be smaller than that in the reduced model. In addition, we have assumed the frailty to follow the Gamma distribution. In practice, this might not be true, resulting in an inaccurate estimate for the frailty parameter. Future work can examine model checking procedures for such assumptions on the frailty distribution.

Chapter 4

Extensions to the Additive-Multiplicative Model

4.1 Introduction

The focus of this thesis thus far has been on the additive risk model. We now extend the additive risk frailty model to the more general additive-multiplicative hazard models for the analysis of semi-competing risks data. For univariate survival data, the general class of additive-multiplicative model has been studied. The additive hazards and multiplicative hazards models postulate different relationships between the covariates and the hazard function and the choice between additive or multiplicative hazards can be an empirical decision or based on physical logic. The general class of additive-multiplicative models has the flexibility of allowing some covariate effects to be additive while letting others be multiplicative or allowing certain covariates to have both the additive and multiplicative effects.

A simple additive-multiplicative model was first analysed by Andersen and Værth (1989), which looked at relative and excess mortality in comparison to a

known population mortality, $\lambda^*(t)$. Their model was given as

$$\lambda(t) = \alpha(t) + \beta(t)\lambda^*(t),$$

where $\alpha(t)$ and $\beta(t)$ measured the excess and relative risk respectively, but did not account for covariate effects. The proportional excess hazards model was also examined by Sasieni (1996). Lin and Ying (1995) studied the class of models given by

$$\lambda(t|W, X) = g\{\alpha_0^T W(t)\} + \lambda_0(t)h\{\beta_0^T X(t)\},$$

where g and h are known link functions, W and X are possibly time-varying covariates and λ_0 is an unspecified baseline hazard under $g = 0$ and $h = 1$. It is easy to see that such a class of models encompasses both the additive model ($g(x) = x, h = 1$) and the Cox model ($g = 0, h(x) = e^x$). Under this model, all covariate effects are fixed and not time-varying.

Martinussen and Scheike (2002) and Scheike and Zhang (2002) suggested two different additive-multiplicative models with time-varying covariate effects. In the former, Martinussen and Scheike (2002) examined a model similar to Lin and Ying (1995) with g as the identity link and h as the exponential link, but they allowed α_0 to vary with time while keeping β_0 as time-invariant. Scheike and Zhang (2002) considered a variation from the above models and extended the Cox model by allowing the baseline hazard to depend on covariates through the additive Aalen model. Their model was given as

$$\lambda(t|W, X) = (\alpha_0(t)^T W(t))\lambda_0(t)h\{\beta_0^T X(t)\}.$$

In this chapter, we propose to extend the general additive-multiplicative model

to the semi-competing risks setting discussed in Chapter 3. Conditional on the frailty term, the hazard function is an additive-multiplicative function following that of Lin and Ying (1995) with $g(x) = x$ and $h(x) = e^x$. Similar to Chapter 3, we model the baseline hazard functions via cubic B-splines with fixed knots.

4.2 Proposed Additive-Multiplicative Model

4.2.1 Additive-Multiplicative Model

Using the same notation as in Section 3.2 for the failure times, we have the hazard functions for the i-th individual as:

$$\begin{aligned}
\lambda_1(t|T_{i2} \geq t, \gamma_i, \mathbf{W}_i, \mathbf{Z}_i) &= \gamma_i(\lambda_{01}(t)\exp(\boldsymbol{\alpha}_1^T\mathbf{W}_i(t)) + \boldsymbol{\beta}_1^T\mathbf{Z}_i(t)), \\
\lambda_2(t|T_{i1} \geq t, \gamma_i, \mathbf{W}_i, \mathbf{Z}_i) &= \gamma_i(\lambda_{02}(t)\exp(\boldsymbol{\alpha}_2^T\mathbf{W}_i(t)) + \boldsymbol{\beta}_2^T\mathbf{Z}_i(t)), \\
\lambda_3(t|T_{i1} \leq t, \gamma_i, \mathbf{W}_i, \mathbf{Z}_i) &= \gamma_i(\lambda_{03}(t)\exp(\boldsymbol{\alpha}_3^T\mathbf{W}_i(t)) + \boldsymbol{\beta}_3^T\mathbf{Z}_i(t)), \qquad (4.1)
\end{aligned}$$

where $\mathbf{W}$ and $\mathbf{Z}$ are covariate vectors with corresponding vectors of unknown regression parameters, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and as in Chapter 3, γ_i is the frailty term for each individual and assumed to be independent and identically distributed from the Gamma distribution with shape θ^{-1} and scale θ. Again, the Gamma distribution is assumed for mathematical convenience.

4.2.2 Estimation for Additive-Multiplicative Model

Similar to Chapter 3, we propose to use cubic B-splines to estimate the three baseline hazard functions. The observed loglikelihood can then be written as

$$\begin{aligned} l &= \sum_{i=1}^{n} \Bigg\{ \delta_{i1} \log \Big[\lambda_{01}(X_i) \exp(\boldsymbol{\alpha}_1^T \mathbf{W}_i) + \boldsymbol{\beta}_1^T \mathbf{Z}_i \Big] \\ &\quad + \delta_{i2}(1-\delta_{i1}) \log \Big[\lambda_{02}(Y_i) \exp(\boldsymbol{\alpha}_2^T \mathbf{W}_i) + \boldsymbol{\beta}_2^T \mathbf{Z}_i \Big] \\ &\quad + \delta_{i1}\delta_{i2} \log \Big[\lambda_{03}(Y_i) \exp(\boldsymbol{\alpha}_3^T W_i) + \boldsymbol{\beta}_3^T \mathbf{Z}_i \Big] + \delta_{i1}\delta_{i2} log(1+\theta) \\ &\quad - (\frac{1}{\theta} + \delta_{i1} + \delta_{i2}) \log \Bigg[1 + \theta \Bigg\{ \Lambda_{01}(X_i) \exp(\boldsymbol{\alpha}_1^T W_i) + (\boldsymbol{\beta}_1^T \mathbf{Z}_i) X_i \\ &\qquad + \Lambda_{02}(X_i) \exp(\boldsymbol{\alpha}_2^T W_i) + \Big(\boldsymbol{\beta}_2^T \mathbf{Z}_i \Big) X_i \\ &\qquad + (\Lambda_{03}(Y_i) - \Lambda_{03}(X_i)) \exp(\boldsymbol{\alpha}_3^T W_i) + (\boldsymbol{\beta}_3^T \mathbf{Z}_i)(Y_i - X_i) \Bigg\} \Bigg] \Bigg\}, \quad (4.2) \end{aligned}$$

where $\lambda_{0m}(t)$ is estimated by Equation (3.2) and $\Lambda_{0m}(t)$ $(m = 1, 2, 3)$ is estimated by Equation (3.4).

This loglikelihood function can be maximised over the finite dimensional space for the parameters $\psi = \{\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \boldsymbol{\alpha}_3^T, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T, \theta, c_{11}, \ldots, c_{1J}, c_{21}, \ldots, c_{2J}, c_{31}, \ldots, c_{3J}\}$.

4.3 Theoretical Properties

In this section we use the assumptions listed in Section 3.3. The proofs of consistency and asymptotic normality are similar to Chapter 3 and follow in outline to that of Murphy (1995) and Xue *et al.* (2004) and are not given here. Let ψ_0 be the true values of the parameters.

Theorem 4.3.1. (Consistency). *If the conditions 1–3 in Section 3.3 hold, then* $\hat{\psi} \xrightarrow{P} \psi_0$.

Theorem 4.3.2. (Asymptotic normality). *If the conditions 1–3 in Section 3.3 hold, then* $\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0, \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \hat{\theta} - \theta_0)$ *is asymptotically normal, where* $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)$ *and* $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)$.

4.4 Simulation Results for Extended Model

For simplicity, we consider the reduced model where $\lambda_2 = \lambda_3$, that is, $\lambda_{02} = \lambda_{03}$ and $\beta_2 = \beta_3$.

4.4.1 Data Generation of Semi-Competing Risks under Extended Model

The algorithm to generate survival times T_1 and T_2 and the observed data is as follows:

1. Generate T_1 and T_2 from distributions based on λ_1 and λ_2 respectively. That is, T_1 is generated from a distribution F_1, whose hazard function is λ_1. T_2 is generated in a similar fashion.

2. If $T_1 > T_2$, generate T* from distribution with hazard function λ_2 and set $T_2 = T_1 + T^*$.

3. Generate censoring time, C, from Uniform[0,a] for some $a > 0$.

4. If $T_2 < C$, set $\delta_2 = 1$ and $Y = \min(T_2, C)$.

5. If $T_1 < \min(T_2, C)$, set $\delta_1 = 1$ and $X = \min(T_1, Y)$.

The observed data for each sample of size n is then $\{X_i, \delta_{1i}, Y_i, \delta_{2i}, Z_i\}_{i=1}^n$.

4.4.2 Simulation Results for Additive-Multiplicative Hazards on Semi-Competing Risks

For the reduced model, we consider the case of a single discrete covariate Z generated from the Bernoulli(0.5) distribution for the additive component and a single continuous covariate W from the standard Normal distribution for the multiplicative component. Simulations were conducted for a sample size of 400, and at two degrees of censoring, 10% and 30%. We fix the baseline hazard to be from the exponential distribution with mean 1. Two sets of true values are considered for the regression coefficients of the respective hazards — $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0, 0, 1, 1)$ and $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (-0.5, 0, 0.5, 1)$; and we vary the true values of θ for the Gamma frailty distribution — 0.5, 0.95, 1.5.

We generate 1000 simulations for each combination of censoring, regression coefficients and θ. Tables 4.1, 4.2, 4.3 show the results and for each parameter combination give: (i) the Estimate (Est),the average of 1000 estimates; (ii) the Empirical Standard Error (EmpSE), the standard deviation of the 1000 estimators; (iii) the Model Standard Error (ModSE), the average of 1000 standard error estimators, using the information matrix.

The results show that the method works reasonably well, with small biases of about 0.1 or less. Empirical and estimated variances based on the information matrix were relatively similar. Also, larger values of coefficients result in larger biases and standard errors. However, they give smaller absolute biases per unit of standard error. The exception is $\hat{\theta}$, which has a mostly constant value of 0.5 of absolute bias per unit of standard error. Larger values of θ resulted in larger standard error estimates for all parameters.

Table 4.1: Estimators for Additive-Multiplicative Risk Frailty Model for Reduced Model of Semi-Competing Risks Data with single W from standard Normal and single Z from Bernoulli(0.5) and $\theta = 0.5$, with 10% and 30% censoring and varying α and β.

			$(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0, 0, 1, 1)$		
[0,a]	*Cens*		**Est**	**EmpSE**	**ModSE**
$[0, 6.6]$	10%	α_1	0.0022	0.137	0.127
		α_2	0.0049	0.152	0.152
		β_1	1.0997	0.346	0.327
		β_2	1.1000	0.279	0.272
		θ	0.5735	0.130	0.125
$[0, 1.7]$	30%	α_1	0.0050	0.139	0.136
		α_2	-0.0015	0.164	0.157
		β_1	1.1195	0.366	0.369
		β_2	1.1192	0.303	0.306
		θ	0.5948	0.176	0.182

			$(\alpha_1, \alpha_2, \beta_1, \beta_2) = (-0.5, 0, 0.5, 1)$		
[0,a]	*Cens*		**Est**	**EmpSE**	**ModSE**
$[0, 6.6]$	10%	α_1	-0.5137	0.125	0.126
		α_2	-0.0014	0.157	0.155
		β_1	0.5543	0.251	0.254
		β_2	1.1100	0.277	0.277
		θ	0.5779	0.133	0.133
$[0, 1.7]$	30%	α_1	-0.5152	0.136	0.135
		α_2	0.0006	0.147	0.157
		β_1	0.5566	0.275	0.284
		β_2	1.1073	0.305	0.308
		θ	0.5794	0.173	0.191

Table 4.2: Estimators for Additive-Multiplicative Risk Frailty Model for Reduced Model of Semi-Competing Risks Data with single W from standard Normal and single Z from Bernoulli(0.5) and $\theta = 0.95$, with 10% and 30% censoring and varying α and β.

			$(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0, 0, 1, 1)$		
[0,a]	*Cens*		**Est**	**EmpSE**	**ModSE**
$[0, 12]$	10%	α_1	-0.0015	0.172	0.158
		α_2	0.0026	0.204	0.191
		β_1	1.1135	0.374	0.367
		β_2	1.1134	0.321	0.317
		θ	1.0382	0.164	0.165
$[0, 2.4]$	30%	α_1	-0.0032	0.161	0.154
		α_2	-0.0054	0.168	0.179
		β_1	1.0984	0.387	0.387
		β_2	1.1017	0.339	0.334
		θ	1.0287	0.201	0.214

			$(\alpha_1, \alpha_2, \beta_1, \beta_2) = (-0.5, 0, 0.5, 1)$		
[0,a]	*Cens*		**Est**	**EmpSE**	**ModSE**
$[0, 12]$	10%	α_1	-0.5143	0.160	0.153
		α_2	0.0064	0.196	0.195
		β_1	0.5640	0.282	0.285
		β_2	1.1218	0.329	0.323
		θ	1.0440	0.173	0.173
$[0, 2.4]$	30%	α_1	-0.5201	0.167	0.153
		α_2	0.0007	0.201	0.189
		β_1	0.5852	0.298	0.311
		β_2	1.1376	0.345	0.346
		θ	1.0654	0.221	0.230

Table 4.3: Estimators for Additive-Multiplicative Risk Frailty Model for Reduced Model of Semi-Competing Risks Data with single W from standard Normal and single Z from Bernoulli(0.5) and $\theta = 1.5$, with 10% and 30% censoring and varying α and β.

			$(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0, 0, 1, 1)$		
[0,a]	*Cens*		**Est**	**EmpSE**	**ModSE**
$[0, 30]$	10%	α_1	0.0065	0.211	0.217
		α_2	0.0128	0.260	0.265
		β_1	1.1072	0.398	0.408
		β_2	1.1050	0.367	0.363
		θ	1.6045	0.209	0.217
$[0, 3.9]$	30%	α_1	0.0054	0.190	0.185
		α_2	0.0060	0.220	0.220
		β_1	1.1193	0.422	0.431
		β_2	1.1091	0.371	0.375
		θ	1.6100	0.254	0.259

			$(\alpha_1, \alpha_2, \beta_1, \beta_2) = (-0.5, 0, 0.5, 1)$		
[0,a]	*Cens*		**Est**	**EmpSE**	**ModSE**
$[0, 30]$	10%	α_1	-0.5120	0.204	0.214
		α_2	0.0009	0.266	0.283
		β_1	0.6033	0.307	0.324
		β_2	1.1485	0.380	0.374
		θ	1.6199	0.223	0.228
$[0, 3.9]$	30%	α_1	-0.5123	0.182	0.181
		α_2	0.0143	0.231	0.229
		β_1	0.5892	0.319	0.337
		β_2	1.1248	0.384	0.381
		θ	1.6427	0.275	0.276

4.5 Application to NP01 Dataset

We apply this general additive-multiplicative frailty model to the NP01 dataset described in Section 3.5. In the analysis of the previous chapter, the additive risk frailty model that analysed treatment effect while adjusting for nodal status and TNM staging showed little improvement over the model that analysed treatment effect alone. Hence, we apply the reduced model used in the simulations and adjust for the same covariates. We imposed a multiplicative effect on the treatment covariate and additive effects on nodal status and TNM staging. This is to compare our results obtained with those of Xu *et al.* (2010) and the thesis written by Lim (2010), where both analysed the treatment covariate as having a multiplicative effect. Table 4.4 shows the results of the modelling.

Table 4.4: Estimators for Additive-Multiplicative Risk Frailty Model for Semi-Competing Risks, accounting for treatment, nodal status and TNM staging (data from Wee *et al.*, 2005).

	Relapse	Death
Multiplicative component		
Treatment [a]	1.581 *	2.243 *
	(0.513)	(1.023)
Additive component		
Nodal status [b]	0.343 *	0.103 *
	(0.169)	(0.048)
TNM staging [c]	−0.190 *	0.063
	(0.022)	(0.084)

Standard errors given in parentheses and * indicates significance at 5% level.
[a]CRT as reference group
[b]N0–2 as reference group
[c]Stage 3 as reference group

The association between relapse and death was measured through the frailty

and the variance of the assumed Gamma distribution was estimated as 6.205 (SE=1.325), which is highly significant. This indicates a strong association between relapse and death.

After accounting for the association in the model, the covariate effects are now significant. Patients in the CRT group fared better than the RT group in both the relapse arm and death arm, with hazard ratios $e^{-1.581} = 0.206$ and $e^{-2.243} = 0.106$ respectively. This confirms the significant survival benefit of CRT treatment over the RT treatment. In addition, the adjusted variables also have a significant effect on survival, in terms of excess risk. Patients with higher nodal status have a higher risk of relapse and of death. As for TNM staging, the effect differs in the relapse and death arm. Patients with a higher TNM Stage have a lower risk of relapse, but higher risk of death. However, the effect of TNM staging on death is not significant. In contrast, the estimated effects of the adjusted variables were not significant in Chapter 3.

Figures 4.1 and 4.2 show the curves for disease-free survival (time to relapse) and overall survival (time to death). In contrast to the additive model where the effect on survival is cumulative and increases over time, the additive-multiplicative model has effects which do not necessarily increase over time, as can be seen in the graphs. Even after accounting for nodal status and TNM staging, patients in the CRT group are observed to have better survival chances with regard to both death and relapse. In all combinations of nodal status and TNM staging, the difference between the survival chances of the CRT and RT groups becomes constant after about 2 years. Thus, the effect of adding chemotherapy to radiotherapy is largely seen within the first 2 years of randomisation.

The comparison of survival rates also varied for patients in different nodal status

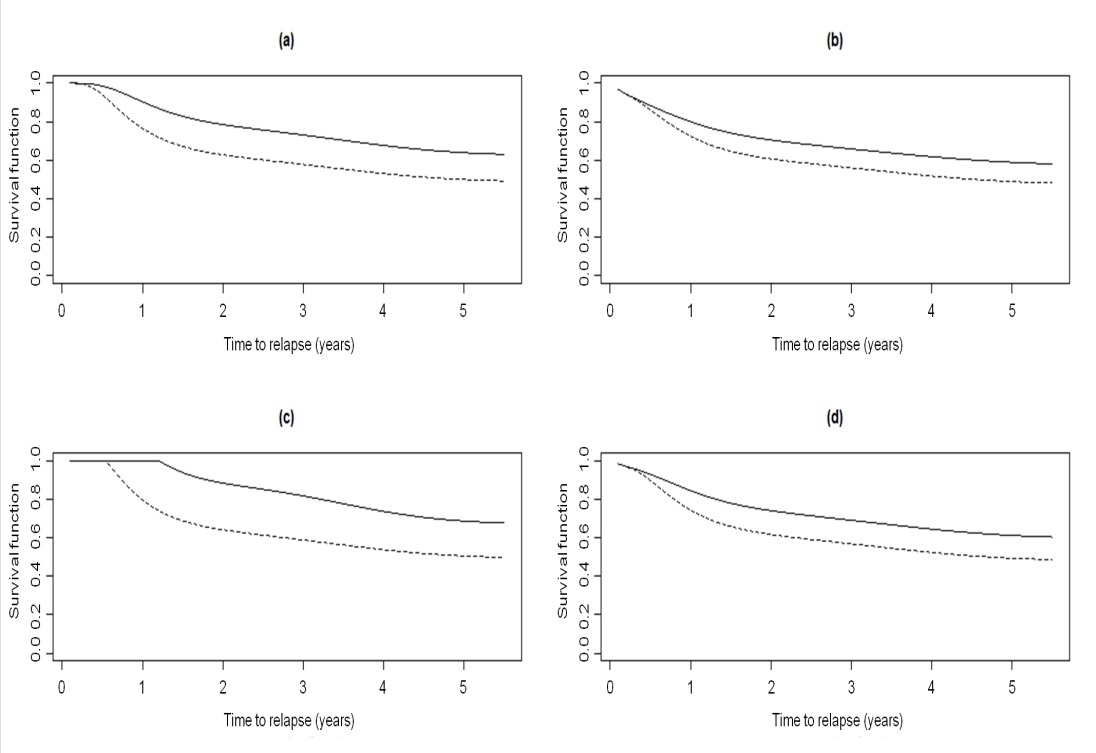

Figure 4.1: Survival functions comparing treatment effect on time to relapse for an individual under additive-multiplicative reduced model, stratifying for nodal status and TNM staging: (a) Nodal status N0–2, TNM Stage 2–3; (b) Nodal status N3, TNM Stage 2–3; (c) Nodal status N0–2, TNM Stage 4; (d) Nodal status N3; TNM Stage 4 (data from Wee *et al.*, 2005). — gives the survival function for patients receiving CRT; - - - gives the survival function for patients receiving RT.

groups. For patients with nodal status N0–N2, the 2-year disease-free survival rate was about 85% for the CRT group, while the rate were about 60% for the RT group. For patients with nodal status N3, the 2-year disease-free survival rate was about 73% for the CRT group compared to 60% for the RT group. The same analysis can be made for overall survival rates. For patients with nodal status N0–N2, the 2-year overall survival rate was about 90% for the CRT group and 78% for the RT group. Patients with nodal status N3 had estimated 2-year survival rates of 83% if they were in the CRT group and 60% if they were in the RT group.

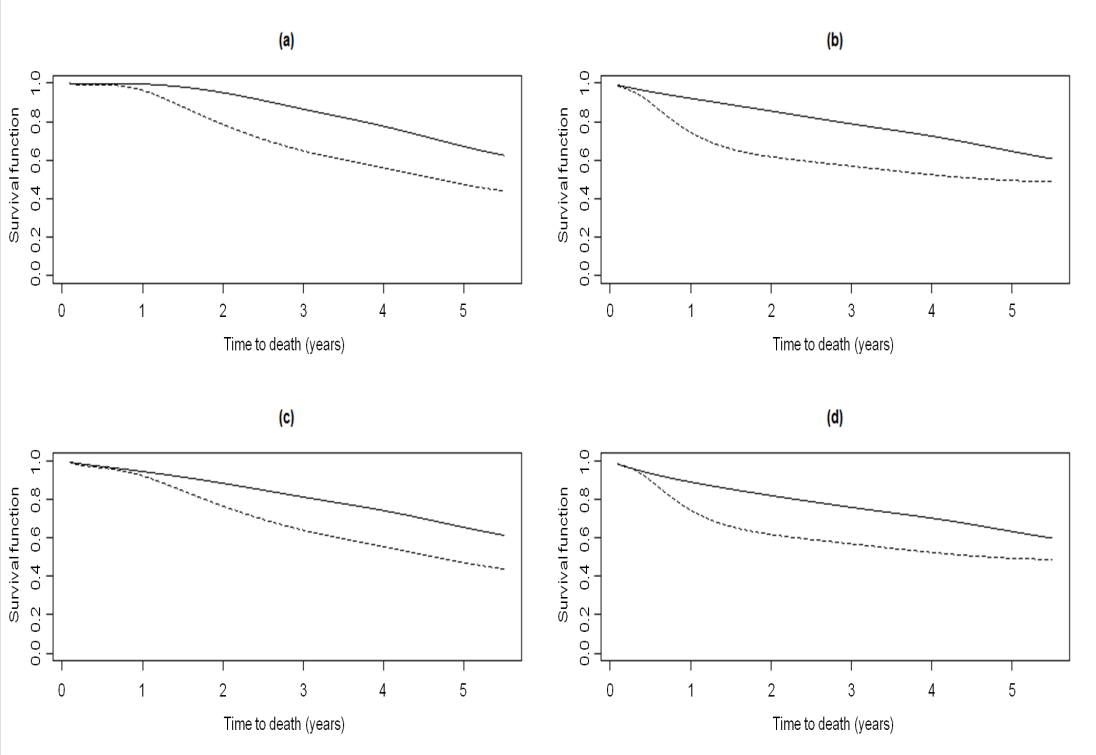

Figure 4.2: Survival functions comparing treatment effect on time to death for an individual under additive-multiplicative reduced model, stratifying for nodal status and TNM staging: (a) Nodal status N0–2, TNM Stage 2–3; (b) Nodal status N3, TNM Stage 2–3; (c) Nodal status N0–2, TNM Stage 4; (d) Nodal status N3; TNM Stage 4 (data from Wee *et al.*, 2005). — gives the survival function for patients receiving CRT; - - - gives the survival function for patients receiving RT.

From the graphs, we can also see that within the treatment groups, patients were more likely to suffer a relapse than death.

4.6 Discussion

In this chapter, we generalise the frailty model for semi-competing risks data to include both the additive and multiplicative components. Simulations on the reduced model show the method works well for moderate sample sizes. The estimation for the multiplicative component seems to fare better than the estimation for the additive coefficients. This could be due to the fact that there is no

constraint on the regression coefficients for the multiplicative component while the additive coefficients need to be constrained such that the hazards are non-negative hazards, ie., that $\lambda_m(t) \geq 0$ for $m = 1, 2, 3$. Application to the NPC dataset provides insight for the covariate effects, where the effects of treatment, nodal status and TNM staging are all significant. Under the reduced compartment model, adjuvant chemotherapy is observed to have significant protective effect, although the estimated effect appears to be implausibly large. This could be due to the restrictive assumptions of the model made in this chapter. In contrast, the estimate of the treatment effect under the restricted proportional hazards model proposed by Lim (2010) is smaller but not significant, while the estimates of Xu *et al.* (2010) under their restricted model produced a hazard ratio of about 0.345 when comparing CRT to RT and accounting for tumour size and nodal status. The estimate of the frailty parameter obtained in this chapter ($\hat{\theta} = 6.2$ with SE=1.3) is similar to the estimate of 7.0 obtained by Xu *et al.* (2010), indicating the strong relationship between relapse and death. Further work on the dataset could look into additive-multiplicative effects for different combinations of the covariates to see which gives the best fit. Models with less restrictive assumptions can also be explored to see if more plausible estimates of multiplicative treatment effect can be obtained. Model checking procedures can be developed to check the assumptions of the restricted model.

Chapter 5

Conclusion and Further Research

5.1 Conclusion

In biomedical studies, it is often of interest to evaluate drug efficacy in clinical trials in diseases. Although death is an important endpoint, it is also essential to study intermediate events like disease relapse, as they can provide additional information. This area of semi-competing risks has often been analysed based on a competing risks framework, due to the lack of an appropriate methodology. Methods for analysing such data have been proposed in the existing literature and were discussed in Chapter 1. These methods involve the use of copula models and assumptions on the existence of the marginal distribution for the time to the non-terminal event. In contrast, our proposed frailty model based on additive hazards does not make such assumptions and our analysis is focused only on the observable range of the data.

With the frailty model, covariate effects can be explicitly modelled and have a direct interpretation, as compared to the copula models. While the proportional

hazards model is the most commonly used regression model in univariate and multivariate survival analysis, we propose the additive risk model as a complementary measure. As mentioned in Chapter 1, it may make more biological sense in some cases to consider excess risk of a covariate instead of its relative risk. For example, the latent period for the risk of cancer following exposure to low doses of ionizing radiation can be better understood in terms of an additive risk model (Huffer and McKeague, 1991). Buckley (1984) shows that assuming a multiplicative model for analysis can have very misleading results when the data is from an additive model.

As such, Chapter 2 looked at the setting of competing risks and applied the additive risk model to the two widely-applied approaches for handling competing risks data — cause-specific hazards and subdistribution hazards respectively. We also proposed an additive risk model with time-varying coefficients for the subdistribution hazards model due to model limitations of the model with constant coefficients. Although there was a lack of fit in the subdistribution hazards model when applied to the prostate cancer dataset, this could be due to the specification of the time-varying form. If the time-varying form was correctly specified, then the proposed model would work well in practical settings, as demonstrated in the simulations. Other time-varying forms we could consider include $\alpha(t) = \beta/t$.

For the additive risk frailty model in Chapter 3, while we can allow the baseline hazard to be estimated nonparametrically, we propose the use of spline approximations to model the baseline hazard to reduce the complexity of the model. Splines have been widely used in modelling and are known for their flexibility. Our simulation studies indicate that the use of cubic B-splines to approximate the baseline hazard functions do not affect the estimation of the regression coef-

ficients and are flexible enough to estimate the continuous form of the unknown baselines.

When fitting the additive hazards model, it needs to be ensured that the overall hazard is non-negative. One way to account for this constraint would be to reparameterize $\beta^T Z$ to become $\exp(\beta^T Z)$, but this makes interpretation of the coefficients less straightforward. Hence, we choose to retain the original form of the additive risk model and work with contrained optimization. There are many packages in statistical software readily available to cope with constrained optimization. Also note that the contraints here are not on specific parameters but only on the overall hazard.

We applied our proposed model in Chapter 3 to analyse a real dataset of patients with endemic nasopharyngeal cancer. Results from the restricted model show similar observations as the original clinical paper (Wee *et al.*, 2005). Fitting of the additive risk frailty hazards to the general compartment model showed significant protective effect of CRT as compared to RT in the relapse and death after relapse arms. However, patients in the CRT group experienced an increase in risk of death without relapse as compared to those in the RT group. However, the frailty variance was found to be small and close to 0, indicating a lack of association between relapse and death.

In Chapter 4, we extended the model to the general additive-multiplicative frailty model to include the conditional proportional hazards and additive hazards as special cases. Simulations on the reduced model indicate reasonable performance for moderate sample sizes. Analysis on a real dataset of patients with endemic NPC using the restricted additive-multiplicative frailty model showed significant protective effect of CRT as compared to RT in both the

relapse and death outcomes. The estimated frailty variance also indicated a significant relationship between relapse and death.

5.2 Further Work

This thesis has attempted to shed some light on the modelling of semi-competing risks data through the use of shared frailties to model the dependence between the terminal and non-terminal events and an additive risk model to capture covariate effects. Further work in this area could include:

1. The consideration of other frailty distributions, such as the log-normal or positive stable distributions. The positive stable distribution has been shown to preserve proportionality of the hazards in the marginal distribution (Hougaard, 1986b). It would be worth investigating how these distributions behave in an additive risk setting and to analyse their properties.

2. Model-checking procedures to analyse goodness-of-fit of the restricted and general compartment models and also for model selection. Procedures could also be developed to check frailty assumptions.

3. Extension of spline approximations to tensor splines, to account for the bivariate nature of the terminal and non-terminal event times observed in the death after relapse arm of the compartment model. This would require a more in-depth study of the nature of splines, their uses and theoretical properties.

4. Extension of the proposed additive and additive-multiplicative models to accommodate data of other censoring structures, such as bivariate data, multivariate data and recurrent failure time data.

Bibliography

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. *Lecture Notes in Statistics*, 2:1 – 25.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. New York: Springer-Verlag.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10:1100 – 1120.

Andersen, P. K. and Værth, M. (1989). Simple parametric and nonparametric models for excess and relative mortality. *Biometrics*, 45:523 – 535.

Beyersmann, J., Latouche, A., Buchholz, A., and Schumacher, M. (2009). Simulating competing risks data in survival analysis. *Statistics in Medicine*, 28:956 – 971.

Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research, **1**, The Analysis of Case-Control Studies*. Lyon: IARC.

Buckley, J. D. (1984). Additive and multiplicative models for relative survival rates. *International Biometric Society*, 40:51 – 62.

Cai, J., Fan, J., Jiang, J., and Zhou, H. (2007). Partially linear hazard regression for multivariate survival data. *Journal of the American Statistical Association*, 102:538 – 551.

Cheng, S. C., Fine, J. P., and Wei, L. J. (1998). Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*, 54:219 – 228.

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

Clayton, D. G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society, Series A*, 148:82 – 117.

Couban, S., Simpson, D. R., Barnett, M. J., Bredeson, C., Hubesch, L., Kang, H. L., B, S. T., Walker, I. R., Browett, P., Messner, H. A., Panzarella, T., and Lipton, J. H. (2002). A randomized multicenter comparison of bone marrow and peripheral blood in recipients of matched sibling allogenic transplants for myeloid malignancies. *Blood*, 100:1525 – 1531.

de Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer-Verlag.

Delta Coordinating Committee (1996). Delta: a randomized double-blind controlled trial comparing combinations of zidovudine plus didanosine or zalcitablne with zidovudine alone in HIV-infected individuals. *Lancet*, 348:283 – 291.

Etezadi-Amoli, J. and Ciampi, A. (1987). Extended hazard regression for censored survival data with covariates: a spline approximation for the baseline hazard function. *Biometrics*, 43:191 – 192.

Fine, J. P. (1999). Analysing competing risks data with transformation models. *Journal of the Royal Statistical Society, Series B*, 61:817 – 830.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94:496 – 509.

Fine, J. P., Jiang, H., and Chappell, R. (2001). On semi-competing risks data. *Biometrika*, 88:907 – 919.

Gooley, T. A., Leisenring, W., Crowley, J., and Storer, B. E. (1999). Estimation of failure probabilities in the presence of competing risks. *Statistics in Medicine*, 18:695 – 706.

Gorfine, M., Zucker, D. M., and Hsu, L. (2006). Prospective survival analysis with a general semiparametric shared frailty model: A pseudo full likelihood approach. *Biometrika*, 93:735 – 741.

Gray, R. J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, 16:1141 – 1154.

Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, 87:942 – 951.

Gray, R. J. (1994). Spline-based tests in survival analysis. *Biometrics*, 50:640 – 652.

Green, S. B. and Byar, D. P. (1980). The choice of treatment for cacner patients based on covariate information: Application to prostate cancer. *Bulletin Cancer, Paris*, 67:477 – 488.

Grenander, U. (1981). *Abstract Inference*. New York: Wiley.

Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika*, 71:75 – 83.

Hougaard, P. (1986a). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73:387 – 396.

Hougaard, P. (1986b). A class of multivariate failure time distributions. *Biometrika*, 73:671 – 678.

Hsieh, J. and Wang, W. (2008). Regression analysis based on semi-competing risks data. *Journal of the Royal Statistical Society, Series B*, 70:3 – 20.

Huffer, F. W. and McKeague, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association*, 86:38 – 53.

Huster, W. J., Brookmeyer, R., and Self, S. G. (1989). Modelling paired survival data with covariates. *Biometrics*, 45:145 – 156.

Jiang, H., Fine, J. P., Korosok, R., and Chappell, R. (2005). Pseudo self-consistent estimation of a copula model with informative censoring. *Scandinavian Journal of Statistics*, 32:1 – 20.

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

Kay, R. (1986). Treatment effects in competing-risk analysis of prostate cancer data. *Biometrics*, 42:203 – 211.

Lakhal, L., Rivest, L. P., and Abdous, B. (2008). Estimating survival and association in a semi-competing risks model. *Biometrics*, 64:180 – 188.

Lam, K. F. and Kuk, A. (1997). A marginal likelihood approach to estimation in frailty models. *Journal of the American Statistical Association*, 92:985 – 990.

Larson, M. G. (1984). Covariate analysis of competing-risks data with log-linear models. *Biometrics*, 40:459 – 469.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 72:13 – 22.

Lim, G. H. (2010). Frailty modeling of semi-competing risks data. Master's thesis, National University of Singapore.

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, 81:61 – 71.

Lin, D. Y. and Ying, Z. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *The Annals of Statistic*, 23:1712 – 1734.

Martinussen, T. and Schieke, T. H. (2002). A flexible additive multiplicative hazard model. *Biometrika*, 89:283 – 298.

McKeague, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika*, 81:501 – 514.

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

Ng, S. K. and McLachlan, G. J. (2003). An EM-based semiparametric mixture model approach to the regression analysis of competing-risks data. *Statistics in Medicine*, 22:1097 – 1111.

Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sorensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, 19:25 – 43.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84:487 – 493.

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

Pepe, M. S. (1991). Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association*, 86:770 – 778.

Pintilie, M. (2006). *Competing Risks: A Practical Perspective*. New York: Wiley.

Pipper, C. B. and Martinussen, T. (2004). An estimating equation for parametric shared frailty models with marginal additive hazards. *Journal of the Royal Statistical Society, Series B*, 66:207 – 220.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag.

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

Rivest, L. P. and Wells, M. T. (2001). A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *Journal of Multivariate Analysis*, 79:138 – 155.

Rosenberg, P. S. (1995). Hazard function estimation using B-splines. *Biometrics*, 51:874 – 887.

Sasieni, P. D. (1996). Proportional excess hazards. *Biometrika*, 83:127 – 141.

Scheike, T. H. and Zhang, M. (2002). An additive-multiplicative Cox-Aalen regression model. *Scandinavian Journal of Statistics*, 29:75 – 88.

Schumaker, L. L. (2007). *Spline Functions: Basic Theory*. Cambridge University Press.

Shen, Y. and Cheng, S. C. (1999). Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics*, 55:1093 – 1100.

Sleeper, L. A. and Harrington, D. P. (1990). Regression splines in the cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association*, 85:941 – 949.

Sun, L., Liu, J., Sun, J., and Zhang, M. (2006). Modelling the subdistribution of a competing risk. *Statistica Sinica*, 16:1367 – 1385.

Tai, B. C., De Stavola, B. L., De Gruttola, V., Gebski, V., and Machin, D. (2008). First event or marginal estimation of cause-specific hazards for analysing correlated multivariate failure time data. *Statistics in Medicine*, 27:922 – 936.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

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 Tan, H. S., 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 3 and 4 nasopharyngeal cancer of the endemic variety. *Journal of Clinical Oncology*, 23:6730 – 6738.

Xu, J., Kalbfleisch, J. D., and Tai, B. C. (2010). Statistical analysis of illness-death processes and semi-competing risks data. *Biometrics*, 66:716 – 725.

Xue, H., Lam, K. F., and Li, G. (2004). Sieve maximum likelihood estimator for semiparametric regression models with current status data. *Journal of the American Statistical Association*, 99:346 – 356.

Yau, K. W. (2001). Multilevel models for survival analysis with random effects. *Biometrics*, 57:96 – 102.

Zheng, M. and Klein, J. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, 82:127 – 138.