Imputation and Dynamic Models in Semiparametric Survival Analysis

by

Xiaohong Liu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2011

Doctoral Committee:

Associate Professor Susan Murray, Co-chair
Professor Alexander Tsodikov, Co-chair
Professor Roderick J.A. Little
Professor Jeremy M. G. Taylor
Associate Professor Ji Zhu

To my parents.

ACKNOWLEDGMENTS

I sincerely thank my advisors, Dr. Susan Murray and Dr. Alex Tsodikov, for their guidance and support throughout my doctoral study. I am also grateful to Dr. Rod Little, Dr. Jeremy Taylor, and Dr. Ji Zhu for their valuable comments and inputs, which greatly improve three manuscripts developed based on this thesis.

The Department of Biostatistics is such an adorable place to stay. The faculties are always helpful and knowledgable. The five years here will forever be a cherished memory in my life.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

LIST OF APPENDICES

Appendix

CHAPTER I

Introduction

One of the most common research questions in survival analysis is how the observed covariates, Z, affect survival when data is right censored. Usually treated as a finite-dimensional vector in Euclidean space, Z can be baseline covariates that stay constant over time, or time changing factors. The presence of censoring complicates the estimation since care has to be taken with respect to unknown cumulative baseline hazard function H. Though regarded as absolutely continuous, H is either estimated non-parametrically via approximation by a infinite dimensional step function, or is imposed with a fully parametric distribution form. The latter approach is less frequently used in practice compared to the former, the semiparametric approach, since the inference is less robust due to the parametric assumptions made. The challenge of handling infinite dimensional nuisance parameters inspires various works in semiparametric models and that forms an important research area in survival analysis. Cox proportional hazard (PH) model is perhaps the most popular semiparametric model used in survival analysis, Cox (1972). Various extensionsof the PH model avoid the proportionality assumption or any distributional form of the hazard. For example, imputation methods, as in Faucett et al. (2002), Andersen et al. (2004), Hsu et al. (2006). These methods directly fill in missing data using multiple imputation methods; frailty models that represent heterogeneity of mortality, among other non linear

transformation models. In this thesis, we further explore the possibility of improving the existing semiparametric analysis methods in two areas. First we develop a multiple imputation method that better utilizes covariates to recover statistical information; Second area focuses on transformation models that generalizes frailty models when time dependent covariates are observed, or, in another word, when exposures are dynamic.

Currently most multiple imputation methods for censored survival data either ignore covariate information when imputing a likely event time, or place quite restrictive modeling assumptions on the survival distributions used for imputation. Several researchers modeled censored restricted lifetimes as a function of patient characteristics. Karrison (1987) used a generalized Cox model approach with piecewise constant baseline hazards, and made appropriate transformations to the restricted mean scale that indirectly linked covariates to the restricted means. Extensions of this approach, also centered around a Cox model, were given by Chen and Tsiatis (2001) and Zucker (1988). Andersen et al. (2004) link covariate effects with the mean restricted lifetimes by using pseudo-observations in lieu of the original outcomes and applying generalized estimating equations (GEE) by Liang and Zeger (1986). None of these authors suggested the restricted mean as a tool for imputation of censored lifetimes. In Chapter II, we propose a more robust multiple imputation approach that directly imputes restricted lifetimes over the study period based on a model of the mean restricted life as a linear function of covariates. This method has the advantages of retaining patient characteristics when making imputation choices, while avoiding parametric assumptions on the shapes of hazards or survival functions. Simulation results show that our method outperforms its closest competitor for modeling restricted mean lifetimes in terms of bias and efficiency in both independent censoring and dependent censoring scenarios. Survival estimates of restricted lifetime model parameters and marginal survival estimates regain much of the precision lost due to censoring. The

proposed method is also much less subject to dependent censoring bias captured by co-variates in the restricted mean model. This particular feature is observed in a full statistical analysis conducted in the context of the International Breast Cancer Study Group Ludwig Trial V using the proposed methodology. The work is also published as Liu et al. (2011).

Frailty models are popular when unmeasured predictors cause heterogeneity in mortality. Most often frailties represent an unobserved time-independent predictor known at baseline, for instance, Zeng and Lin (2007), Kosorok et al. (2004). The large sample properties of estimates are usually justified using empirical process theories, see Kosorok et al. (2004), Zeng and Lin (2007), Zeng and Lin (2010), or martingale-based arguments (Chen (2009) and Chen (2010)). Meanwhile there are some research propose that, when time changing factors are present, it is quite natural to believe a latent stochastic process is implied and it also affects survival. Typical examples are Gjessing et al. (2003) and Yashin and Manton (1997). Those works also discuss asymptotics of estimates. In Chapter III, we extend the modeling of frailty as a stochastic process in a general setting. We provide the estimation and inference framework for this type of model based on full likelihood approach, and build more rigorous theoretical justifications for large sample property using Martingale theory. We also discuss the asymptotic properties of estimates using a profile likelihood approach as in Murphy (1994), Murphy (1995), and Murphy and van der Vaart (2000). We show that in general no explicit covariance function form are available. Simulation studies are presented to illustrate the method, assuming frailty as a non-homogeneous Poisson process (NHPP), using a time-dependent covariate. We also present simulation studies that, assuming an oberved frailty process, Cox model gives biased estimate.

In Chapter IV, we develop a mechanistic modeling approach to explain dynamic effect when the disease (not necessarily cancer) has a latent development period before diagnosis.

This paper is motivated by an observational study in Tsodikov and Muller (1998), where the natural process of tumor initiation/promotion/progression is modulated by radiation. Radiation triggers lesions (initiation) that adds to the process of spontaneous formation of lesions in the host. Those lesions then compete as they develop into tumor during promotion. The same radiation may also kill some of the lesions. This non-linear interplay of these complex effects may result in improved survival of the subjects. Classical semi-parametric models with time dependent risks, including Cox model, fail to account for the dynamic irradiation effects on latent tumor process. We propose an incidence model based on the lesion initiation, promotion, competition and death modulated by the radiation to interpret the diversity of such complex exposure effects. The estimation procedure is based on the idea of iterative weighted algorithm introduced by Chen (2009). The advantage of this modeling approach is that we can interpret the time changing exposure effect as a stochastic process while retaining the power of rigorous statistical inference. As a result, we develop statistically and numerically efficient NPMLE techniques essentially as convenient as those for Cox models.

CHAPTER II

Multiple Imputation Based on Restricted Mean Models for Censored Data

2.1 Introduction

In survival analysis, estimation of expected life over a fixed time window is often of interest, either non-parametrically or as a function of covariates. In addition, it is common to desire estimates of survival probabilities within particular subgroups. For example, in the International Breast Cancer Study Group (IBCSG) Ludwig Trial V, investigators would like to estimate a long-duration treatment effect on patient lifetimes over the 9 year study period, adjusting for tumor size, estrogen receptor (ER) status, number of positive nodes and age. They also want to compare marginal survival curves for the two treatment groups. The presence of right censoring makes standard analysis methods for fully observed data inappropriate, although they would be much simpler to implement if available.

We propose that since restricted means are of interest and may be modeled already as part of a thorough analysis of the IBCSG study, that we take advantage of the restricted mean model structure to augment censored outcomes via Multiple Imputation (MI). The resulting final analyses (regression parameters, estimated restricted means and non-parametric quantities) are based on more standard analytical tools using multiply imputed datasets and are hypothesized to be more efficient since imputes better utilize co-

variate information.

Several researchers have given attention to modeling censored restricted lifetimes as a function of patient characteristics. Karrison (1987) used a generalized Cox model approach Cox (1972) with piecewise constant baseline hazards, and made appropriate transformations to the restricted mean scale that indirectly linked covariates to the restricted means. Extensions of this approach, also centered around a Cox model, were given by Chen and Tsiatis (2001) and Zucker (1988). Andersen et al. (2004) link covariate effects with the mean restricted lifetimes by using pseudo observations in lieu of the original outcomes and applying generalized linear models (GLM) Liang and Zeger (1986). None of these authors suggested the restricted mean as a tool for imputation of censored lifetimes.

Meanwhile, increasingly researchers have come to view censored data in the more traditional role of missing data, where multiple imputation is a popular strategy for appropriately addressing missing information in an analysis. For example, Faucett et al. (2002) multiply impute survival outcomes via joint modeling of a change-point model and a time-dependent Cox proportional hazards model. Taylor et al. (2002) develop non-parametric MI methods that reproduce Kaplan Meier Kaplan and Meier (1958) estimates when no covariate information is available. Hsu et al. (2006) use Cox models to more selectively build risk sets of individuals with similar hazards for multiple imputation, utilizing a non-parametric imputation procedure within this risk set. The advantages of imputation are longstanding, because many different analyses may be conducted using the multiply imputed datasets once they are obtained. An overview of several effective imputation strategies based upon observed data is given in Rubin (1987) and Little and Rubin (2002). Most existing MI methods either assume parametric models acting on (and linking) the hazards of interest or are non-parametric in nature.

Our goal in this research is to produce multiply imputed datasets that directly model

the missing outcomes of interest via a restricted mean structure. The resulting multiply imputed datasets incorporate individual information to gain efficiency in restricted mean model parameter estimation as well as in other analyses of interest, such as survival curve estimation and two sample testing. The rest of the manuscript is structured as follows: in Section 2.2, we describe the mean structure for restricted lifetimes given covariates. Section 2.3 introduces the restricted mean lifetime based MI algorithm with some technical details of implementation included in an appendix. In Section 2.4 we summarize several commonly used standard analyses applied to multiply imputed datasets. Section 2.5 presents finite sample simulation results. We return to the IBCSG study in Section 2.6 and report various analyses of interest. Discussion follows in Section 2.7.

This work is also published as Liu et al. (2011).

2.2 Structure for restricted mean lifetimes used in multiple imputation

Suppose lifetime, T, has survival function, $S_T(t)$, with mean life $E(T) = \int_0^\infty S_T(t)dt$. With right censored data, the data tends to support only estimated lifetimes restricted to the study period, or restricted means, $E\{min(\tau, T)\} = \int_0^\tau S_T(t)dt$, where τ is within the range of the observed data Irwin (1949).

To examine the restricted mean as function of covariates, Z, regression models have been developed. For instance, one approach is to assume a transformation model taking the form, $g(T) = \beta^T Z + \epsilon$, where $\beta = (\beta_{[0]}, \beta_{[1]}, ..., \beta_{[p]})$ is a $(p+1)$-dimensional vector, ϵ is residual vector with mean zero, and g is some link function (e.g., Buckley and James (1979), Dabrowska and Doksum (1979), Fine et al. (1998)). Fully parametric models can be implemented if the residual distribution is known. Andersen et al. (2004) use pseudo observations to model this mean structure, with the added advantage that few assumptions

are required on the distribution of ϵ for their model to hold.

In particular, for each individual, pseudo observation i ($i = 1, ..., n$) is defined as:

$$n \int_0^\tau \widehat{S}^{KM}(t)dt - (n-1) \int_0^\tau \widehat{S}^{KM(-i)}(t)dt,$$

where $\widehat{S}^{KM}(t)$ is the Kaplan Meier (KM) estimate for survival and $\widehat{S}^{KM(-i)}(t)$ is the KM estimate excluding patient i. These pseudo observations are comparable in expectation to the distribution of the original restricted failure times, similar to the jackknife. Hence this modeling approach addresses the censoring issue through transformation to uncensored values with identical restricted mean regression parameters. Andersen et al. recommend GLM analysis on log transformed pseudo observations using an identity link.

We assume a similar mean structure with the idea of imputing for $\log min(\tau, T)$ rather than using log transformed pseudo observations. That is:

(2.1) $$E[\log\{min(\tau, T)\}|Z] = \beta^T Z.$$

The log transformation of $min(\tau, T)$ continues to ensure that regression parameters apply to the real line rather than merely to positive values. Also, transforming $min(\tau, T)$ before model fitting seems to produce better estimates of the intercept than if a log link were applied, which is useful in the context of imputation. We suspect that this is the case due to Jensen's Inequality since $\log E\{min(\tau, T)|Z\} \geq E[\log\{min(\tau, T)\}|Z]$ and we impute on the scale of $\log\{min(\tau, T)\}$. Standard linear models can be used to fit (2.1) once multiply imputed datasets are created.

2.3 Multiple Imputation Algorithm

With the mean structure in (2.1), we have a natural way to fill in missing event times during the study window. We achieve this goal by developing restricted mean Multiple Imputation (MI) algorithm. The algorithm has two parts: first, we obtain desired parameter

estimates as in (2.1); Second, we append appropriate residuals to the estimated means to form an impute that better approximate variability of the original data.

The proposed algorithm is summarized in *Step 1-4*, with further details of implementation following.

Step 1: By fitting GLM model (2.1) treating censored data as failures, we obtain initial parameter estimates $\widehat{\beta}^{(0)}$. Next we use a pseudo EM algorithm described in the Appendix (A.1), to obtain a converged and improved $\widehat{\beta}$. This algorithm takes into account the current estimate of $\widehat{\beta}$, its variability and the observed censoring time C_i for each value requiring imputation.

Step 2: We form imputes for censored patients by adding error terms to the estimated means $\widehat{\beta}^T Z$ where $\widehat{\beta}$ is obtained in *Step 1*. For patients with similar $\widehat{\beta}^T Z$, observed residuals are sampled; the detailed sampling procedure is described in Appendix (A.2). Sampled error terms are required to yield an impute larger than the original censored value.

Step 3: Repeat *Step 2* until we have M imputes for each censored value.

Step 4: Combine analysis from M imputed datasets to get the final parameter estimates and the associated variances.

Next we describe the details of the algorithm. Suppose $T_1, ..., T_n$ come from a non negative random variable, T, with survival function, S_T, and $C_1, ..., C_n$ come from a random variable, C, that may or may not depend on covariates in model (2.1), but are otherwise independent of T. Let $X_i = min(T_i, C_i), i = 1, ..., n$ be the observed times to event. Let $Y = \log\{min(\tau, T)\}$ with τ a fixed positive constant.

In *Step 1*, we fit (2.1) to obtain initial values $\widehat{\beta}^{(0)}$ treating all observed data as failures. In practice we have not found the initial value to have much influence on final parameter estimates, although several alternative choices for obtaining $\widehat{\beta}^{(0)}$ were explored. Naturally the more censoring in the data, the further away $\widehat{\beta}^{(0)}$ is from the true β when censored

values are treated as failures. The next part of the algorithm is an iterative procedure to obtain $\widehat{\beta}^{(0)}, \widehat{\beta}^{(1)}, \widehat{\beta}^{(2)}, ..., \widehat{\beta}^{(k)}$ where the procedure is said to converge when $max_j(|\hat{\beta}_{[j]}^{(k)} - \hat{\beta}_{[j]}^{(k-1)}|) < a$ for some small tolerance a, with $\hat{\beta}_{[j]}^{(k)}$ being the jth element of the vector $\widehat{\beta}^{(k)}$. Steps of the iterative procedure are located in Appendix (A.1).

In *Step 2*, with the converged parameter $\widehat{\beta}$, $\widehat{E}(Y_i|C_i, Z_i)$ is calculated as $\widehat{\beta}^T Z_i$ for a censored patient i, then residual errors are added to $\widehat{E}(Y_i|C_i, Z_i)$ to reproduce appropriate variability. Details of sampling residuals are in Appendix (A.2). Our assumption for constructing residuals is that patients with similar $\widehat{E}(Y_i|Z_i)$ will have a similar distribution of residuals and can be used to create an appropriate pool for selection. Residuals that result in imputes of $\log\{min(\tau, T)\} < \log C_i$ are removed from further consideration. In the case of models with discrete covariates only, this residual pool reduces to patients with covariates identical to Z_i, and the impute ($\widehat{\beta}^T Z_i$ + residual) will essentially select one of the failed patients' death times from the pool as the imputed value. For continuous covariates, we sample residuals from patients whose $\widehat{E}(Y|Z)$ fall within some small $b-$margin of the censored patient's $\widehat{E}(Y_i|Z_i)$, so that the impute does not necessarily match any observed failure time from the original dataset, but is shifted higher or lower depending on $\widehat{\beta}^T Z_i$. In either case, when a patient from the residual pool is selected with event times $> \tau$, we use $\log \tau$ for the imputed value since a reasonable shift from $\widehat{\beta}^T Z_i$ is not available in this case.

Finally in *Step 4*, for each censored value, we sample M of those residuals and add them to $\widehat{E}(Y_i|Z_i)$ as described above, resulting in a total of M imputed datasets to be analyzed.

2.4 Analyse multiply imputed datasets

Since we fill in the missing outcomes for censored people, many research problems become complete data problems and we can apply standard procedures to analyze M

imputed datasets. In practice, $M = 10$ multiply imputed datasets are usually sufficient. Next we summarize some most commonly desired analyses.

2.4.1 Estimating regression parameters based on multiply imputed datasets

Using a standard GLM modeling approach, each of the M imputed dataset yields estimates $\widehat{\beta}_m^{MI} = (\hat{\beta}_{m,[0]}^{MI}, \hat{\beta}_{m,[1]}^{MI}, ..., \hat{\beta}_{m,[p]}^{MI})$, $m = 1, ..., M$ under model (2.1). The final estimates based on the multiply imputed datasets are $\widehat{\beta}^{MI} = \sum_{m=1}^{M} \widehat{\beta}_m^{MI}/M$. The associated variances are composed of within imputation variances W and between imputation variances B respectively, see Rubin (1987) and Little and Rubin (2002).

The variances become

$$Var(\widehat{\beta}^{MI}) = W + (1 + M^{-1})B.$$

where $W = \sum_{m=1}^{M} Var(\widehat{\beta}_m^{MI})/M$ and $B = \sum_{m=1}^{M} (\widehat{\beta}_m^{MI} - \widehat{\beta}^{MI})^2/(M - 1)$.

Similarly, covariances between the jth and kth elements of $\widehat{\beta}^{MI} = (\hat{\beta}_{[0]}^{MI}, \hat{\beta}_{[1]}^{MI}, ..., \hat{\beta}_{[p]}^{MI})$ are calculated as in Rubin (1987) and Little and Rubin (2002):

$$
\begin{aligned}
Cov(\hat{\beta}_{[j]}^{MI}, \hat{\beta}_{[k]}^{MI}) &= \sum_{m=1}^{M} Cov\{\hat{\beta}_{m,[j]}^{MI}, \hat{\beta}_{m,[k]}^{MI}\}/M \\
&\quad + (1 + M^{-1}) \sum_{m=1}^{M} [\{\hat{\beta}_{m,[j]}^{MI} - \hat{\beta}_{[j]}^{MI}\}\{\hat{\beta}_{m,[k]}^{MI} - \hat{\beta}_{[k]}^{MI}\}]/(M - 1),
\end{aligned}
$$

where $\hat{\beta}_{m,[j]}^{MI}$, $j = 0, 1, ..., p$, $m = 1, ..., M$ is the jth element of estimate $\widehat{\beta}_m^{MI}$ from mth dataset.

The hypothesis test for $\widehat{\beta}^{MI}$ and significance level is determined by the t distribution: $(\widehat{\beta}^{MI} - \beta)Var(\widehat{\beta}^{MI})^{-1/2} \sim t_\upsilon$, where $\upsilon = \{1 + (M + 1)^{-1}W/B\}^2 (M - 1)$ based on Satterthwaite approximation (Rubin (1987), Little and Rubin (2002)).

2.4.2 Survival curve estimates based on multiply imputed datasets

Using the M imputed datasets, we calculate KM survival estimates $\widehat{S}_1^{KM}(t), ..., \widehat{S}_M^{KM}(t)$, and obtain associated variances $\widehat{V}_1(t), ..., \widehat{V}_M(t)$ based on Greenwood's formula. For complete datasets, the KM survival estimates reduce to simple sample proportions of those alive at times t. The combined survival estimate $\widehat{S}^{MI}(t)$ and $\widehat{V}^{MI}(t)$ are calculated to be

$$\widehat{S}^{MI}(t) = M^{-1} \sum_{m=1}^{M} \widehat{S}_m^{KM}(t)$$

$$\widehat{V}^{MI}(t) = M^{-1} \sum_{m=1}^{M} \widehat{V}_m(t) + (1 + M^{-1}) \sum_{m=1}^{M} \{\widehat{S}_m^{KM}(t) - \widehat{S}^{MI}(t)\}/(M-1).$$

2.4.3 Log-rank test based on multiply imputed datasets

For the mth imputed dataset $(m = 1, ..., M)$, let $T_1^m < ... < T_L^m$ denote the ordered failure times , D_{ik}^m and Y_{ik}^m denote the number of failures and number at risk for group i at time T_k^m, $i = 1, 2, k = 1, ..., L$. Furthermore, let D_k^m and Y_k^m denote the corresponding values in whole sample, $E_{1k}^m = D_k^m Y_{1k}^m / Y_k^m$ be expected failures in group 1 and $V_{1k}^m = D_k^m Y_{1k}^m Y_{2k}^m (Y_k^m - D_k^m)/(Y_k^m)^2(Y_k^m - 1)$. The log-rank statistic for the mth imputed dataset is given by:

$$Q_m^{MI} = (\sum_{k=1}^{L} V_{1k}^m)^{-1/2} \sum_{k=1}^{L} (D_{1k}^m - E_{1k}^m).$$

The combined log-rank statistic and corresponding variance then become:

$$Q^{MI} = \sum_{m=1}^{M} Q_m^{MI}/M$$

$$V^{MI} = 1 + (1 + M^{-1}) \sum_{m=1}^{M} (Q_m^{MI} - Q^{MI})/(M-1),$$

since Q_m^{MI} is asymptotic standard normal for large samples.

The hypothesis test for treatment difference and significance level is determined by the t distribution: $Q^{MI}/\sqrt{V^{MI}} \sim t_v = \left\{1 + (M+1)^{-1}/\sum_{m=1}^{M}(Q_m^{MI} - Q^{MI})\right\}^2 (M-1)$ based on Satterthwaite approximation (Rubin (1987), Little and Rubin (2002)).

2.5 Simulation Study

In this section, we study finite sample properties of selected analyses based on restricted mean MI datasets, including restricted mean regression parameter estimation and marginal survival curve estimation.

GLM parameter estimates for the restricted mean model are produced using the following methods:

(a) model (2.1) with an identity link where $\log\{min(\tau, T)\}$ is multiply imputed for censored observations using the restricted mean MI approach,

(b) model (2.1) with an identity link applied to log transformed pseudo observations (PO) as in Andersen, Hansen and Klein Andersen et al. (2004),

(c) the model in (a) applied to the uncensored (fully observed) data.

We first study the independent censoring case. In each of 1000 simulations, we perform the following procedure with a sample of size $n = 100$ and τ fixed at 1.5:

Step 1: For covariates, we generate bivariate normal $(0, 1)$ pairs $(Z_{1i} = z_{1i}, Z_{2i} = z_{2i})$, $i = 1, ..., n$ with correlation 0.3. We then transform one of these into a Uniform$(0, 1)$ distributed covariate by applying the inverse transform method, that is, $U_i = P(Z_{1i} \leq z_{1i})$. The second normal is transformed into a Bernoulli(0.5) covariate, $B_i = I(Z_{2i} \geq 0)$.

Step 2: We obtain the outcome of interest, $min(\tau, T_i)$, $i = 1, ..., n$. Each failure time T_i is simulated from an exponential distribution with hazard rate, λ_i, that satisfies the mean structure (2.1) for a pre-specified $\beta = (\beta_0, \beta_1, \beta_2)$, B_i, U_i and τ. That is, for this simulation

setup,

$$
\begin{aligned}
E(Y_i|C_i, B_i, U_i) &= \int_{-\infty}^{+\infty} y \, dF_{Y_i}(y) \\
&= \int_{-\infty}^{\log \tau^-} y f_{T_i}(e^y) dy + P(Y_i \geq \log \tau) \times \log \tau \\
&= \int_{-\infty}^{\log \tau^-} y \lambda_i e^y e^{-\lambda_i e^y} dy + e^{-\lambda_i \tau} \times \log \tau.
\end{aligned}
$$

Therefore λ_i is a numerical solution to

$$
\int_{-\infty}^{\log \tau^-} y \lambda_i e^y e^{-\lambda_i e^y} dy + e^{-\lambda_i \tau} \times \log \tau = \beta_0 + \beta_1 \times B_i + \beta_2 \times U_i,
$$

in terms of β_0, β_1, β_2, B_i, U_i and τ. This step gives us an uncensored dataset for analyses using method (c).

Step 3: We generate independent censoring from an exponential distribution with rate chosen to yield approximately 30% censoring prior to τ. Censored data analyses are based on $min(\tau, X_i)$ where $X_i = min(T_i, C_i)$.

Step 4: Using the censored dataset generated from *Step 3*, we apply the MI algorithm described in Section 3 and obtain multiply imputed datasets. Then we estimate model (2.1) regression parameters as in Section 4.1, and survival percentages as in Section 4.2.

Step 5: Using the censored dataset generated from *Step 3*, we apply the PO approach to estimate model (2.1) regression parameters. We estimate survival percentages using KM method. These analyses will be compared to those in *Step 4*.

Step 6: Using the uncensored dataset generated from *Step 2*, we estimate model (2.1) regression parameters and survival percentages. These analyses represent the upper bound of available efficiency that is attainable for this setting.

The simulation for the dependent censoring case uses $n = 1000$ and $\tau = 2$. C_i is simulated using the exponential distribution $Exp(\lambda \times U_i)$, where λ is chosen to give approximately 30% censoring. Otherwise the procedure for simulation is similar to the

independent case.

Table 2.1 displays the simulation results in the independent censoring case, where the true values of β are $(-1, 0.5, 0.5)$. For each approach (a), (b) and (c), we present bias as $\widehat{\beta} - \beta$, and the corresponding average estimated standard errors (SE). We also calculate the empirical standard deviation (ESD) of the 1000 estimates and the proportion of simulations that cover the true values (empirical coverage probability, CP). To assess gains in efficiency, we give average 95% confidence interval (95% CI) widths over the 1000 simulations. The asymptotic relative efficiency (ARE) is defined as $Var(\widehat{\beta}_{PO})/Var(\widehat{\beta} \text{ of interest})$, where $Var(\widehat{\beta}_{PO})$, the variance of estimates using the PO approach, is used as the reference variance.

The results show that the MI approach and uncensored data analysis yield approximately unbiased estimates. Aside from the intercept term, the PO method also appears unbiased. Because of difficulty in estimating the intercept term, the PO approach tends to have a lower coverage rate for the true $E[\log\{min(\tau, T)\}]$ (83% for the PO approach as opposed to 93% for the MI method and 94% for the uncensored data case).

The 95% CI widths for regression parameters are around 12% narrower using the MI method compared to the PO approach. Furthermore, they are very close to the 95% CI widths based on uncensored data. The MI parameter estimates are $27-28\%$ more efficient in terms of the ARE than the PO approach. Hence by assuming the mean structure as in (2.1) and including minimal assumptions on the variance, we are able to recover much of the efficiency lost due to censoring.

Additional simulations under different parameter settings show similar patterns of results. For simplicity, we present one additional scenario where the true values of β are $(-1, 0, 0.5)$, $(-1, 0.5, 0)$ and $(-1, 0, 0)$ as in Table 2.2, 2.3 and 2.4 and we skip the bootstrap stages for all these scenarios. An interesting fact worth mentioning is that even

Table 2.1: Independent censoring: Comparison of estimates based on model (1) using method (a) restricted mean MI (MI) approach, method (b) pseudo observation (PO) approach, method (c) uncensored observations (Uncensored), and method (d) restricted mean MI (MIS) approach with a bootstrap procedure added, method ($n = 100$).

Parameter	Method	Bias[1]	SE[2]	ESD[3]	CP[4]	Width of CI[5]	ARE[6]
$\beta_0 = -1$	Uncensored	-0.005	0.286	0.314	91.7%	1.12	1.36
	PO	-0.193	0.333	0.383	89.1%	1.31	1.00
	MI	0.002	0.295	0.316	92.8%	1.15	1.27
	MIS	-0.012	0.262	0.297	91.1%	1.03	1.62
$\beta_1 = 0.5$	Uncensored	0.005	0.220	0.233	93.8%	0.86	1.35
	PO	0.055	0.256	0.273	92.9%	1.01	1.00
	MI	0.011	0.227	0.240	93.5%	0.89	1.27
	MIS	0.002	0.228	0.250	91.4%	0.90	1.26
$\beta_2 = 0.5$	Uncensored	0.005	0.437	0.456	94.4%	1.71	1.36
	PO	0.059	0.510	0.537	93.8%	2.00	1.00
	MI	-0.042	0.451	0.452	95.1%	1.77	1.28
	MIS	0.010	0.391	0.413	93.4%	1.77	1.28

1. Bias is the average of $\widehat{\beta} - \beta$ over the simulations. I.e., the average estimated parameter is the shown bias plus the true parameter in column 1.

2. SE is the average estimated standard errors over the simulations.

3. ESD is empirical standard deviation (ESD) of the 1000 estimates.

4. CP is the empirical coverage probability, i.e., the proportion of simulations that cover the true values.

5. Width of CI is the average 95% confidence interval (95% CI) widths over the 1000 simulations.

6. ARE is asymptotic relative efficiency, defined as $Var(\widehat{\beta}_{PO})/Var(\widehat{\beta})$, where $Var(\widehat{\beta}_{PO})$, the variance of estimates using the PO approach, is used as the reference variance.

Table 2.2: Independent censoring II: Comparison of estimates based on model (1) using method (a) restricted mean MI (MI) approach, method (b) pseudo observation (PO) approach, method (c) uncensored observations (Uncensored), method ($n = 100$).

Parameter	Method	Bias[1]	SE[2]	ESD[3]	CP[4]	Width of CI[5]	ARE[6]
$\beta_0 = -1$	Uncensored	0.010	0.254	0.281	91.7%	1.00	1.37
	PO	-0.176	0.298	0.339	88.5%	1.17	1.00
	MI	0.000	0.262	0.302	91.0%	1.03	1.29
$\beta_1 = 0$	Uncensored	-0.011	0.221	0.236	92.1%	0.87	1.37
	PO	0.041	0.259	0.284	92.0%	1.02	1.00
	MI	-0.002	0.228	0.249	91.6%	0.89	1.29
$\beta_2 = 0.5$	Uncensored	-0.001	0.380	0.387	94.3%	1.49	1.37
	PO	0.046	0.444	0.448	94.7%	1.74	1.00
	MI	0.000	0.390	0.419	93.3%	1.53	1.30

1. Bias is the average of $\widehat{\beta} - \beta$ over the simulations. I.e., the average estimated parameter is the shown bias plus the true parameter in column 1.

2. SE is the average estimated standard errors over the simulations.

3. ESD is empirical standard deviation (ESD) of the 1000 estimates.

4. CP is the empirical coverage probability, i.e., the proportion of simulations that cover the true values.

5. Width of CI is the average 95% confidence interval (95% CI) widths over the 1000 simulations.

6. ARE is asymptotic relative efficiency, defined as $Var(\widehat{\beta}_{PO})/Var(\widehat{\beta})$, where $Var(\widehat{\beta}_{PO})$, the variance of estimates using the PO approach, is used as the reference variance.

though the covariates, in the scenario where true values of β are $(-1, 0, 0)$, are not predictive of the outcome, efficiency gain is still observed. This is because the algorithm utilizes the covariates information regardless their predictability. In another word, the model itself will not be able to tell whether the covariates are predictive or not.

The results for dependent censoring case are presented in Table 2.5, where the true values of β are $(-1, 1, 0.5)$. The sample size used (1000) is comparable to the sample size in the IBCSG example in Section 6. Parameter estimates using the restricted mean MI

Table 2.3: Independent censoring III: Comparison of estimates based on model (1) using method (a) restricted mean MI (MI) approach, method (b) pseudo observation (PO) approach, method (c) uncensored observations (Uncensored), method ($n = 100$).

Parameter	Method	Bias[1]	SE[2]	ESD[3]	CP[4]	Width of CI[5]	ARE[6]
$\beta_0 = -1$	Uncensored	0.004	0.283	0.292	94.2%	1.11	1.37
	PO	-0.185	0.330	0.350	91.4%	1.29	1.00
	MI	-0.011	0.292	0.310	93.1%	1.14	1.28
$\beta_1 = 0.5$	Uncensored	-0.012	0.246	0.250	93.9%	0.96	1.36
	PO	0.034	0.287	0.301	93.7%	1.13	1.00
	MI	-0.002	0.254	0.266	93.4%	0.99	1.28
$\beta_2 = 0$	Uncensored	0.012	0.422	0.407	95.9%	1.66	1.36
	PO	0.012	0.493	0.472	96.1%	1.93	1.00
	MI	0.019	0.436	0.430	94.7%	1.71	1.28

1. Bias is the average of $\widehat{\beta} - \beta$ over the simulations. I.e., the average estimated parameter is the shown bias plus the true parameter in column 1.

2. SE is the average estimated standard errors over the simulations.

3. ESD is empirical standard deviation (ESD) of the 1000 estimates.

4. CP is the empirical coverage probability, i.e., the proportion of simulations that cover the true values.

5. Width of CI is the average 95% confidence interval (95% CI) widths over the 1000 simulations.

6. ARE is asymptotic relative efficiency, defined as $Var(\widehat{\beta}_{PO})/Var(\widehat{\beta})$, where $Var(\widehat{\beta}_{PO})$, the variance of estimates using the PO approach, is used as the reference variance.

Table 2.4: Independent censoring IV: Comparison of estimates based on model (1) using method (a) restricted mean MI (MI) approach, method (b) pseudo observation (PO) approach, method (c) uncensored observations (Uncensored), method ($n = 100$).

Parameter	Method	Bias[1]	SE[2]	ESD[3]	CP[4]	Width of CI[5]	ARE[6]
$\beta_0 = -1$	Uncensored	-0.019	0.281	0.275	94.3%	1.10	1.17
	PO	-0.216	0.330	0.333	92.2%	1.29	1.00
	MI	-0.093	0.280	0.285	93.3%	1.10	1.17
$\beta_1 = 0$	Uncensored	0.005	0.254	0.259	95.0%	1.00	1.17
	PO	0.012	0.298	0.312	94.3%	1.17	1.00
	MI	0.003	0.254	0.268	93.5%	0.99	1.17
$\beta_2 = 0$	Uncensored	0.004	0.503	0.493	94.8%	1.97	1.17
	PO	-0.005	0.591	0.580	95.4%	2.32	1.00
	MI	0.001	0.503	0.516	93.2%	1.97	1.18

1. Bias is the average of $\widehat{\beta} - \beta$ over the simulations. I.e., the average estimated parameter is the shown bias plus the true parameter in column 1.

2. SE is the average estimated standard errors over the simulations.

3. ESD is empirical standard deviation (ESD) of the 1000 estimates.

4. CP is the empirical coverage probability, i.e., the proportion of simulations that cover the true values.

5. Width of CI is the average 95% confidence interval (95% CI) widths over the 1000 simulations.

6. ARE is asymptotic relative efficiency, defined as $Var(\widehat{\beta}_{PO})/Var(\widehat{\beta})$, where $Var(\widehat{\beta}_{PO})$, the variance of estimates using the PO approach, is used as the reference variance.

Table 2.5: Dependent censoring: Comparison of estimates based on model (1) using method (a) restricted mean MI (MI) approach, method (b) pseudo observation (PO) approach, method (c) uncensored observations (Uncensored), method ($n = 1000$).

Parameter	Method	Bias[1]	SE[2]	ESD[3]	CP[4]	Width of CI[5]	ARE[6]
$\beta_0 = -1$	Uncensored	0.001	0.078	0.091	90.6%	0.31	1.49
	PO	-0.163	0.095	0.119	57.3%	0.37	1.00
	MI	0.002	0.082	0.099	90.0%	0.32	1.33
$\beta_1 = 1$	Uncensored	0.001	0.066	0.071	93.5%	0.26	1.49
	PO	0.131	0.081	0.093	63.7%	0.32	1.00
	MI	0.004	0.069	0.074	93.5%	0.27	1.38
$\beta_2 = 0.5$	Uncensored	-0.004	0.117	0.121	92.9%	0.46	1.49
	PO	-0.083	0.142	0.148	89.2%	0.56	1.00
	MI	-0.008	0.122	0.132	91.7%	0.48	1.36

1. Bias is the average of $\widehat{\beta} - \beta$ over the simulations. I.e., the average estimated parameter is the shown bias plus the true parameter in column 1.

2. SE is the average estimated standard errors over the simulations.

3. ESD is empirical standard deviation (ESD) of the 1000 estimates.

4. CP is the empirical coverage probability, i.e., the proportion of simulations that cover the true values.

5. Width of CI is the average 95% confidence interval (95% CI) widths over the 1000 simulations.

6. ARE is asymptotic relative efficiency, defined as $Var(\widehat{\beta}_{PO})/Var(\widehat{\beta})$, where $Var(\widehat{\beta}_{PO})$, the variance of estimates using the PO approach, is used as the reference variance.

approach are essentially unbiased, while the PO method was subject to bias as large as one standard deviation.

As mentioned in Section 1, various analyses can be conducted based on multiply imputed datasets. One example is to produce marginal survival estimates. Simulation results in the independent censoring case are shown in Table 2.6 at survival quantiles 60% through $40\% \approx S(\tau)$. Survival estimates using the MI method approximate true quantiles well, and as expected, the level of efficiency gain increases with increased censoring. Survival

quantiles 100% through 60%, where censoring was minimal (around 16%), showed only negligible differences in efficiency (less than 1%). For this reason, we did not present the results for this range.

Simulations conducted in the dependent censoring case gave unbiased results for the MI estimated quantiles whereas the KM method overestimated survival by approximately 3% after the 60th quantile. Although gains in efficiency for survival estimates were seen using MI method in the dependent censoring case, gains were not nearly as attractive as in the independent censoring case [data not shown].

We also conducted simulation using PO method as described in Andersen and Perme (2010) for both censoring scenario. In these settings, PO is created according to weighted KM estimates. We first generate Z^* as 0 if $U \in (0, 0.25]$, 1 if $U \in (0.25, 0.5]$, 2 if $U \in (0.5, 0.75]$ and 3 otherwise. In another word, Z^* is the original covariate U categorized into 4 bins with the same size. Then we create weighted KM estiamte $n^{-1} \sum_{g=1}^{G} n_g \widehat{S}_g^{KM}(t)$, where $\widehat{S}_g^{KM}(t)$ is a Kaplan-Meier estimate in subgroup g of patients with categorical covariate $Z^* = g$ and n_g is the number of patients in subgroup g ($g = 1, ..., 4$). The results using this approach and using restricted MI method are similar as shown in Table 2.1 and Table 2.5. The discussion of comparing the two methods in this context is in final section.

Some simulations were repeated using an additional bootstrap step to provide a further level of variability in the selection of imputed values. That is, each of the M imputed datasets was produced from a different bootstrap sample of the original observed data, which further varied the distribution of parameter estimates in Section 3 Step 1 as well as the observed residual distribution in Section 3 Step 2. The results are presented in Table 2.1 and Table 2.6. Although some authors, for example, Rubins and Schenker (1991), Heitjan and Little (1991), Taylor et al. (2002), have found improved coverage using this approach, coverage probabilities in the simulations did not appreciably change Table 2.6.

Table 2.6: Independent censoring: Comparison of survival estimates using restricted mean MI (MI) approach, restricted mean MI (MI) approach with bootstrap procedure (MIS), KM with censored observations (Censored), and KM with uncensored observations (Uncensored) ($n = 100$).

Quantile	Censoring %[1]	Method	Bias[2]	SE[3]	ARE[4]
60%	21.3%	Uncensored	0.006	0.049	1.18
		Censored	0.006	0.053	1.00
		MI	-0.005	0.052	1.01
		MIS	0.003	0.052	1.03
55%	23.7%	Uncensored	0.005	0.049	1.22
		Censored	0.005	0.055	1.00
		MI	-0.008	0.054	1.03
		MIS	0.002	0.053	1.05
50%	26.0%	Uncensored	0.007	0.050	1.27
		Censored	0.007	0.056	1.00
		MI	-0.007	0.054	1.06
		MIS	0.002	0.054	1.07
45%	28.1%	Uncensored	0.006	0.050	1.33
		Censored	0.006	0.057	1.00
		MI	-0.009	0.054	1.09
		MIS	-0.002	0.054	1.10
40%	30.1%	Uncensored	0.006	0.049	1.40
		Censored	0.004	0.058	1.00
		MI	-0.008	0.054	1.13
		MIS	0.001	0.054	1.14

1. Censoring % is the average censoring percentage up to corresponding quantile over the simulations.

2. Bias is the average of true survival percentages minus estimated survival percentages over the simulations.

3. SE is the average estimated standard errors over the simulations.

4. ARE is asymptotic relative efficiency, defined as $Var(\widehat{S}_{PO}^{KM})/Var(\widehat{S}^{KM})$, where $Var(\widehat{S}_{PO}^{KM})$, the variance of estimates using the PO approach, is used as reference variance.

A conservative recommendation would be to perform the analysis both with and without the bootstrap step, particularly for smaller sample sizes than the simulations shown here ($n < 100$).

2.6 IBCSG Ludwig Trial V Example

We now apply standard analyses to restricted mean-based multiply imputed datasets from the IBCSG Ludwig Trial V study. The data consist of 1229 patients, where 59 patients are still at risk at 108 months, 551 have died and 669 are censored prior to 108 months of followup. Observed covariates include long-duration (LD) or short-duration (SD) treatment assignment, estrogen-receptor ER status (positive vs. negative/unknown), tumor size (greater or less than 2cm), number of nodes (0-3, 4-9 or 10+), and age (in decades).

The primary interest of the study is to examine the treatment effect over the study period. Just as in a traditional medical journal results section, we first describe marginal treatment effects via plots and point estimates using the KM method and our restricted mean MI method. Then we test for treatment differences using the traditional log-rank test and the MI augmented log-rank test. Multivariate analysis results then assess adjusted treatment effects and other useful predictors.

As part of performing the MI procedure we estimate the restricted mean in terms of available data as follows:

$$
\begin{aligned}
E[\log\{min(108,T)\}] =\ & \beta_0 + \beta_1 \times I(\text{LD treatment}) + \beta_2 \times I(\text{ER positive}) \\
+\ & \beta_3 \times I(\text{Tumor} \geq 2cm) + \beta_4 \times I(\text{positive nodes 4-9}) \\
+\ & \beta_5 \times I(\text{positive nodes 10+}) + \beta_6 \times (\text{Age in decades}).
\end{aligned}
$$

In other words, using the methods described in Section (2.3), for censored patients we multiply imputed failure times incorporating information on treatment assignment, ER status, node group categories, tumor size and age.

Marginal survival curves based on the KM method and the MI method are shown in Figure (2.1), with confidence intervals at year marks shown in Table 2.7. The rightmost column of Table 2.7 summarizes the differences in point estimates of $\widehat{S}(t)$, according to method. Estimates between the two methods are similar during the first couple of years, but as censoring increases, so do differences between survival estimates across time. Particularly in the short-duration therapy, the difference in $\widehat{S}(8 \text{ years})$ approaches 6%. This pattern was similar both with and without a bootstrap step included in the analysis. When investigating possible reasons for differences in tail survival estimates, we identified dependent censoring captured by age (Hazard Ratio for censoring: 0.81 per decade age increase, 95% CI: (0.70, 0.93), p-value=0.004). That is, older patients who entered the trial were both less likely to be censored and had longer restricted lifetimes (to be discussed shortly in Table 2.8 as part of the multivariate analyses). The MI procedure accounts for this setting, giving lower survival estimates over time when compared to the KM method. The treatment differences are much larger once the dependent censoring bias related to age is accounted for. This is reflected in the much higher significance of the logrank analysis on the restricted mean MI datasets (without bootstrap: p=7×10^{-9}; with bootstrap: p=2×10^{-9}) compared to the traditional logrank test (p=0.0001). The marginal survival plots also indicate some non proportionality early in the study duration, perhaps arguing the merits of a multivariate model not dependent on proportional hazard shapes to hold.

The significant treatment difference is maintained once we adjust for other risk factors (full results in Table 2.8). The first 3 columns of Table 2.8 give the GLM model fit using the pseudo observation (PO) approach as in Andersen et al. (2004). The remaining
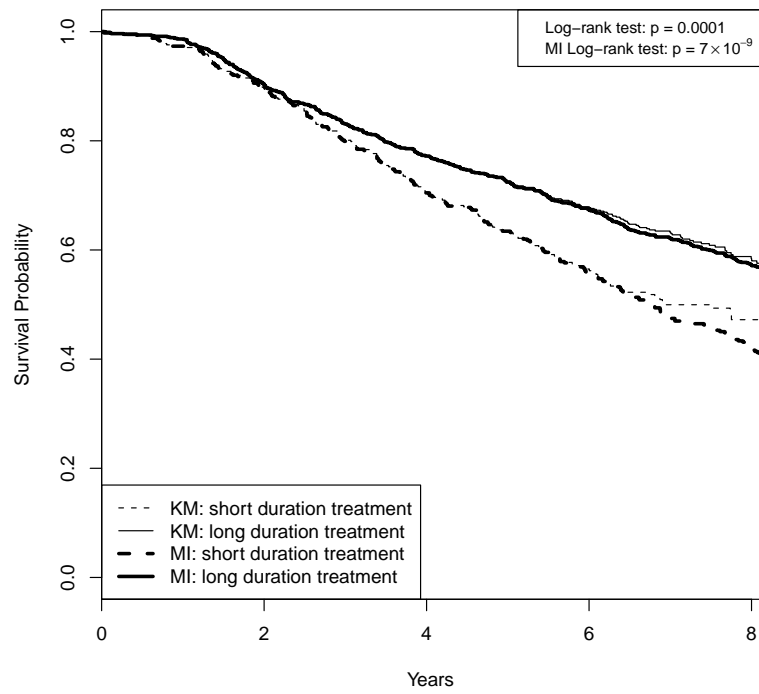
Figure 2.1: Treatment Specific Marginal Survival by Method.

columns give the analysis based on our restricted mean MI procedure. We report $e^{\widehat{\beta}}$, where $\widehat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6)$ with associated 95% CIs and p-values. The parameters $e^{\widehat{\beta}}$ are interpreted as multiplicative effects, so that parameter estimates higher than one give longer estimated restricted lifetimes and estimates smaller than one give shorter estimated restricted lifetimes. For example, using the MI approach, long-duration treatment tends to prolong restricted lifetime by a factor of $e^{\hat{\beta}_1} = 1.14$ (95% CI: 1.06, 1.23), compared to the restricted lifetime on the SD arm, adjusted for other factors. Across covariate effects, we observe similar or slightly narrower 95% CI widths and 1% efficiency gains for the MI method versus the PO approach. These minor efficiency gains in the variability of the parameter estimates were also seen when including a bootstrap step. The multivariate analysis confirms the role of age as a factor causing some dependent censoring since age is both related to the event time as well as the censoring time as noted above. The significance of age straddles the p-value=0.05 depending on methodology used. According to all methods used, older age is associated with longer restricted lifetimes. The MI method without a bootstrap step indicates that for each decade increase in age, the estimated restricted lifetime increases by 4.1% $\{e^{\hat{\beta}_6^{MI}} = 1.041, 95\%$ CI: (1.002, 1.082), p-value=0.040 $\}$. This seems to be capturing the known risk of more aggressive breast cancer tumors diagnosed in younger patients. Incorporating a bootstrap step into the MI procedure gave corresponding results for age as $\{1.036, 95\%$ CI: (0.996, 1.076), p-value=0.076$\}$, while the PO method gave $\{e^{\hat{\beta}_6^{PO}} = 1.038, 95\%$ CI: (0.998, 1.079), p-value=0.060$\}$.

The MI procedure seems to be accounting for dependent censoring through slightly better estimation of the intercept, which affects estimation of restricted lifetimes, and therefore the values imputed in the MI algorithm. For instance, we may estimate restricted lifetimes for a typical patient in the SD treatment arm, using average patient profile values for other risk factors, i.e., $(58.3)(0.79)^{0.27}(0.55)^{0.16}(0.88)^{0.77}(1.23)^{0.54}(1.04)^{5.00} = 61.6$

months for the MI method (95% CI: 57.9, 65.5), and similarly 62.1 months for the PO method (95% CI: 58.3, 66.0). That is, a 50 year old patient with a 27% chance of having 4-9 positive nodes, a 16% of having 10+ positive nodes, a 54% chance of being ER positive, and a 77% chance of having tumor greater than 2cm is expected to live 61.6 months out of a possible 108 months on study based on the MI method. The PO approach estimates a slightly longer restricted mean for this type of patient, likely connected to dependent censoring biases associated with age. Since younger (sicker) patients are being censored more often, the PO approach seems to be slightly overestimating the expected number of months lived during the 108 month duration.

2.7   Discussion

Using the restricted mean formulation, the shapes of the survival curves in relation to one another are not specified. Our method merely requires the mean structure to be correctly specified, i.e., the area under the survival curve to τ follows (2.1). It is also possible, of course, to recover restricted mean estimates from a Cox modeling framework using $\hat{S}(t|Z) = \hat{S}_0(t)^{e^{\hat{\beta}^T Z}}$, and in cases where hazards are truly proportional these estimates should be fully efficient. In two sample testing literature when hazards are not proportional, Pepe and Fleming (1998) indicated a substantial improvement in detecting treatment effects using differences in restricted means as opposed to the logrank test (score test for Cox model). Previous authors advocating restricted mean models have not clearly laid out inference performance issues in relation to hazard based models beyond the intuition gleaned from the two sample testing setting.

Andersen et al. (2004) use pseudo observations to create a modified dataset and apply a similar mean structure as (2.1) for analysis. In studying the PO method and how it might be modified to provide imputes larger than the observed censoring times, we

discovered a potential loss of statistical information available from Z. We considered imputes for C_i that add C_i plus a conditional pseudo observation created from the patients at risk at C_i. Suppose $\widehat{S}^{KM(-i)}(t|T \geq C_i)$ is the KM estimate with the person censored at C_i left out among those otherwise at risk at C_i. Then a conditional pseudo observation defined as $n \int_{C_i}^{\tau} \widehat{S}^{KM}(t|T \geq C_i)dt - (n-1) \int_{C_i}^{\tau} \widehat{S}^{KM(-i)}(t|T \geq C_i)dt$ would reduce to $\int_{C_i}^{\tau} \widehat{S}^{KM}(t|T \geq C_i)dt$. The non-parametric estimate $\widehat{S}^{KM}(t|T \geq C_i)$ does not fully utilize covariate information from Z. Pseudo observation calculation for this special case may indicate why our restricted mean MI approach outperforms the traditional PO method in simulation with respect to efficiency. Looking at this special case of pseudo observation creation may also suggest why dependent censoring might influence the PO method in terms of bias. That is, a conditional pseudo observation $\widehat{S}^{KM}(t|T \geq C_i)$ may still be biased if censoring depends on Z, and the traditional PO method seems also subject to this same source of bias.

More recently, Andersen and Perme (2010) suggested the use of pseudo observations based upon weighted Kaplan-Meier estimates (Murray and Tsiatis (1996)), $n^{-1} \sum_{g=1}^{G} n_g \widehat{S}_g^{KM}(t)$, where $\widehat{S}_g^{KM}(t)$ is a Kaplan-Meier estimate in subgroup g of patients with categorical covariate $Z = g$ and n_g is the number of patients in subgroup g ($g = 1, ..., G$). When only a single categorical covariate is associated with the survival and censoring distributions, corresponding calculations of conditional pseudovalues created from those at risk at C_i reduce to $\int_{C_i}^{\tau} \widehat{S}_g^{KM}(t|T \geq C_i)dt$, which is a maximum likelihood estimate of the restricted mean life for someone from group g surviving past C_i. Since all available covariate information is utilized in this scenario, we suspect that the pseudo observation approach will be fully efficient and unbiased for estimation of parameters in (1). Simulations using categorical covariates yield very similar results for parameter estimates based on weighted Kaplan-Meier based pseudo observations and restricted mean MI method, see Table 2.1.

The approach of creating pseudo observations with Kaplan-Meier estimates averaged across categorical covariate strata may be impractical in regression settings with many covariates. As covariate strata become more finely partitioned, technical difficulties of estimating survival consistently in the tails of the distribution arise since each stratum specific Kaplan-Meier curve is only guaranteed consistency in the range of the observed outcomes for that stratum. When only a single continuous covariate is related to survival and censoring distributions, averaging Kaplan-Meier estimates across categorical strata of the continuous covariate may still be viable since simulations by Murray and Tsiatis demonstrated that most efficiency gain and bias correction may be captured using roughly 3 to 5 strata.

With a single categorical covariate our restricted mean MI approach reduces to that of Hsu et al. (2006) since risk set groupings based on either similar restricted means, as in our work, or similar hazards, as in Hsu et al.'s work, will result in groups with the same categorical covariate to impute from. As indicated in Hsu et al., this special case produces marginal survival estimates similar in expectation to the weighted Kaplan-Meier estimate described by Murray and Tsiatis as well as the survival estimates proposed by Malani (1995). Similarly, rank-based tests based on MI analyses with categorical covariates would be expected to perform similarly to those proposed by Mackenzie and Abrahamowicz (2005).

To our knowledge, this is the first instance when the restricted mean lifetime has been used to impute censored survival data based on risk factors, increasing efficiency. Efficiency gains when using covariate information in marginal survival curve estimation have been seen in many other contexts by authors including Finkelstein and Schoenfeld (1994),Gray (1994) and Robins and Rotnitzky (1992) . The MI method retains essential characteristics of the observed data and approximates the original distribution well. Final

parameter estimates have good operating characteristics and have improved finite sample properties. Particularly appealing to clinicians is that the interpretation of the parameter estimates apply directly to days, months or years of life saved for different risk profiles. In addition, our method preserves the traditional benefits of MI such as transparency of variance calculation and availability of standard statistical software to analyze the augmented datasets.

Table 2.7: IBCSG Ludwig Trial V results I: Survival estimates using Kaplan Meier (KM) approach and the restricted mean multiple imputation (MI) approach.

Treatment	T	Censoring%	$\widehat{S}^{(KM)}(t)$ (95% CI)[1]	$\widehat{S}^{(MI)}(t)$ (95% CI)[1]	$\Delta^{[2]}$
SD					
	1 year	0.0%	0.97 (0.96,0.99)	0.97 (0.96,0.99)	0.000
	2 year	0.2%	0.90 (0.87,0.93)	0.89 (0.87,0.92)	-0.001
	3 year	0.5%	0.80 (0.76,0.84)	0.80 (0.76,0.84)	-0.001
	4 year	0.7%	0.71 (0.66,0.75)	0.71 (0.66,0.75)	-0.001
	5 year	2.7%	0.63 (0.59,0.68)	0.63 (0.59,0.68)	-0.001
	6 year	14.8%	0.57 (0.52,0.61)	0.55 (0.51,0.60)	-0.010
	7 year	28.8%	0.50 (0.45,0.55)	0.46 (0.41,0.51)	-0.036
	8 year	38.5%	0.47 (0.42,0.53)	0.41 (0.31,0.41)	-0.063
LD					
	1 year	0.0%	0.99 (0.98,0.99)	0.99 (0.98,0.99)	0.000
	2 year	0.1%	0.90 (0.88,0.92)	0.90 (0.88,0.92)	0.000
	3 year	0.3%	0.83 (0.80,0.86)	0.83 (0.81,0.86)	0.000
	4 year	0.8%	0.77 (0.74,0.80)	0.77 (0.74,0.80)	0.000
	5 year	2.8%	0.73 (0.70,0.76)	0.73 (0.70,0.76)	-0.001
	6 year	14.2%	0.68 (0.65,0.71)	0.68 (0.65,0.71)	-0.003
	7 year	31.7%	0.63 (0.60,0.67)	0.62 (0.60,0.66)	-0.008
	8 year	45.6%	0.58 (0.54,0.62)	0.57 (0.54,0.61)	-0.008

1. Statistical significance is defined when 95% CI does not cover the true values with Type I error rate 5%.

2. Δ is defined as $\widehat{S}^{(KM)}(t) - \widehat{S}^{(MI)}(t)$, at different t.

Table 2.8: IBCSG Ludwig Trial V results II: estimated restricted mean model parameters using a pseudo observation (PO) approach and the restricted mean multiple imputation (MI) approach.

Parameter	Mean	PO method Estimate[1]	95% CI[2]	p-value	MI method Estimate[1]	95% CI[2]	p-value	ARE[3]
Intercept	–	59.3	47.7, 73.7	<0.0001	58.3	46.9, 72.3	<0.0001	1.01
long-duration arm(LD)	0.66	1.12	1.04, 1.21	0.00300	1.14	1.06, 1.23	<0.0001	1.01
Positive nodes (4-9)	0.27	0.80	0.73, 0.87	<0.0001	0.79	0.72, 0.85	<0.0001	1.01
Positive nodes (10+)	0.16	0.56	0.50, 0.62	<0.0001	0.55	0.50, 0.61	<0.0001	1.01
Tumor($\geq 2cm$)	0.77	0.88	0.81, 0.96	0.00400	0.88	0.81, 0.96	0.00400	1.01
ER positive	0.54	1.23	1.14, 1.32	<0.0001	1.23	1.14, 1.32	<0.0001	1.01
Age (in decades)	5.00	1.04	0.99, 1.08	0.06000	1.04	1.00, 1.08	0.04000	1.01

1. Parameter estimates from mean structure (2.1) are presented on the scale of e^{β} so that estimates higher than one give longer mean restricted lifetimes and estimates smaller than one give shorter mean restricted lifetimes, i.e., multiplicative effects.

2. Statistical significance is defined when 95% CI does not cover the true values with Type I error rate 5%.

3. ARE is asymptotic relative efficiency, defined as $Var(\hat{\beta}_{PO})/Var(\hat{\beta}^{MI})$, where $Var(\hat{\beta}^{PO})$, the variance of estimates using the PO approach, is used as reference variance.

CHAPTER III

Stochastic Process Frailty Model

3.1   Introduction

Frailty models are commonly used when population heterogeneity is not fully explained by individuals' observed risk factors. The idea of using a non-negative random variable, U, to represent additional variability in the population was first introduced by Vaupel et al. (1979). Frailties are often incorporated into variations of the Cox proportional hazard (PH) model Cox (1972); In the case of Vaupel et al., the survival function becomes $S(t|z) = E(e^{-Ue^{\beta^T z} \int_0^t dH_s})$, where H_s is a baseline cumulative hazard function at time s, z is an observed time-independent covariate vector, and β is a parameter vector associated with the influence of z on survival. U is typically assumed to follow a non-negative parametric distribution, such as gamma {Klein (1992), Nielson et al. (1992) and Hougaard (1996), compound Poisson family {Aalen (1992)}, among other distributions Hougaard (2000).

Several authors extend frailty models by allowing z to be a function of time t. For example, Kosorok et al. (2004) considered the following survival model:

$$(3.1) \qquad\qquad S(t|z[0,t]) = E(e^{-U \int_0^t e^{\beta^T z_s} dH_s}),$$

where $z[0,t]$ represents the history of z_s, $s \in [0,t]$. Here U remains a time-independent frailty unrelated to $z[0,t]$. As shown in Appendix (B.1), the corresponding hazard function

33

becomes

$$
\begin{aligned}
\lambda(t|z[0,t]) &= \frac{E\{U\exp(-U\int_0^t e^{\beta^T z_s}dH_s)\}}{E\{\exp(-U\int_0^t e^{\beta^T z_s}dH_s)\}}e^{\beta^T z_t}dH_t/dt \\
&= E(U|T \geq t, z[0,t])e^{\beta^T z_t}dH_t/dt,
\end{aligned}
$$

where T is the failure time distribution.

Zeng and Lin (2007) considered an extended model in the same spirit where time accelerates according to a function of the observed time-dependent covariate, z_s for $s \in [0,t]$. In the context of univariate frailty models, the work of Zeng and Lin (2007) reduces to (3.1) extended by an additional power transformation, and where the distribution of U has no unknown parameters. The frailty, U, is characterized by the Laplace transform, $\mathcal{L}(x) = E(e^{-Ux})$. Their transformation function $G(x) = -\log E(e^{-Ux})$ acts on the cumulative hazard. That is, suppose $\tilde{z}$ is the subset of observed time independent covariates in z_s. The cumulative hazard becomes $\Lambda(t|z[0,t]) = G(x)$, where $x = \left(\int_0^t e^{\beta^T z_s}dH_s\right)^{exp(\gamma^T \tilde{z})}$.

Various authors have contributed to the literature on estimation and inference for models with time-independent frailties. Murphy (1994) and Murphy (1995) provide asymptotic theory in the case U has a gamma distribution and covariates are recorded at baseline. For general U, Tsodikov (2003) proposes a Quasi-EM procedure for estimation of (3.1). Kosorok et al. (2004), Zeng and Lin (2007) and Zeng and Lin (2010) study properties of estimated hazards and model parameters for their respective models using empirical processes. Bagdonnavicius and Nikulin (1999) and Martinussen and Scheike (2006) use a modified partial likelihood approach to argue normality of estimates in (1). Chen (2009) and Chen (2010) outline a weighted Breslow-type estimation approach for (3.1), basing associated asymptotic results on martingale theory.

In the majority of the existing literature, U is treated as time-independent. However, with observed z_t being dynamic, there is no reason to assume that the unobserved explana-

tory variable U is not. We will use the notation U_t to represent an unobserved stochastic process that changes over time $t \in [0, \tau]$. The history of U_s for $s \in [0, t]$ will hereafter be denoted as $U[0, t]$. For instance, Tsodikov and Muller (1998) use a non-homogeneous Poisson process (NHPP) to incorporate a frailty, U_t, that captures induction of cancerous lesions in mice by a time changing radiation exposure. Yashin and Manton (1997) propose a Gaussian frailty process for U_t. Gjessing et al. (2003) model U_t as a Levy process. Yue and Chan (1997) extend stochastic frailty modeling to the bivariate survival setting.

In this work, we consider a general process U_t and an non parametric maximum likelihood estimate (NPMLE) framework for the corresponding models for $S(t|z[0, t])$. Estimation and inferential tools are developed within this general framework, where the distribution of U_t is left unspecified except for Laplace functionals Shiryaev (1960) of the process. Section 3.2 outlines the stochastic process frailty model. Estimation procedures outlined in Section 3.4 and asymptotic derivations in Section 3.5 continue to keep U_t in its most general form; this latter section gives asymptotic properties of estimates for model parameters and the cumulative hazard function. A special case of the stochastic process frailty model is elucidated in Section 3.6, the non-homogeneous Poisson process (NHPP) frailty, assuming observed covariate is a time-dependent binary factor. Examples of a known frailty, Poisson frailty and Poisson process frailty are given in Section 3.7, assuming observed covariate is a time-independent binary factor. A simulation conducted in this special case is given in Section 3.8. A discussion follows in Section 3.9.

3.2 Stochastic Process Frailty Model

Suppose the frailty process, U_t, follows an unspecified distribution that is a functional of the observed history of time-dependent covariates, $z[0, t]$. The survival function $S(t|z[0, t])$ will be a weighted average over possible histories of U_s, $s \in [0, t]$, or $U[0, t]$.

That is,

$$(3.2) \qquad S(t|z[0,t]) = E_{U[0,t]}\left(e^{-\int_0^t U_s dH_s}\right).$$

with corresponding cumulative hazard function

$$(3.3) \qquad \Lambda(t|z[0,t]) = -\log E_{U[0,t]}(e^{-\int_0^t U_s dH_s}).$$

In the case when $U_s = Ue^{\beta^T z_s}$, (3.2) reduces to (3.1), however other functional forms of U_s in terms of z_s or $z[0,s]$ would also fit into this framework.

Based on (3.2), the hazard function becomes

$$
\begin{aligned}
\lambda(t|z[0,t]) &= \frac{E_{U[0,t]}\left\{U_t \exp(-\int_0^t U_s dH_s)\right\}}{E_{U[0,t]}\left\{\exp(-\int_0^t U_s dH_s)\right\}} dH_t/dt \\
(3.4) \qquad &= E(U_t|T \geq t, z[0,t])dH_t/dt,
\end{aligned}
$$

by an argument shown in Appendix (B.1). Thus, the hazard is the product of the baseline hazard and an expectation that can be interpreted as an imputation of the unobserved value of the frailty process at time t, given the individual is alive at t. In the case of the traditional Cox model, $U_s = e^{\beta^T z_s}$, so that the traditional Cox model can be thought of as having an (imputed) non-random frailty.

Let $H[0,t]$ denote the history of baseline hazard values from 0 to t. To emphasize that the imputed frailty, $E(U_t|T \geq t, z[0,t])$, depends on $H[0,t]$, as well as the history of observed covariates, $z[0,t]$, we introduce the notation

$$\Theta^0(t, H[0,t]|z[0,t]) = E(U_t|T \geq t, z[0,t]),$$

so that $\lambda(t|z[0,t]) = \Theta^0(t, H[0,t]|z[0,t])dH_t/dt$. Importantly, under independent censoring, conditioning upon $T \geq t$ is equivalent to conditioning upon $X = t, \Delta = 0$. This corresponds to the usual idea of relating net and crude hazards under independent censoring.

We will see that in the more general stochastic process case the impute, $\Theta^0(t, H[0, t]|z[0, t])$, allows the ability to multiplicatively increase or decrease $\lambda(t|z[0, t])$. This differs from the more standard case with a time-independent frailty, where the imputes only decrease the hazard multiplicatively over time Vaupel et al. (1979). In Section 3.6 we will see an example where U_t gives non monotonicity of the imputes.

3.3 Imputation of U_t

In this section, we expand our understanding of the role of imputed frailties, their interpretation within likelihoods, estimating equations and their NPMLEs.

3.3.1 Data structure and notation

Suppose failure time, T, and right censoring time, C, are independent random variables. C is also independent of U_s, $s \in [0, \tau]$. We observe a vector of time-dependent covariates, $z(t)$, through time t, with $t \in [0, \tau]$ and $\tau < \infty$, appropriately constrained to be during the study period. Let $X = min(T, C)$ and $\Delta = I(T \le C)$. Each independent individual, i, contributes observed data $\{X_i, \Delta_i, z_i[0, X_i]\}$, $i = 1, \ldots, n$, where $z_i[0, X_i]$ is the subject-specific trajectory of the covariate process in the interval where the subject is at risk. The observed data can be represented equivalently in counting process notation as $\{N_i(t), N_i^c(t), Y_i(t), z_i(t)\}$, $i = 1, \ldots, n$, where $N_i(t) = I(X_i \le t)\Delta_i$ gives the failure counting process, $N_i^c(t) = I(X_i \le t)(1 - \Delta_i)$ gives the censoring counting process, and $Y_i(t) = I(X_i \ge t)$ gives the at risk process Nielson et al. (1992).

3.3.2 General form of imputes, U_x, given knowledge to time t

In this paper we begin by considering imputes of U_t based on full knowledge of death and censoring information for the subject, including the future beyond t; that is, $E(U_t|X, \Delta, z[0, X])$, or equivalently, $E(U_t|N(s), Y(s), z(s), s \in [0, \tau])$. In other words,

the impute of U_t involves information not only at t, but also observed death or censoring information in the future. In Appendix B.1, Bayes rule is used to obtain the distribution of U_x given $\{X, \Delta, z[0, X]\}$, or more particularly, given $\{X = t, \Delta, z[0, t]\}$, for $t > x$. The resulting expectations become

$$(3.5) \qquad E(U_x | X = t, \Delta = 0, z[0, t]) \;=\; \frac{E_{U[0,t]}\left\{U_x \exp(-\int_0^t U_s dH_s)\right\}}{E_{U[0,t]}\left\{\exp(-\int_0^t U_s dH_s)\right\}}$$

$$(3.6) \quad \text{and } E(U_x | X = t, \Delta = 1, z[0, t]) \;=\; \frac{E_{U[0,t]}\left\{U_x U_t \exp(-\int_0^t U_s dH_s)\right\}}{E_{U[0,t]}\left\{U_t \exp(-\int_0^t U_s dH_s)\right\}}.$$

For simplicity, we introduce the following notation for the above expressions,

$$\Theta^\Delta(x, H[0, t] | z[0, t]) \;=\; \frac{E_{U[0,t]}\left\{U_x U_t^\Delta \exp(-\int_0^t U_s dH_s)\right\}}{E_{U[0,t]}\left\{U_t^\Delta \exp(-\int_0^t U_s dH_s)\right\}}$$

$$(3.7) \qquad\qquad\qquad\qquad\quad =\; E(U_x | X = t, \Delta, z[0, t]).$$

Recall that in the case when $\Delta = 0$, this definition of $\Theta^0(x, H[0, t] | z[0, t])$ is identical to that in equation (3.5), so that imputes based upon knowledge of observed censoring at time t are the same as imputes for survivors in the absence of additional information. These imputes, $\Theta^\Delta(x, H[0, t] | z[0, t])$, become useful in describing the form of the NPMLE for H_t.

For $t > x$, define

$$\Theta^0(x, t, H[0, t] | z[0, t]) \;=\; E(U_x U_t | X = t, \Delta = 0, z[0, t])$$
$$=\; \frac{E_{U[0,t]}\left\{U_x U_t \exp(-\int_0^t U_s dH_s)\right\}}{E_{U[0,t]}\left\{\exp(-\int_0^t U_s dH_s)\right\}}.$$

This term algebraically relates $\Theta^0(t, H[0, t] | z[0, t])$ to $\Theta^1(x, H[0, t] | z[0, t])$, since

$$\frac{\Theta^0(x, t, H[0, t] | z[0, t])}{\Theta^0(t, H[0, t] | z[0, t])} = \Theta^1(x, H[0, t] | z[0, t]).$$

In the following section, it will be convenient to have the functional derivatives of $\Theta^0(t, H[0,t]|z[0,t])$ and $\log \Theta^0(t, H[0,t]|z[0,t])$ with respect to the baseline hazard at time x, dH_x.

First, for $t > x$,

$$\frac{\partial}{\partial dH_x} \Theta^0(t, H[0,t]|z[0,t])$$

$$= \frac{\partial}{\partial dH_x} \frac{E_{U[0,t]} \left\{ U_t \exp(-\int_0^t U_s dH_s) \right\}}{E_{U[0,t]} \left\{ \exp(-\int_0^t U_s dH_s) \right\}}$$

$$= -\frac{E_{U[0,t]} \left\{ U_x U_t \exp(-\int_0^t U_s dH_s) \right\}}{E_{U[0,t]} \left\{ \exp(-\int_0^t U_s dH_s) \right\}}$$

$$+ \frac{E_{U[0,t]} \left\{ U_t \exp(-\int_0^t U_s dH_s) \right\}}{E_{U[0,t]} \left\{ \exp(-\int_0^t U_s dH_s) \right\}} \frac{E_{U[0,t]} \left\{ U_x \exp(-\int_0^t U_s dH_s) \right\}}{E_{U[0,t]} \left\{ \exp(-\int_0^t U_s dH_s) \right\}}$$

$$= -\Theta^0(x, t, H[0,t]|z[0,t]) + \Theta^0(x, H[0,t]|z[0,t])\Theta^0(t, H[0,t]|z[0,t])$$

$$= -E(U_x U_t | X = t, \Delta = 0, z[0,t])$$

$$+ E(U_x | X = t, \Delta = 0, z[0,t]) E(U_t | X = t, \Delta = 0, z[0,t])$$

$$(3.8) \qquad = -cov(U_x, U_t | X = t, \Delta = 0, z[0,t]).$$

Hence this term is known based on the covariance structure of the stochastic process, U_t, without requiring an explicit form for U_t. Using this result, we also get for $t > x$

$$\frac{\partial}{\partial dH_x} \log \Theta^0(t, H[0,t]|z[0,t])$$

$$= -\frac{\Theta^0(x, t, H[0,t]|z[0,t]) - \Theta^0(x, H[0,t]|z[0,t])\Theta^0(t, H[0,t]|z[0,t])}{\Theta^0(t, H[0,t]|z[0,t])}$$

$$(3.9) \qquad = -\Theta^1(x, H[0,t]|z[0,t]) + \Theta^0(x, H[0,t]|z[0,t]).$$

The above relationship expresses the difference in the value of the imputed frailty process U_x at time $x < t$ for a subject known to survive past t ($\Theta^0(x, H[0,t]|z[0,t])$) vs. the one known to fail at t ($\Theta^1(x, H[0,t]|z[0,t])$), the history on $[0,x]$ being the same for both.

In the time-independent frailty case, $cov(U_x, U_t | X = t, \Delta = 0, z[0, t]) = Var(U | X = t, \Delta = 0, z[0, t]) > 0$, and consequently from (3.8) and (3.9) the failing subject always has larger imputed risk. This is still the case with frailty being a stochastic process with positive conditional covariance. That is, $cov(U_x, U_t | X = t, \Delta = 0, z[0, t]) > 0$. However, when this covariance is negative, smaller U_x is generally associated with larger future U_t leading to potentially greater risk of failure. Inverting the previous sentence, information that subject failed at t may lead to smaller imputed U_x compared to the subject known to survive past t, $\Theta^1(x, H[0, t] | z[0, t]) < \Theta^0(x, H[0, t] | z[0, t])$. Consequently, the frailty process has a sense of a measure of relative risk only when the conditional covariance is positive, and the risk of failure is positively "correlated" with the frailty.

3.3.3 NPMLE based on imputes of U_t

Using $S(t | z[0, t])$ as in (3.2), $\Lambda(t | z[0, t])$ as in (3.3) and $\lambda(t | z[0, t])$ as in (3.4), the log-likelihood becomes

$$
\begin{aligned}
l &= \sum_{i=1}^{n} \int_0^\tau \left\{ \log \lambda(t | z_i[0, t]) \right\} dN_i(t) - \left\{ dN_i^c(t) + dN_i(t) \right\} \Lambda(t | z_i[0, t]) \\
&= \sum_{i=1}^{n} \int_0^\tau \left\{ \log \Theta^0(t, H[0, t] | z_i[0, t]) + \log dH_t \right\} dN_i(t) \\
&\quad + \left\{ dN_i^c(t) + dN_i(t) \right\} \log E(e^{-\int_0^t U_s dH_s}).
\end{aligned}
$$

(3.10)

The role of the impute, $\Theta^0(t, H[0, t] | z_i[0, t])$, in the likelihood is to represent the frailty of a person with a similar history to individual i given this person survived beyond time t; in the Cox model, the non-random impute, $e^{\beta^T z_{s,i}}$ would stand in for this individual.

Based on taking the functional derivative with respect to dH_x in (3.10), the correspond-

ing estimating equation for H_t becomes

$$
\begin{aligned}
U_H^*(H[0,\tau],t,\beta) &= \sum_{i=1}^{n} \int_0^t \Bigg(dN_i(x) \\
&\quad + \Bigg[\int_x^\tau \Big\{ \Theta^0(x,H[0,u]|z_i[0,u]) - \Theta^1(x,H[0,u]|z_i[0,u]) \Big\} dN_i(u) \\
&\quad - \int_x^\tau \Theta^0(x,H[0,u]|z_i[0,u]) \{dN_i(u) + dN_i^c(u)\} \Bigg] dH_x \Bigg) \\
&= \sum_{i=1}^{n} \int_0^t \Bigg(dN_i(x) \\
&\quad - \Bigg\{ \int_x^\tau \Theta^1(x,H[0,u]|z_i[0,u])dN_i(u) \\
&\quad + \int_x^\tau \Theta^0(x,H[0,u]|z_i[0,u])dN_i^c(u) \Bigg\} dH_x \Bigg) \\
&= \sum_{i=1}^{n} \int_0^t \Bigg(dN_i(x) \\
&\quad - \Bigg[\int_x^\tau \Theta^{\Delta_i}(x,H[0,u]|z_i[0,u]) \{dN_i(u) + dN_i^c(u)\} \Bigg] dH_x \Bigg).
\end{aligned}
$$

Set this equation to zero, then take derivative with respect to t, we may obtain an equation for dH_x. Integrating this equation for dH_x from 0 to t, we have the following NPMLE equation for H_t.

$$
(3.11) \quad d\widehat{H}_t = \sum_{i=1}^{n} \int_0^t \frac{dN_i(x)}{\sum_{i=1}^{n} \int_x^\tau \Theta^{\Delta_i}(x,H[0,u]|z_i[0,u]) \{dN_i(u) + dN_i^c(u)\}}.
$$

The same form of NPMLE can be found in traditional Cox models, models with time independent frailty as in Tsodikov (2003), Kosorok et al. (2004), Zeng and Lin (2007), Chen (2009), and Chen (2010). In the special case where $U[0,t]$ is fully observed, (3.2) reduces to a traditional Cox model with observed time dependent covariates and (3.35) reduces to the Breslow estimator for that case. In the special case where the frailty is time-independent, (3.35) reduces to the estimator developed by Tsodikov (2003), Kosorok et al. (2004), Zeng and Lin (2007), 9Chen (2009), and Chen (2010). In each case, the imputes $\Theta^{\Delta_i}(x,H[0,u]|z_i[0,u])$ in (3.35) capture the full information available on the subject.

We build our estimation procedure and asymptotic theory on the log-likelihood represented in next section as that form yields convenient martingale property of the score.

3.4   Estimation of model parameters and baseline cumulative hazard

Consider the log-likelihood (3.10) in Section 3.3. The trailing term involving $-\int_0^\tau \log E(e^{-\int_0^t U_s dH_s}) \{dN_i^c(t) + dN_i(t)\}$ can be rewritten as $\int_0^\tau Y_i(t)\Theta^0(t, H[0,t]|z[0,t])dH_t$ upon noting that $\Lambda(t|z[0,t]) = \int_0^t \Theta^0(s, H[0,s]|z[0,s])dH_s$. So the log-likelihood in (3.10) becomes

$$
\begin{aligned}
l \ = \ & \sum_{i=1}^n \int_0^\tau \left\{ \log dH_t + \log \Theta^0(t, H[0,t]|z_i[0,t]) \right\} dN_i(t) \\
& -Y_i(t)\Theta^0(t, H[0,t]|z_i[0,t])dH_t,
\end{aligned}
$$

(3.12)

where here and below the subscript i denotes individual level information, and

$$
\Theta^\beta(t, H[0,t])|z[0,x] \ = \ \frac{\partial}{\partial \beta}\Theta^0(t, H[0,t]|z[0,t]).
$$

Also, let

(3.13)
$$
dM(t) \ = \ dN(t) - Y(t)\Theta^0(t, H[0,t]|z[0,t])dH_t.
$$

Then differentiating the likelihood we can obtain the following score functionals for

(H_t, β):

$$
\begin{aligned}
U_H(H[0,\tau], t, \beta) &= \sum_{i=1}^{n} \int_0^t \bigg[dN_i(x) + dH_x \int_x^\tau \Big\{ \Theta^0(x, H[0,u]|z_i[0,u]) \\
&\quad - \Theta^1(x, H[0,u]|z_i[0,u]) \Big\} dN_i(u) \\
&\quad - Y_i(x)\Theta^0(x, H[0,x]|z_i[0,x]) dH_x \\
&\quad + dH_x \int_x^\tau Y_i(u) \Big\{ - \Theta^0(x, t, H[0,t]|z[0,t]) \\
&\quad + \Theta^0(x, H[0,t]|z[0,t])\Theta^0(t, H[0,t]|z[0,t]) \Big\} dH_u \bigg] \\
&= \sum_{i=1}^{n} \int_0^t \bigg(dM_i(x) + \Big[\int_x^\tau \Big\{ \Theta^0(x, H[0,u]|z_i[0,u]) \\
&\quad - \Theta^1(x, H[0,u]|z_i[0,u]) \Big\} dM_i(u) \Big] dH_x \bigg)
\end{aligned}
$$

(3.14)

(3.15) $\quad U_\beta(H[0,\tau], \beta) = \sum_{i=1}^{n} \int_0^\tau \dfrac{\Theta^\beta(x, H[0,x]|z_i[0,x])}{\Theta^0(x, H[0,x]|z_i[0,x])} dM_i(x).$

Solutions to the above estimating equations, when set to zero, give $\widehat{H}_t$ and $\widehat{\beta}$.

3.5 Asymptotics of estimates

To demonstrate the consistency and asymptotic normality of $(\widehat{H}_t, \widehat{\beta})$, for a time point t, we follow the ideas presented in Chen (2009) and Chen (2010). For models like (3.1), Kosorok et al. (2004), Zeng and Lin (2007), and Zeng and Lin (2010) establish the asymptotic properties of the NPMLE using empirical processes. We follow an alternative arguments of Chen (2009) and Chen (2010), who elucidate the weak convergence of the NPMLE using martingales.

3.5.1 Martingale properties of estimating equations

We show that (3.14) is a multivariate martingale if (H_t, β) are evaluated at their true values (H_t, β_0).

After exchanging integrals and some algebra, (3.14), scaled by $n^{-1/2}$, can be written

as:

$$
\begin{aligned}
n^{-1/2}U_H(H[0,\tau],t,\beta) &= n^{-1/2}\sum_{i=1}^{n}\int_0^t\left[1+\int_0^u\left\{\Theta^0(x,H[0,u]|z_i[0,u])\right.\right. \\
&\qquad\left.\left.-\Theta^1(x,H[0,u]|z_i[0,u])\right\}dH_x\right]dM_i(u) \\
&\quad+n^{-1/2}\int_{t+}^{\tau}\int_0^t\left\{\Theta^0(x,H[0,u]|z_i[0,u])\right. \\
&\qquad\left.-\Theta^1(x,H[0,u]|z_i[0,u])\right\}dH_xdM_i(u) \\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\xi_i(u,t)dM_i(u)
\end{aligned}
$$

(3.16)

where $\xi_i(u,t) = I(u\leq t)+\int_0^{u\wedge t}\left\{\Theta^0(x,H[0,u]|z_i[0,u])-\Theta^1(x,H[0,u]|z_i[0,u])\right\}dH_x$,

is a bounded and predictable process up to u.

Prvious authors have studied stochastic integrals of the form $\int_0^{\tau}\xi(u,t)dM(u)$ Lachout (2001), calling them linear transformation of Wiener process. In general, $\int_0^{\tau}\xi(u,t)dM(u)$ is not a martingale unless $\xi(u,t)$ does not depend on t for $u\leq t$, as in this case. That is, $\xi(u,t)=\xi(u)=1+\int_0^u\left\{\Theta^0(x,H[0,u]|z_i[0,u])-\Theta^1(x,H[0,u]|z_i[0,u])\right\}dH_x$, for $u\leq t$. Further details are provided in Appendix (B.3).

The corresponding predictable variation process for (3.16) is:

(3.17) $$\qquad n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\xi_i^2(u,t)Y_i(u)\Theta^0(u,H[0,u]|z_i[0,u])dH_u,$$

that converges in probability to the following deterministic functional as $n\to\infty$:

$$\int_0^{\tau}\xi^2(u,t)P(T\geq u)\Theta^0(u,H[0,u]|z[0,u])dH_u,$$

Assume $\xi(u,t)$ and $\Theta^0(u,H[0,u]|z_i[0,u])$ are bounded and smooth, under certain regularity conditions (listed in Appendix (B.2), Martingale Central Limit Theorum (MCLT) gives the following weak convergence:

(3.18) $$\qquad n^{-1/2}U_H(H[0,\tau],t,\beta)\Rightarrow\tilde{W}_{U_H}(t,\beta),$$

where $\tilde{W}_{U_H}(t, \beta)$ is a zero-mean Gaussian process with the following covariance function:

$$(3.19) \qquad \sigma_H^2(s, t) \;=\; \int_0^\tau \xi(u, s)\xi(u, t)P(T \geq u)\Theta^0(u, H[0, u]|z[0, u])dH_u,$$

for $t, s \in [0, \tau]$.

Similarly, assuming $\frac{\Theta^\beta(x, H[0,x]|z_x)}{\Theta^0(x, H[0,x]|z_x)}$ is bounded and smooth, $n^{-1/2}U_\beta(H[0, \tau], \beta_0)$ is also a martingale with predictable variation process:

$$(3.20) \qquad n^{-1} \sum_{i=1}^n \int_0^\tau \frac{\{\Theta^\beta(u, H[0, u]|z_i[0, u])\}^2}{\Theta^0(u, H[0, u]|z_i[0, u])}Y_i(u)dH_u,$$

that converges to the following deterministic functional $\sigma_\beta^2(\beta)$ as $n \to \infty$:

$$(3.21) \qquad \sigma_\beta^2(\beta) \;=\; \int_0^\tau \frac{\{\Theta^\beta(u, H[0, u]|z[0, u])\}^2}{\Theta^0(u, H[0, u]|z[0, u])}P(T \geq u)dH_u.$$

By MCLT,

$$(3.22) \qquad n^{-1/2}U(H[0, \tau], \beta_0) \Rightarrow \tilde{W}_{U_\beta}(\beta),$$

where $\tilde{W}_{U_\beta}$ is a zero-mean Gaussian process with the covariance $\sigma_\beta^2(\beta)$.

Meanwhile, $n^{-1/2}U(H[0, \tau], t, \beta_0)$, for some t, and $n^{-1/2}U(H[0, \tau], \beta_0)$ is also a martingale that has the following predictable covariation process:

$$(3.23) \qquad n^{-1} \sum_{i=1}^n \int_0^\tau \xi(u, t)\Theta^\beta(u, H[0, u]|z_i[0, u])Y_i(u)dH_u,$$

that converges to the following deterministic functional $\sigma_{H,\beta}^2(t, \beta)$, as $n \to \infty$:

$$\sigma_{H,\beta}^2(t, \beta) \;=\; \int_0^\tau \xi(u, t)\Theta^\beta(u, H[0, u]|z_i[0, u])P(T \geq u)dH_u.$$

Hence, by MCLT, $n^{-1/2}\{U(H[0, \tau], t, \beta_0), U(H[0, \tau], \beta_0)\}$ converges weakly to a Gaussian process with covariance function $\sigma_{H,\beta}^2(t, \beta)$.

Therefore, (3.14) is a martingale that coverges to a zero-mean Gaussian process $\tilde{W}_U(t, \beta)$. That is,

$$(3.24) \qquad n^{-1/2}\begin{bmatrix} U_H(H[0, \tau], t, \beta) \\[2mm] U_\beta(H[0, \tau], \beta) \end{bmatrix} \Rightarrow \tilde{W}_U(t, \beta)$$

The covariance function of $\tilde{W}_U(t, \beta)$ is characterized by the following:

$$(3.25) \qquad \Sigma(s, t, \beta) = \begin{bmatrix} \sigma_H^2(s, s) & \sigma_H^2(s, t) & \sigma_{H,\beta}^2(s, \beta) \\ \sigma_H^2(t, s) & \sigma_H^2(t, t) & \sigma_{H,\beta}^2(t, \beta) \\ \sigma_{H,\beta}^2(s, \beta) & \sigma_{H,\beta}^2(t, \beta) & \sigma_\beta^2(\beta) \end{bmatrix}$$

for $s, t \in [0, \tau]$.

The components in $\Sigma(s, t, \beta)$ can be consistently estimated by the corresponding predictable variation (covariation) processes as in (3.17), (3.20) and (3.23) by replacing dH with $d\widehat{H}$ and replacing β with $\widehat{\beta}$.

Remark 1: We can also write (3.14) as

$$U_H(H[0, \tau], x, \beta) = \sum_{i=1}^n dN_i(x) - \sum_{i=1}^n Y_i(x)\Theta^0(x, H[0, x]|z_i[0, x])w_i(x, \tau)dH_x,$$

$$U_\beta(\beta) = \sum_{i=1}^n \int_0^\tau \frac{\Theta^\beta(x, H[0, x]x|z_i[0, x])}{\Theta^0(x, H[0, x]|z_i[0, x])}dM_i(x),$$

where

$$w_i(x) = 1 - \frac{\int_x^\tau \left\{ \Theta^0(x, H[0, u]|z_i[0, u]) - \Theta^1(x, H[0, u]|z_i[0, u]) \right\} dM_i(u)}{\Theta^0(x, H[0, x]|z_i[0, x])}$$

As discussed in Chen (2009) and Chen (2010), $w(x)$ has unit expectation, which is a consequence of the martingale property of the score. Also as discussed in Chen (2009), the martingale suggests:

$$E\left\{ \sum_{i=1}^n dN_i(t)|\mathcal{F}_{t-} \right\} = \sum_{i=1}^n Y_i(t)\Theta^0(t, H[0, t]|z_i[0, t])dH_t,$$

that leads to inefficient NPMLE as:

$$(3.26) \qquad d\widehat{H}_t = \sum_{i=1}^n \frac{dN_i(x)}{\sum_{i=1}^n Y_i(x)\Theta^0(t, H[0, t]|z_i[0, t])}.$$

In this way, we can interpret the weight w as correcting for the loss of information by using only past histories in the imputation operators vs. the full available information on

each subject. In expectation, this quantity is 1, suggesting that both (3.35) and (3.26) lead to consistent estimate for H.

Remark 2: The implication of weak convergence of (3.14) is that $n^{-1}(U_H, U_\beta)$ converges to zero in probability by Lenglart Inequality. The consistency of $\{(\widehat{H}, \widehat{\beta})\}$ follows immediately, see, Fleming and Harrington (2005).

Remark 3: Murphy (1994), Murphy (1995) and Murphy and van der Vaart (2000) showed weak convergence of the NPMLE based on the similar models as (3.1). In Appendix (B.8), we present the works using profile likelihood and show that the converging process for $\widehat{H}$ satisfies second order non-homogeneous stochastic equation. In general, there is no explicit solution to this type of equation. The difference between profile likelihood approach and the approaches used by Bagdonnavicius and Nikulin (1999) and Martinussen and Scheike (2006), is that the latter used modified profile likelihood similar to partial likelihood as in Cox (1972). The parital likelihood approach is not fully efficient for models (3.1) and (3.2), because the hazard functions are usually not linear in H.

3.5.2 Weak Convergence

In this section we derive the asymptotic distribution for $\sqrt{n}(\widehat{H}_t - H_t, \widehat{\beta} - \beta_0)$.

Taylor expansion suggests the following, ignore $o_p(1)$ terms:

$$
n^{-1/2}
\begin{bmatrix}
U_H(\widehat{H}[0,\tau], t, \widehat{\beta}) - U_H(H[0,\tau], t, \beta) \\
U_\beta(\widehat{H}[0,\tau], \widehat{\beta}) - U_\beta(H[0,\tau], \beta)
\end{bmatrix}
$$

$$
= \sqrt{n}
\begin{bmatrix}
\int_0^t n^{-1}U_H^H(H[0,\tau], t, y, \beta)d(\widehat{H} - H)(y) + n^{-1}U_H^\beta(H[0,\tau], t, \beta)(\widehat{\beta} - \beta) \\
\int_0^\tau n^{-1}U_\beta^H(H[0,\tau], y, \beta)d(\widehat{H} - H)(y) + n^{-1}U_\beta^\beta(H[0,\tau], \beta)(\widehat{\beta} - \beta)
\end{bmatrix}
$$

$$
= \sqrt{n}
\begin{bmatrix}
\int_0^t \phi_H^H(H[0,\tau], t, y, \beta)d(\widehat{H} - H)(y) + \phi_H^\beta(H[0,\tau], t, \beta)(\widehat{\beta} - \beta) \\
\int_0^\tau \phi_\beta^H(H[0,\tau], y, \beta)d(\widehat{H} - H)(y) + \phi_\beta^\beta(H[0,\tau], \beta)(\widehat{\beta} - \beta)
\end{bmatrix}
$$

where U_H^H, U_H^β, U_β^H and U_β^β, together with their corresponding converging functions, ϕ_H^H,

ϕ_H^β, ϕ_β^H and ϕ_β^β are given in Appendix B.5.

First, $U_H(\widehat{H}[0,\tau], t, \widehat{\beta})$ and $U_\beta(\widehat{H}[0,\tau], \widehat{\beta})$ are zero. Meanwhile, based on (3.24) in Section 3.5.1, we have

$$\sqrt{n} \begin{bmatrix} \int_0^t \phi_H^H(H[0,\tau], t, y, \beta) d(\widehat{H} - H)(y) + \phi_H^\beta(H[0,\tau], t, \beta)(\widehat{\beta} - \beta) \\ \int_0^\tau \phi_\beta^H(H[0,\tau], y, \beta) d(\widehat{H} - H)(y) + \phi_\beta^\beta(H[0,\tau], \beta)(\widehat{\beta} - \beta) \end{bmatrix}$$

(3.27) $\qquad \Rightarrow \tilde{W}_U(t, \beta),$

where $\tilde{W}_U(t, \beta)$ has covariance matrix $\Sigma(s, t, \beta)$ in (3.25).

Let $W_1(t)$ denote $\int_0^t \phi_H^H(H[0,\tau], t, y, \beta) dV(y) + \phi_H^\beta(H[0,\tau], t, \beta)(\widehat{\beta} - \beta)$, where $V(y) = (\widehat{H} - H)y$ and W_2 denote $\int_0^\tau \phi_\beta^H(H[0,\tau], y, \beta) dV(y) + \phi_\beta^\beta(H[0,\tau], \beta)(\widehat{\beta} - \beta)$.

The covariance for the left hand side of (3.27), as shown in Appendix C.2, is

$$\Sigma^*(s, t, \beta) = \begin{bmatrix} \sigma_{W_1}^{*,2}(s, s) & \sigma_{W_1}^{*,2}(s, t) & \sigma_{W_1, W_2}^{*,2}(s, \beta) \\ \sigma_{W_1}^{*,2}(t, s) & \sigma_{W_1}^{*,2}(t, t) & \sigma_{W_1, W_2}^{*,2}(t, \beta) \\ \sigma_{W_1, W_2}^{*,2}(s, \beta) & \sigma_{W_1, W_2}^{*,2}(t, \beta) & \sigma_{W_2}^{*,2}(\beta) \end{bmatrix}.$$

Based on (3.27), we have $\Sigma^*(s, t, \beta) = \Sigma(s, t, \beta)$. As a result, the covariance function of $\sqrt{n}(\widehat{H}_t - H_t, \widehat{\beta} - \beta_0)$ are expressed as solutions to this system of Volterra integral equations.

The observed information matrix I will be used to give estimate of the covariance of the estimating equations. The form of I is given in Appendix (B.7).

3.6 Example: U_t as a non homogeneous Poisson process (NHPP)

In this section we illustrate our method in a special case where U_t is a non-homogeneous Poisson process (NHPP) characterized by cumulative intensity $\int_0^t \mu_s ds$ with $\mu_s = e^{\beta^T z_s}$. This process is well suited to describe accumulating lesions or other damages in a system induced by time-varying environmental exposure linked to the NHPP intensity. For now

we keep z_s general; in the simulation section that follows z_s is a time-dependent indicator function of progression that turns on at a random point during followup.

As in (3.2), $S(t|z[0,t]) = E_{U[0,t]}(e^{-\int_0^t U_s dH_s})$. In the case of NHPP, U_s is a count over time taking values, $0, 1, 2, \ldots$ at times $0, T_1, T_2, \ldots$. Let N_t be the number of events in $[0,t]$. Thus, $S(t|z[0,t])$ becomes

$$E_{U[0,t]}(e^{-\int_0^t U_s dH_s})$$

$$= E_{U[0,t]}\left\{E\left(e^{-\int_0^t U_s dH_s}|U_s, s \in [0,t]\right)\right\}$$

$$(3.28) \quad = E_{U[0,t]}E\left\{\exp\left(-\sum_{u=1}^{N_t-1}\int_{T_u}^{T_{u+1}} u dH_s - \int_{T_{N_t}}^t N_t dH_s\right)|U_s, s \in [0,t]\right\},$$

where the inner expectation is taken with respect to event times $T_1, T_2, \ldots, T_{N_t}$. And since after cancellations of like terms

$$-\sum_{u=1}^{N_t-1}\int_{T_u}^{T_{u+1}} u dH_s - \int_{T_{N_t}}^t N_t dH_s = \sum_{u=1}^{N_t} H_{T_u} - N_t H_t,$$

(3.28) becomes

$$(3.29) \quad E_{U[0,t]}\left[\exp(-N_t H_t)E\left\{\exp\left(\sum_{u=1}^{N_t} H_{T_u}\right)|U_s, s \in [0,t]\right\}\right].$$

From McDonald (1947), given N_t events in $[0,t]$, the unordered events $T_1^*, T_2^*, \ldots, T_{N_t}^*$ are independent and identically distributed random variables with density $\frac{\mu_x}{\int_0^t \mu_s ds} = e^{\beta^T z_x}(\int_0^t e^{\beta^T z_s} ds)^{-1}$, for $x \in [0,t]$, so that (3.29) becomes

$$E_{U[0,t]}\left[\exp(-N_t H_t)E\left\{\exp\left(\sum_{u=1}^{N_t} H_{T_u}\right)|U_s, s \in [0,t]\right\}\right]$$

$$= E_{U[0,t]}\left[\exp(-N_t H_t)\left\{E(e^{H_{T^*}}|U_s, s \in [0,t])\right\}^{N_t}\right]$$

$$(3.30) \quad = E_{U[0,t]}\left\{\left(\frac{e^{-H_t}\int_0^t e^{H_s}\mu_s ds}{\int_0^t \mu_s ds}\right)^{N_t}\right\}.$$

Let $B_t = \frac{\int_0^t e^{H_s - H_t}\mu_s ds}{\int_0^t \mu_s ds}$. The remaining random variable in the expectation of (3.30),

N_t, is distributed as a Poisson random variable with mean $\int_0^t \mu_s ds$. Hence (3.30) becomes:

$$
\begin{aligned}
E_{U([0,t])}(B_t^{N_t}) &= \sum_{N_t=0}^{\infty} B_t^{N_t} \frac{1}{\{N_t\}!} \left\{ \int_0^t \mu_s ds \right\}^{N_t} e^{-\int_0^t \mu_s ds} \\
&= \sum_{N_t=0}^{\infty} \frac{1}{\{N_t\}!} \left\{ B_t \int_0^t \mu_s ds \right\}^{N_t} \exp\left(-B_t \int_0^t \mu_s ds \right) \\
&\quad \times \exp\left(B_t \int_0^t \mu_s ds - \int_0^t \mu_s ds \right) \\
&= \exp\left(B_t \int_0^t \mu_s ds - \int_0^t \mu_s ds \right).
\end{aligned}
$$

After minor calculations when substituting B_t into the above, the survival function becomes:

(3.31)
$$
S(t|z[0,t]) = exp\left\{ -\int_0^t \mu_s(1 - e^{H_s - H_t}) ds \right\},
$$

with corresponding hazard function :

$$
\lambda(t|z[0,t]) = \left(e^{-H_t} \int_0^t \mu_s e^{H_s} ds \right) \times \left(\frac{dH_t}{dt} \right).
$$

The imputed frailties for this case, according to (3.7), become

$$
\begin{aligned}
\Theta^0(x, H[0,t]|z[0,t]) &= S(t|z[0,t])^{-1} \\
&\quad \times \frac{\partial}{\partial dH_x} \left[-exp\left\{ -\int_0^t \mu_s(1 - e^{H_s - H_t}) ds \right\} \right] \\
&= \int_0^t \frac{\partial}{\partial dH_x} \mu_s(1 - 1 - e^{H_s - H_t}) ds \\
&= \int_0^x \mu_s e^{H_s - H_t} ds \\
\Theta^1(x, H[0,t]|z[0,t]) &= \int_0^x \mu_s e^{H_s - H_t} ds \times \left(1 + \frac{1}{\int_0^t \mu_s e^{H_s - H_t} ds} \right).
\end{aligned}
$$

Based on (3.14), the estimating equations become

$$
\begin{aligned}
U_H(H[0,\tau],t,\beta) \;=\; & \sum_{i=1}^{n} \int_0^t dN_i(x) - Y_i(x)(e^{-H_x}\int_0^x \mu_{s,i}e^{H_s}ds)dH_x \\
& + \left[\int_x^\tau \frac{\int_0^x \mu_{s,i}e^{H_s}ds}{\int_0^u \mu_{s,i}e^{H_s}ds}\Big\{ dN_i(u) \right. \\
& \left. -Y_i(u)\left(e^{-H_u}\int_0^u \mu_{s,i}e^{H_s}ds \right)dH_u \Big\} \right]dH_x, \\
U(\beta) \;=\; & \sum_{i=1}^{n} \int_0^\tau \frac{\int_0^x (\frac{d\mu_{s,i}}{d\beta})e^{H_s}ds}{\int_0^x \mu_{s,i}e^{H_s}ds}\Big\{ dN_i(x) \\
& -Y_i(x)\left(e^{-H_x}\int_0^x \mu_{s,i}e^{H_s}ds \right)dH_x \Big\}.
\end{aligned}
$$

(3.32)

$\widehat{H}$ and $\widehat{\beta}$ are solutions to the above equations when they are zero. The associated variances can be obtained using the Hessian information matrix as shown in Appendix (B.7). In practice, it is common to numerically solve for $\widehat{H}$ and $\widehat{\beta}$ in estimating equations like (3.32). Similarly, it is common to numerically compute second order derivatives based on the full likelihood when calculating variances. This is the approach we adopt in Section 3.8, for the NHPP example.

Remark 4: When U_t is a Poisson process with a constant mean function $\mu_s = \mu$, the results obtained in this section also apply. Recall that in deriving the form of survival function based on (3.29), given N_t events in $[0,t]$, the unordered events $T_1^*, T_2^*, \ldots, T_{N_t}^*$ are independent and identically distributed random variables with density $1/t$, so that (3.29) becomes, for U_t as a Poisson process,

$$
E_{U[0,t]}\left[\exp(-N_t H_t)E\left\{ \exp\left(\sum_{u=1}^{N_t} H_{T_u} \right) |U_s, s\in[0,t] \right\} \right]
$$

(3.33)
$$
= E\left\{ \left(t^{-1}e^{-H_t}\int_0^t e^{H_s}ds \right)^{N_t} \right\}.
$$

This will be useful in next section where we further look at survival function and hazard function of U_t when it is a Poisson process.

3.7 Other examples of known U, time-independent U

As mentioned in Section 3.2, Cox models and time-independent frailty models are special cases under our modeling framework in (3.2). For simplicity and clarity, we give some simple examples where observed covariate, z, is assumed to be a time-independent binary factor, we compare the different behaviors of survival functions and hazard functions based on the same z and baseline hazard function $H_t = \log((1+t)$.

3.7.1 U is known

When U is fully observed, (3.2) is simply Cox PH model, where $U = e^{\beta z}$. The survival function and hazard function are

$$
\begin{aligned}
S(t|z) &= \exp\left\{-e^{\beta z}\log(1+t)\right\} \\
\lambda(t|z) &= (1+t)^{-1}e^{\beta z}
\end{aligned}
$$

3.7.2 U is Poisson random variable

When U is a Poisson distributed random variable, where the mean of U is $e^{\beta z}$. We can see that (3.2) will also reduce to a Cox PH model, The survival function and hazard function are

$$
\begin{aligned}
S(t|z) &= E(e^{-UH_t}) \\
&= \exp\left\{e^{\beta z}(e^{-H_t}-1)\right\} \\
&= \exp\left\{-e^{\beta z}t/(1+t)\right\} \\
\lambda(t|z) &= (1+t)^{-2}e^{\beta z}
\end{aligned}
$$

3.7.3 U_t is a Poisson process

Interestingly, when U_t is a Poisson process with mean $e^{\beta z}$, (3.2) is still a Cox PH model. Based on (3.33), using the similar arguments as shown in Section 3.6, the survival

function is

$$
\begin{aligned}
S(t|z) &= E\left\{ \left(t^{-1}e^{-H_t} \int_0^t e^{H_s} ds \right)^U \right\} \\
&= \sum_{U=0}^{\infty} \frac{1}{U!} \left\{ B_t e^{\beta z} t \right\}^U \exp\left(-B_t e^{\beta z} t \right) \\
&\quad \times \exp\left(B_t e^{\beta z} t - e^{\beta z} t \right) \\
&= \exp\left\{ (B_t - 1) e^{\beta z} t \right\},
\end{aligned}
$$

where $B_t = t^{-1} \int_0^t e^{H_s - H_t} ds$. The corresponding hazard function becomes

$$
\begin{aligned}
\lambda(t|z) &= (1+t)^{-1} e^{\beta z} \int_0^t \frac{1+s}{1+t} ds \\
&= (1+t)^{-1} e^{\beta z} \left(\frac{t}{1+t} + \frac{t^2}{2+2t} \right)
\end{aligned}
$$

The figures for a $\beta = -0.5$ are presented to compare the different forms of survival functions (Figure 3.1) and hazard functions (Figure 3.2) under three assumptions of U or U_t.
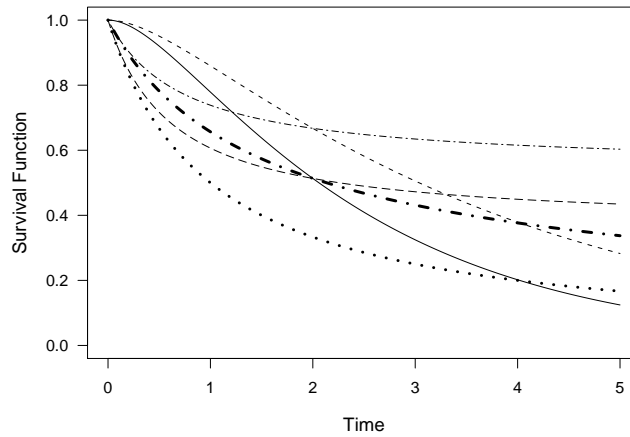

Figure 3.1: Survival functions with observed covariate z=0 and z=1, for known frailty (dotted and dot-dash), Poisson frailty (long-dash and two-dash) and Poisson process fraitly (solid and dash)
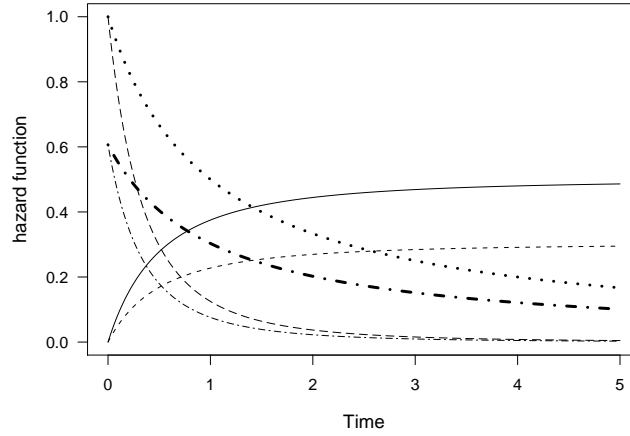
Figure 3.2: Hazard functions with observed covariate z=0 and z=1, for known frailty (dotted and dot-dash), Poisson frailty (long-dash and two-dash) and Poisson process fraitly (solid and dash)

3.8 Simulation results for NHPP special case

We take the NHPP setting described in the previous section as the context for our simulation, where $H_s = \log(1 + s)$ and $z_s = I(A > s)$. The unobserved frailty U_t is an NHPP with cumulative intensity, $\int_0^t e^{\beta^T I(A>s)} ds$. For instance this frailty can be thought of as capturing accumulating, and unobserved, metastasis sites. A is generated from an exponential distribution with rates chosen to produce $P(z_\tau = 0)$ between 70% to 90%; A can be thought of as the time of observed progression. Hence, survival function, $S(t|z[0, t])$, becomes $exp\left\{ -\int_0^t e^{\beta^T I(A>s)}(\frac{t-s}{1+t}) ds \right\}$. That is, when $t < A$, $S(t|z[0, t]) = \exp\left\{ -e^\beta t^2/(2 + 2t) \right\}$. When $t \geq A$, $S(t|z[0, t])$ equals to $\exp\left\{ -e^\beta At/(1 + t) + e^\beta A^2/(2 + 2t) - (t^2 - At)/(1 + t) + (t^2 - A^2)/(2 + 2t) \right\}$.

Based on the inverse transform method, the survival time T is created by solving $S(t|z[0, t]) - B = 0$ for t, where B is generated from a $Uniform(0, 1)$ distribution. Censoring is simulated from a $Uniform(0, \tau)$ distribution with τ chosen to yield the desired

degree of censoring (10%, 30% or 50%).

Table 3.1 and Table 3.2 give results for scenarios varying $\beta = (0, -0.5)$, $N = (200, 300, 500)$ and censoring rates $= (10\%, 30\%, 50\%)$. For purposes of comparison, we estimate model parameters based on (3.31) (NHPP method) as well as parameters from the more traditional Cox model with observed time-dependent $z_s = I(A > s)$ (Cox method).

Recall that when $\beta = 0$, model (3.31) reduces to Poisson process frailty model, which turns out to be a Cox PH model as we discussed in Section 3.7. In this case, β stands for the log odds ratio for both NHPP method and Cox method. Hence it is not surprising that they give similar results for $\widehat{\beta}$.

When $\beta = -0.5$, (3.31) is no longer a Cox model. In order to compare to a Cox model assuming observed U_t with the same interpretation of β, we simulate from the following Cox model

$$
\begin{aligned}
\lambda(t|z[0,t]) &= \left(\int_0^t \mu_s ds \right) \times dH_t/dt \\
&= \left(\int_0^t e^{\beta^T I(A>s)} ds \right) \times dH_t/dt
\end{aligned}
$$

(3.34)

with log-partial likelihood

$$
l = \sum_{i=1}^n \Delta_i \left\{ \log \int_0^{X_i} e^{\beta I(A_i>s)} ds - \log \sum_{j=1}^n Y_j(X_i) \int_0^{X_i} e^{\beta I(A_j>s)} ds \right\},
$$

and NPMLE is:

$$
d\widehat{H}_t = \frac{\sum_{i=1}^n dN_i(t)}{\sum_{j=1}^n Y_j(t) \int_0^t e^{\widehat{\beta} I(A_j>s)} ds}.
$$

when accounting for the unobserved frailty $U[0,t]$ using our NHPP model, bias is negligible, estimated variability is close to empirically derived variability and coverage

Table 3.1: Simulation results for the NHPP example using proposed method (NHPP) and the traditional Cox model with time-dependent covariate (Cox) for $\beta = 0$, based on 1000 simulated datasets with different sample sizes (N) and censoring rates.

N	censoring %	$\beta = 0$							
		NHPP				Cox			
		Estimate[1]	SE[2]	ESD[3]	CP[4]	Estimate[1]	SE[2]	ESD[3]	CP[4]
200	10%	-0.503	0.277	0.285	95.0%	-0.513	0.257	0.273	93.7%
300	10%	-0.495	0.226	0.232	94.7%	-0.511	0.209	0.216	93.7%
500	10%	-0.500	0.175	0.177	95.0%	-0.502	0.161	0.157	94.6%
200	30%	-0.484	0.295	0.288	96.5%	-0.513	0.268	0.285	93.8%
300	30%	-0.509	0.240	0.247	94.4%	-0.512	0.217	0.227	94.6%
500	30%	-0.495	0.186	0.184	94.8%	-0.502	0.168	0.166	95.0%
200	50%	-0.502	0.323	0.339	94.5%	-0.514	0.284	0.301	94.8%
300	50%	-0.498	0.262	0.264	95.1%	-0.514	0.231	0.242	93.6%
500	50%	-0.499	0.203	0.206	95.1%	-0.503	0.178	0.178	94.7%

1. Estimate is the average of estimates of the true parameter values over the 1000 simulations.

2. SE is the average of estimated standard errors over the 1000 simulations.

3. ESD is empirical standard deviation (ESD) of the 1000 estimates.

4. CP is the empirical coverage probability, i.e., the proportion of simulations that cover the true parameter values.

rates are within a reasonable range near 95%. When using the Cox model in (3.34) where $z[0, t]$ is observed, but $U[0, t]$ is assumed to be fixed at its mean $\int_0^t \mu_s ds$, bias is 10 times bigger than those using NHPP model, variance is over estimated that resulting in coverages over 97% on average.

3.9 Discussion

Oftentimes in medicine, economics, reliability and other fields of science, unobserved processes mature over time before manifesting themselves in the observed failure of the system or organism. Our methodology lays a foundation for statistical analysis of such systems in the univariate survival setting. We present a modeling framework and inferential tools for research scenarios where both observed and unobserved covariates change over time that affect survival responses.

Our methodological development does not require that the unobserved stochastic process (frailty) be explicitly specified. Characterization of the conditional covariances of the frailty process is sufficient to estimate regression coefficients and the corresponding baseline cumulative hazard NPMLE. Connections are made between the martingale and the imputation rational behind the NPMLE score equations.

A formal martingale-based approach for inference is also developed. We borrow from Chen (2009) to elucidate the martingale structure of the estimating equations and subsequently use martingale machinery to study asymptotics. Empirical processes offer an alternative, if more complex, tool that is potentially more powerful, particularly if strong consistency is desired.

As an instructive example, we lay out a special case of a non-homogeneous Poisson process frailty that is currently not available in the literature. Simulations contrast our modeling approach to that of a traditional Cox model using observed time-dependent co-

Table 3.2: Simulation results for the NHPP example using proposed method (NHPP) and the Cox model with time-dependent covariate (Cox) for $\beta = -0.5$, based on 1000 simulated datasets with different sample sizes (N) and censoring rates.

| | | $\beta = -0.5$ | | | | | | | |
| | | NHPP | | | | Cox | | | |
N	censoring %	Estimate[1]	SE[2]	ESD[3]	CP[4]	Estimate[1]	SE[2]	ESD[3]	CP[4]
200	10%	-0.491	0.302	0.304	94.8%	-0.449	0.296	0.240	97.2%
300	10%	-0.511	0.246	0.251	95.2%	-0.468	0.242	0.202	97.4%
500	10%	-0.498	0.190	0.202	93.2%	-0.476	0.187	0.163	97.2%
200	30%	-0.505	0.317	0.326	95.4%	-0.449	0.314	0.248	96.8%
300	30%	-0.510	0.259	0.264	94.2%	-0.470	0.256	0.207	97.9%
500	30%	-0.498	0.200	0.204	94.4%	-0.478	0.198	0.171	97.6%
200	50%	-0.530	0.343	0.352	95.4%	-0.447	0.341	0.265	97.0%
300	50%	-0.499	0.279	0.291	95.6%	-0.468	0.279	0.222	97.4%
500	50%	-0.494	0.216	0.217	94.6%	-0.482	0.215	0.182	97.6%

1. Estimate is the average of estimates of the true parameter values over the 1000 simulations.

2. SE is the average of estimated standard errors over the 1000 simulations.

3. ESD is empirical standard deviation (ESD) of the 1000 estimates.

4. CP is the empirical coverage probability, i.e., the proportion of simulations that cover the true parameter values.

variates. Results indicate that inference using traditional methods can be very misleading when the stochastic frailty process is ignored.

Our methodological framework allows for many additional formulations of stochastic frailties. Development and careful study of mechanistic failure models custom-built for specific subject matter problems may bring further useful literature to the emerging field of frailty models. Extending the approach to multivariate survival and general multivariate process models is a promising avenue for future research.

CHAPTER IV

Estimate Dynamic Exposure Effects Using A Semiparametric Incidence Model

4.1 Introduction

The development of cancer lesions can come from more than one source. In many cases cancer forms spontaneously, unrelated to any exposure. An alternative, and often simultaneous, process that may induce lesion formation is exposure to a carcinogenic agent, for instance radiation prescribed for treatment. Lesions from these two processes, background and induced, continue to progress until one or more of the lesions become detectable. Of course the primary hope of the radiation therapy is that existing cancer cells will be killed, dominating any role that induced cancer cells would have on progression. This complicated, generally favorable, interplay between cancer cell induction and death is often called hormesis Yakovlev et al. Yakovlev et al. (1992). In this paper we develop a general survival modeling framework and corresponding nonparametric maximum likelihood estimation (NPMLE) tools that build in hazards due to background and induced lesion development, with parameterization that allows for mediation of hazards as a function of dose or other predictors.

In particular, this research is motivated by a mouse survival study ($n = 691$) with three radiation exposure groups Tsodikov et al. Tsodikov and Muller (1998). The mice

60

are of a special breed that generally show a high background incidence of lymphomas in the absence of any exposure. Two hundred ninety four (42.5%) of the mice were injected with 0.15 Gy of radiation at baseline (acute dosing), two hundred ninety nine (43.3%) of the mice were given 0.15 Gy of radiation parcelled out over 36-week period (protracted dosing) and the remaining ninety eight mice (14.2%) were not given radiation (control). Mice were followed for periods varying up to approximately 1000 days, depending on the entry into the study; with censoring % was approximately 44.6% (42.3% for acute dosing group, 47.6% for protracted group, 42.9% for control).

Figure 4.1 shows lymphoma-free survival estimated via Kaplan Meier (KM) Kaplan and Meier (1958), with the end of the protracted dosing period shown as a vertical dashed line at 252 days. The hormesis effect can be appreciated when viewing time to lymphoma detection. The acute dosing group has the highest survival rates over time, indicating that, compared to the control group, any induced lesion development from the radiation dose is counteracted by cancer cell death from that single dose at time 0. However, when the same radiation dose was spread over 252 days in the protracted dosing group, survival declined dramatically before eventually recovering to control group levels approximately 500 days after dosing stopped. A semiparametric approach is attractive since we do not have sufficient information about the dynamics of lesion progression and their influence on baseline hazards originating from background and induced lesions. The Cox proportional hazard model Cox (1972) is a popular semiparametric approach that links risk factors to a hazard function, and this is our starting point. Fitting dose as an observed time-dependent covariate fails to capture the non-proportional hazards in the three dosing groups shown in Figure 1. Stratification of the baseline hazard by the three dosing groups with a time-varying dose covariate could be employed to capture non-proportional hazards. For example, one might attempt a model of the form $\lambda(t|\text{dose}[0, t]) = \lambda_{0g}(t) \exp\{\beta \times \text{dose}(t)\}$, where the subscript
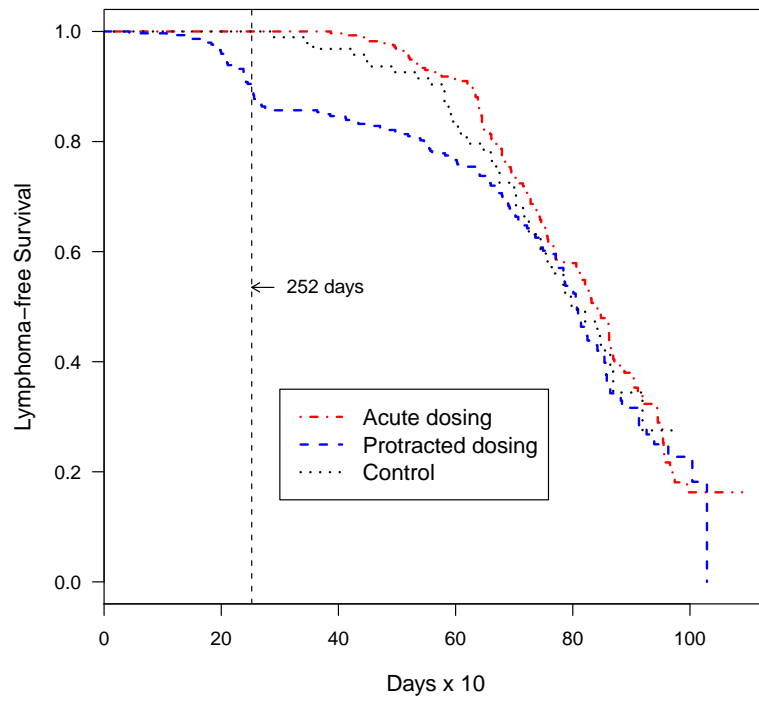
Figure 4.1: Lymphoma Free KM Survivals by Groups

g represents dosing group. However, a shared parameter for β across dosing groups is not plausible for this setting. In the acute dosing group, β would stand in for the effect over the entire study period of total dose at time 0, whereas for the protracted dosing group, β represents the effect of partial dose over smaller time periods. Another parameterization option would be to assume that the total dose over time acts on survival through β as $\lambda(t|\text{dose}[0,t]) = \lambda_{0g}(t)\exp(\beta \times \text{total dose})$. However, as seen in Figure 1, markedly different survival curves for groups with the same total dose over the study period precludes assuming a shared parameter of this form.

Our goal is to provide a modeling framework allows parameterization of a protracted dosing setting where induced hazards as linked to continued exposure to radiation, this framework also allows for simultaneous modeling of hazards of the more traditional proportional hazards form, where for instance, a single dose at baseline affects subsequent hazards.

The rest of manuscript is structured as follows: Section 4.2 introduces intuition and notation driving the model. Section 4.3 gives estimation and inferential procedures. Then we will revisit the mouse study and interpretation of results in Section 4.4. A discussion follows in Section 4.5.

4.2 The incidence model and its generic form

We treat radiation dose as a dynamic time-changing factor. A detailed parametric mechanistic model for dynamic exposure effects was proposed in Tsodikov and Muller (1998), and we generally follow a similar idea here with the aim to arrive at a generic semiparametric form inherrent to such models.

Consider a competing risk with a cumulative hazard H originating at the point in time x. Mechanistically, this can be thought of as a cancer lesion formed at time x and capable

of producing an observed survival response (incidence of observed tumor) at some later time $t > x$ following a progression period $t - x$. Associated with this risk is a survival function

$$(4.1) \qquad F(t - x) \;=\; \exp\{-H(t - x)\} = \exp\left\{ - \int_x^t dH(y - x) \right\},$$

where integration (accumulation) of hazard occurs in follow-up time $y \in [x, t]$, and the lesion-specific risk has a time origin at $y = x$, the point of risk formation. Suppose now that a dynamic exposure has a multiplicative "treatment" effect $\eta(x, y)$ accumulating in follow-up time period $[x, y]$ and affecting the risk of progression. Incorporating this effect into (4.1) we get the dynamic Cox model

$$(4.2) \qquad\qquad F(t - x | \eta) \;=\; \exp\left\{ - \int_x^t \eta(x, y) dH(y - x) \right\}$$

with the hazard function $\eta(x, y) h(y - x)$, $h(t) = dH(t)/dt$.

Now let $\xi(x) dx$ summarize a multiplicative effect (acting on h) of exposure on the strength of the risk induced at time $[x, x + dx]$. The quantity $\xi(x)$ can be thought of as linked to the dose rate with the meaning of the amount of risk induced per unit time exposure to a certain dose in the time interval $[x, x + dx]$.

Finally, assuming the risks at different times x compete with each other. We have the host risk represented by a sum (integral) over the exposure period $x \in [0, t]$ preceding the time point t of interest

$$
\begin{aligned}
\lambda(t | \eta, \xi) \;&=\; \int_0^t \xi(x) \eta(t, x) h(t - x) dx \\
(4.3) \qquad\qquad &=\; \int_0^t \theta(s, t | z[0, t]) dH(s),
\end{aligned}
$$

where $\theta(s, t | z[0, t]) = \xi(t - s) \eta(t, t - s)$ is some parametrically specified functional of the exposure trajectory $z[0, t]$ that θ and η depend on.

Thus, we deduce the key feature of the survival model where risks of exposure are induced over time, competing, and affecting each other multiplicatively, as the hazard function represented as a linear *integral* functional of the baseline hazard with a parametrically specified kernel. Note that in the traditional Cox model, the model hazard λ is related to the baseline hazard h directly rather than λ being a cumulative form of h as in this case.

Reflecting back at the mice lymphoma study that motivated this paper, we describe the background or acutely induced cancer using the traditional time-dependent Cox model with the hazard function of the form

$$(4.4) \qquad \mu(t|z[0,t])dA_t/dt,$$

where A is some generally different baseline cumulative hazard, and μ is a generally time-dependent predictor. Time-dependent μ is needed to describe a possible dynamic "treatment" effect of exposure on background latent disease (the hormesis effect).

Assuming that the risks of different types of exposure or background are additive (that would be the case if they compete and are independent) we arrive at a generic functional transformation model serving a broad range of examples. The class is defined by the following expression for the hazard function:

$$(4.5) \qquad \lambda(t|z[0,t]) = \mu(t|z[0,t])dA_t/dt + \int_0^t \theta(s,t|z[0,t])dH_s,$$

where H_s and A_s are two nonparametrically specified baseline cumulative hazard functions at time s; $\theta(s,t)$ and $\mu(t)$ are functionals of the observed history of dynamic exposure, $z[0,t]$. Incorporated into the parametric kernels θ and μ are regression parameters β measuring sensitivity to exposure and effects of possible other covariates.

4.3 Estimation and inference

4.3.1 Data structure and notation

Suppose failure time, T, and right censoring time, C, are independent random variables. We observe a vector of time-dependent covariates, $z(t)$, through time t, with $t \in [0, \tau]$ and $\tau < \infty$, appropriately constrained to be during the study period. Let $X = min(T, C)$ and $\Delta = I(T \leq C)$. Each independent individual, i, contributes observed data $\{X_i, \Delta_i, z_i(t)\}$, $i = 1, \ldots, n$. The observed data can be represented equivalently in counting process notation as $\{N_i(t), Y_i(t), z_i(t)\}$, $i = 1, \ldots, n$, where $N_i(t) = I(X_i \leq t)\Delta_i$ gives the failure counting process and $Y_i(t) = I(X_i \geq t)$ gives the at risk process, see Nielson et al. (1992).

4.3.2 Estimating Equation

Under independent censoring assumption, the log-likelihood becomes

$$
\begin{aligned}
l \; = \; & \sum_{i=1}^{n} \int_0^{\tau} \Bigg[\log \left\{ \mu(t|z_i[0,t])dA_t + \int_0^t \theta(s,t|z_i[0,t])dH_s \right\} dN_i(t) \\
& -Y_i(t) \left\{ \mu(t|z_i[0,t])dA_t + \int_0^t \theta(s,t)dH_s \right\} \Bigg],
\end{aligned}
$$
(4.6)

where here and below the subscript i denotes individual patient level information.

Differentiating the likelihood we get the following score vector for (H_t, β), for $t \in [0, \tau]$:

$$
\begin{aligned}
U_H(H[0,\tau], A[0,\tau], t, \beta) \; = \; & \sum_{i=1}^{n} \int_{t+}^{\tau} \frac{\theta(t,x|z_i[0,x])}{\lambda(x|z_i[0,x])} dM_i(x) \\
U_A(H[0,t], A[0,t], \beta) \; = \; & \sum_{i=1}^{n} \int_0^t \mu_i(x) dM_i(x) \\
U_\beta(H[0,\tau], A[0,\tau], \beta) \; = \; & \sum_{i=1}^{n} \int_0^{\tau} \{\lambda(t|z_i[0,t])\}^{-1} \left\{ \mu_i^\beta(t)dA_t \right. \\
& \left. + \int_0^t \theta_i^\beta(s,t)dH_s \right\} dM_i(t)
\end{aligned}
$$
(4.7)

where

$$dM_i(t) = dN_i(t) - Y_i(t)\lambda(t|z_i[0,t])dt$$

and

$$\theta^\beta(s,t) = \frac{\partial}{\partial\beta}\theta(s,t)$$

$$\mu^\beta(s,t) = \frac{\partial}{\partial\beta}\mu(s,t)$$

Solutions to the estimating equations setting score to zero give $\widehat{H}_t$, $\widehat{A}_t$ and $\widehat{\beta}$.

4.3.3 Asymptotics of Estimates

In this section we show the consistency and asymptotic normality of $(\widehat{H}_t, \widehat{A}_t, \widehat{\beta})$, for one time t. We follow the ideas described in Chen (2009). The arguments are similar to those developed in Chapter III, therefore, we present the key results without going into technical details.

Since $M(t)$ is martingale with the filtration:

$$\mathcal{F}_{t-} = \sigma\{N(s), Y(s), z(s) : s \in [0,t)\}.$$

Combining terms in (4.7), we obtain

$$U_H(H[0,\tau], A[0,\tau], t, \beta) = \sum_{i=1}^n \int_{t+}^\tau \psi_i(x,t)dM_i(x)$$

$$U_\beta(H[0,\tau], A[0,\tau], \beta) = \sum_{i=1}^n \int_0^\tau \rho_i(x)dM_i(x)$$

where

$$\psi(x,t) = \frac{\theta(t,x)}{\lambda(x|z[0,x])} - \lambda^{-1}(x,|z[0,x]) \left\{ \int_{t+}^x \theta(s,x)dH_s \right\}.$$

Therefore, (4.7) is multivariate martingale if (H_t, A_t, β) are evaluated at their true values. It converges to a zero-mean Gaussian process $\tilde{W}_U(t, \beta)$ with the covariance function

$$(4.8)\ \Sigma(s,t,\beta) = \begin{bmatrix} \sigma_H^2(s,s) & \sigma_H^2(s,t) & \sigma_{H,A}^2(s,s) & 0 & \sigma_{H,\beta}^2(s,\beta) \\ \sigma_H^2(t,s) & \sigma_H^2(t,t) & 0 & \sigma_{H,A}^2(t,t) & \sigma_{H,\beta}^2(t,\beta) \\ \sigma_A^2(s,s) & \sigma_A^2(s,t) & \sigma_{H,A}^2(s,s) & 0 & \sigma_{A,\beta}^2(s,\beta) \\ \sigma_A^2(t,s) & \sigma_A^2(t,t) & 0 & \sigma_{H,A}^2(t,t) & \sigma_{A,\beta}^2(t,\beta) \\ \sigma_{H,\beta}^2(s,\beta) & \sigma_{H,\beta}^2(t,\beta) & \sigma_{A,\beta}^2(s,\beta) & \sigma_{A,\beta}^2(t,\beta) & \sigma_\beta^2(\beta) \end{bmatrix},$$

where the forms of each component in $\Sigma(s,t,\beta)$ are given in Appendix C.1.

Next we outline the study of weak convergence of $\sqrt{n}(\widehat{H}_t - H_t, \widehat{A}_t - A_t, \widehat{\beta} - \beta_0)$.

Taylor expansion gives (ignoring $o_p(1)$ terms):

$$
n^{-1/2}
\begin{bmatrix}
U_H(\widehat{H}[0,\tau], \widehat{A}[0,\tau], t, \widehat{\beta}) - U_H(H[0,\tau], A[0,\tau], t, \beta) \\[6pt]
U_A(\widehat{H}[0,t], \widehat{A}[0,t], \widehat{\beta}) - U_A(H[0,t], A[0,t], \beta) \\[6pt]
U_\beta(\widehat{H}[0,\tau], \widehat{A}[0,\tau], \widehat{\beta}) - U_\beta(H[0,\tau], A[0,\tau], \beta)
\end{bmatrix}
$$

$$
= \sqrt{n}
\begin{bmatrix}
\begin{array}{c}
\int_0^t n^{-1} U_H^H(H[0,\tau], A[0,\tau], t, y, \beta) d(\widehat{H} - H)(y) \\[4pt]
+ \int_0^t n^{-1} U_H^A(H[0,\tau], A[0,\tau], t, y, \beta) d(\widehat{A} - A)(y) \\[4pt]
+ n^{-1} U_H^\beta(H[0,\tau], A[0,\tau], t, \beta)(\widehat{\beta} - \beta) \\[8pt]
\int_0^t n^{-1} U_A^H(H[0,\tau], A[0,\tau], t, y, \beta) d(\widehat{H} - H)(y) \\[4pt]
+ \int_0^t n^{-1} U_A^A(H[0,\tau], A[0,\tau], t, y, \beta) d(\widehat{A} - A)(y) \\[4pt]
+ n^{-1} U_A^\beta(H[0,\tau], A[0,\tau], t, \beta)(\widehat{\beta} - \beta) \\[8pt]
\int_0^\tau n^{-1} U_\beta^H(H[0,\tau], A[0,\tau], y, \beta) d(\widehat{H} - H)(y) \\[4pt]
+ \int_0^\tau n^{-1} U_\beta^A(H[0,\tau], A[0,\tau], y, \beta) d(\widehat{A} - A)(y) \\[4pt]
+ n^{-1} U_\beta^\beta(H[0,\tau], A[0,\tau], \beta)(\widehat{\beta} - \beta)
\end{array}
\end{bmatrix}
$$

$$
(4.9) \qquad = \sqrt{n}
\begin{bmatrix}
\begin{array}{c}
\int_0^t \phi_H^H(H[0,\tau], A[0,\tau], t, y, \beta) d(\widehat{H} - H)(y) \\[4pt]
+ \int_0^t \phi_H^A(H[0,\tau], A[0,\tau], t, y, \beta) d(\widehat{A} - A)(y) \\[4pt]
+ \phi_H^\beta(H[0,\tau], A[0,\tau], t, \beta)(\widehat{\beta} - \beta) \\[8pt]
\int_0^t \phi_A^H(H[0,\tau], A[0,\tau], t, y, \beta) d(\widehat{H} - H)(y) \\[4pt]
+ \int_0^t \phi_A^A(H[0,\tau], A[0,\tau], t, y, \beta) d(\widehat{A} - A)(y) \\[4pt]
+ \phi_A^\beta(H[0,\tau], A[0,\tau], t, \beta)(\widehat{\beta} - \beta) \\[8pt]
\int_0^\tau \phi_\beta^H(H[0,\tau], A[0,\tau], y, \beta) d(\widehat{H} - H)(y) \\[4pt]
+ \int_0^\tau \phi_\beta^A(H[0,\tau], A[0,\tau], y, \beta) d(\widehat{A} - A)(y) \\[4pt]
+ \phi_\beta^\beta(H[0,\tau], A[0,\tau], \beta)(\widehat{\beta} - \beta)
\end{array}
\end{bmatrix} .
$$

Then we have

$$(4.10) \quad \sqrt{n} \left[\begin{array}{c} \int_0^t \phi_H^H(H[0,\tau], A[0,\tau], t, y, \beta) d(\widehat{H} - H)(y) \\ + \int_0^t \phi_H^A(H[0,\tau], A[0,\tau], t, y, \beta) d(\widehat{A} - A)(y) r \\ + \phi_H^\beta(H[0,\tau], A[0,\tau], t, \beta)(\widehat{\beta} - \beta) \\ \int_0^t \phi_A^H(H[0,\tau], A[0,\tau], t, y, \beta) d(\widehat{H} - H)(y) \\ + \int_0^t \phi_A^A(H[0,\tau], A[0,\tau], t, y, \beta) d(\widehat{A} - A)(y) \\ + \phi_A^\beta(H[0,\tau], A[0,\tau], t, \beta)(\widehat{\beta} - \beta) \\ \int_0^\tau \phi_\beta^H(H[0,\tau], A[0,\tau], y, \beta) d(\widehat{H} - H)(y) \\ + \int_0^\tau \phi_\beta^A(H[0,\tau], A[0,\tau], y, \beta) d(\widehat{A} - A)(y) \\ + \phi_\beta^\beta(H[0,\tau], A[0,\tau], \beta)(\widehat{\beta} - \beta) \end{array} \right] \Rightarrow \tilde{W}_U(t, \beta).$$

Let $V_H(y) = (\widehat{H} - H)y$, $V_A(y) = (\widehat{A} - A)y$ and let $W_1(t)$, $W_2(t)$ and W_3 be the following:

$$\begin{aligned} W_1(t) &= \int_0^t \phi_H^H(H[0,\tau], A[0,\tau], t, y, \beta) dV_H(y) \\ &+ \int_0^t \phi_H^A(H[0,\tau], A[0,\tau], t, y, \beta) dV_A(y) \\ &+ \phi_H^\beta(H[0,\tau], A[0,\tau], t, \beta)(\widehat{\beta} - \beta) \\ W_2(t) &= \int_0^t \phi_A^H(H[0,\tau], A[0,\tau], t, y, \beta) dV_H(y) \\ &+ \int_0^t \phi_A^A(H[0,\tau], A[0,\tau], t, y, \beta) dV_A(y) \\ &+ \phi_A^\beta(H[0,\tau], A[0,\tau], t, \beta)(\widehat{\beta} - \beta) \\ W_3 &= \int_0^\tau \phi_\beta^H(H[0,\tau], A[0,\tau], y, \beta) dV_H(y) \\ &+ \int_0^\tau \phi_\beta^A(H[0,\tau], A[0,\tau], y, \beta) dV_A(y) \\ &+ \phi_\beta^\beta(H[0,\tau], A[0,\tau], \beta)(\widehat{\beta} - \beta). \end{aligned}$$

The covariance function of left hand side of (4.10), denoted as $\Sigma^*(s, t, \beta)$, is equivalent to (4.8). The covariance functions of $\{dW_1(t), dW_2(t), W_3\}$ can be found as solutions to

the system of Volterra integral equations expressing this equivalence.

The associated variances can be estimated using the Hessian information matrix, similar to Chapter III. In practice, it is common to numerically solve for $\widehat{H}$, $\widehat{A}$ and $\widehat{\beta}$ in estimating equations like (4.7). Similarly, it is common to numerically compute second order derivatives based on the full likelihood when calculating variances. This is the approach we adopt in Section 4.4 for the Mouse study.

4.4 Mouse Study

We now revisit the mouse study described in Section 4.1. Let $\lambda_c(t)$ denote the hazard for the control group. Let $\lambda_a(t)$ represent the hazard for acute dosing group. And let $\lambda_p(t)$ denote the hazard for the protracted group.

In agreement with the general modeling of Section 4.2, the components of the hazard are specified as follows with each line representing the respective group in each subclass of exposure (Background, $\mathcal{A}$, and induced, $\mathcal{B}$).

$\mathcal{A}$. *Background hazard:*

(4.11)
$$\text{Control } \lambda_c^B(t) = dA_t/dt$$

(4.12)
$$\text{Protracted dosing } \lambda_p^B(t) = \exp\{-\beta(t \wedge 252)\}dA_t/dt$$

(4.13)
$$\text{Acute dosing } \lambda_a^B(t) = \eta dA_t/dt$$

$\mathcal{B}$. *Induced hazard:*

(4.14)
$$\text{Control } \lambda_c^I(t) = 0$$

(4.15)
$$\text{Protracted } \lambda_p^I(t) = \int_0^{252 \wedge t} \exp\{-\alpha(t-x)\}dH_x$$

(4.16)
$$\text{Acute } \lambda_a^I(t) = \gamma dH_t,$$

The hazard models can be summerized as in Table 4.1.

Table 4.1: Hazard models for three dosing groups, background and induced.

Group	Background	Induced
Acute	$\eta dA_t/dt$	γdH_t
Protracted	$e^{-\beta(t\wedge 252)}dA_t/dt$	$\int_0^t e^{-\alpha(t-x)}dH_x$
Control	dA_t/dt	0

Background and induced hazard components are summed up for each group, respectively. For example, the hazard for protracted dosing group is the sum of $e^{-\beta(t\wedge 252)}dA_t/dt$, the component at background process, and $\int_0^t e^{-\alpha(t-x)}dH_x$, the component on the induced process. Baseline hazard H is used to model the risk of latent tumor in induced process while $\mathcal{A}$ is reserved for background process. There is no induced hazard in the control group, hence the corresponding component is set to 0. The beneficial effect, or treatment effect, of the radiation on background cancer for acute dosing group is the hormesis effect we observe. This is the traditional Cox hazard model since only one dose at baseline is involved. The harmful effect of the same dose that induces new lesions is also modeled by a Cox component. The time-dependent component of background cancer in the protracted dosing group, $e^{-\beta(t\wedge 252)}dA_t/dt$, models the "treatment" effect of protracted exposure on the background cancer. Induced cancer hazard function in this group is represented by an integral form of the baseline hazard modeling accumulating harmful effects of exposure over time, as it is the combination of the independent competing risks that originated at different (continuous) times of exposure.

Shown in Figure 4.2 are observed (Kaplan-Meier) and model-predicted survival curves. Overall, the agreement is very good.

The estimates for the model parameters together with the corresponding 95% confidence intervals are presented in Table 4.2. The algorithm converges when $\widehat{\gamma} = 0$, hence
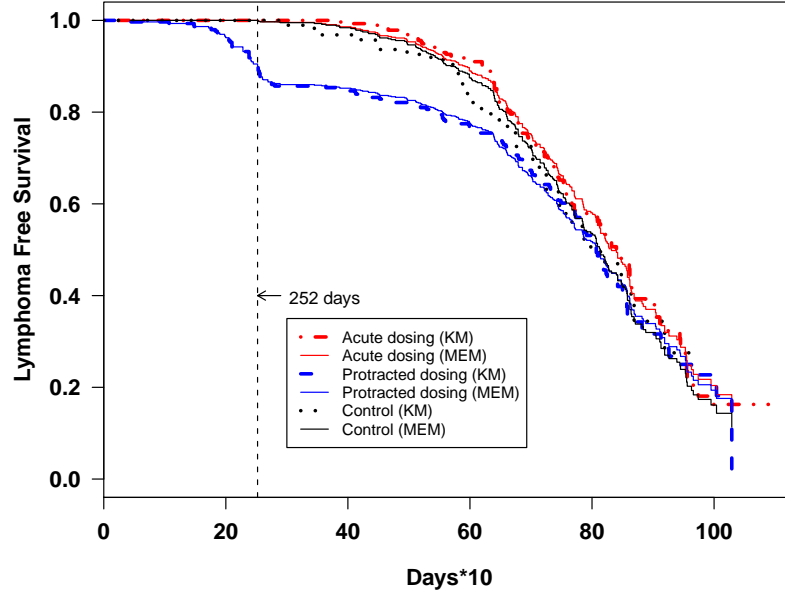
Figure 4.2: Lymphoma Free Survivals by Groups - Predicted vs. Observed

we do not give the associated 95% confidence interval and p-value.

$\widehat{\eta}$ gives the estimate of hormesis effect. That is, the acute dosing group has lower hazard (0.872, 95% CI: 0.684, 1.110, p-value: 0.133) than control group (null hypothesis is $\eta = 1$), resulting higher survival at the beginning of study period. The radiation effect is protracted dosing group is more complicated. The treatment effect of the radiation dose at the background $\widehat{\beta} = 0.008$, (95% CI: 0.001, 0.014, p-value: 0.008) seems to be dominated by the harmful effect on the induced process, where $\widehat{\alpha} = 0.822$, (95% CI: 0.418, 1.226, p-value $<$0.0001) stands for the unit dose harmful effect and overall harmful effect should be integrated over the radiation period.

4.5 Discussion

In this work, we propose a mechanistic model that accounts for complexity of radiation on tumor growth. The model reproduces the diversity of hormesis effect. The rapid decline

Table 4.2: Mouse study model fitting results: model parameter estimates for incidence model.

Group	Background			Induced		
	Estimate[1]	95% CI[2]	p-value	Estimate[1]	95% CI[2]	p-value
Acute	$\widehat{\eta} = 0.872$	0.684, 1.110	0.133	$\widehat{\gamma} = 0$	NA	NA
Protracted	$\widehat{\beta} = 0.008$	0.001, 0.014	0.008	$\widehat{\alpha} = 0.822$	0.418, 1.226	<0.0001

1. Statistical significance is defined when 95% CI does not cover the true values with Type I error rate 5%.

of survival in $[0, 252]$ for protracted dosing group implies that the latent period for lymphoma of the induced process is very short. This can also be confirmed by the observed improved survival soon after time=0 for acute dosing group, where a single radiation dose kills background more than it induces lymphoma, or in another word, the induced lymphoma has such a short latent period that it is killed by the same dose before it surfaces. When the radiation stops, both acute dosing group and protracted dosing group seem to behave similarly as the control group. This indicates that, in the long run, the system is governed by the background process.

In recent years, transformation models have been studied to evaluate the relationship between dynamic risk factors and survival outcomes when Cox PH assumption does not hold. The method we propose can mechanistically model the effect of the radiation in mouse study that demonstrates hormesis effect. We develop a specific transformation model class that is capable to reproduce the diversity of hormesis effect caused by radiation. We provide rigorous theoretical justifications for such models.

The method developed in this chapter will be applicable to many situations when exposure is time changing and effects are complicated. Mice data used as a test for the methodology to be developed in this project are only one example of such data. With this project we are laying the ground work to be able to model any situation where the risk and exposure is time-changing, and the disease has a period of latent development before diagnosis.

The modeling component in (4.5) that models the protracted radiation effect (combination of harmful and beneficial effect) is considered as in integral form since the dose is continuous. However, in some situations, the exposures are not necessarily continuous. For instance, in recent years, as the rapid development of therapy, many cancer patients have better survival and longer intervals between cancer relapse. This creates the discrete

experience for those patients to receive treatments through out their lives. The opportunity of having multiple treatments increases. Our method can be modified for such discretized treatment (exposures) easily by modifying the accumulating risk component in our incidence model as a summation of risks generated at discretized times when the treatment takes place.

Some other examples where the model could be used are: in clinical trials, patients receive drugs on continuous basis, however, when drug cures the disease, its side effect also can be counteracting the treatment effect; bone marrow transplant patients experiencing abnormal platelets, effects of air pollution on pulmonary disorders and other environmental exposure problems.

Though the method developed in this work assumes exposure is continuous, the extension to discretized exposure could also be studied. In modern medical practice, patients with chronic disease who experience relapse usually received multiple treatments during their lifetimes. As the quality of therapy improves, the interval for patients to receive next treatment during relapse is getting longer. This results in discretized exposure (treatment). The incidence model that takes into consideration of the effect of continuous and accumulated radiation, could be modified to account for those situations. A good example is Hodgkin's disease.

CHAPTER V

Conclusion and future work

This thesis explores two topics in semiparametric survival analysis area, 1) prediction of patients' lifetimes based on their risk profiles; (2) estimation of dynamic exposure effects on survival outcomes among patients with chronic disease.

In Chapter II, we develop restricted mean MI method that has several advantages. First, we can retain patients characteristics when making imputation choices, therefore improve precision and efficiency in estimation and inference procedures. This can be observed in both independent censoring and dependent censoring scenarios. Due to the utilization of covariate information, the method can also correct the bias caused by dependent censoring. Second, the MI method does not require assumptions on hazard distribution, so that we bypass the challenge that authors should consider when they use Cox modeling approach. We also maintain the traditional advantages of MI method, for example, we make the censoring data analysis into complete data analysis and standard statistics analytical tools are widely available.

The method serves as a base for MI choices when dependent censoring is caused by a time-changing factor. It also will be modified to impute Quality-of-Life outcomes, based on observed information. I am one of the co-authors in both of these projects.

In Chapter III, we lay a foundation for statistical analysis of any dynamic systems

where observed covariate z_t dynamically affect the latent process that links to the response of the system. The method may well be suited for many cancer research that usually involves tumor growth process before diagnosis. The method can also be extended to non-survival outcomes. For instance, the imputation framework developed for a stochastic process that is informed by some external dynamic factors can be used to describe the status of a system. For instance, if we observe a changing PSA score for a prostate cancer patient, it may well be associated with the underlying tumor growth process that links to the disease status of a patient. In this case, the imputed frailty, given the dynamic PSA score, could be interpreted as how sick the patient is at the time of observed. The imputation of a process can also be applied in fields other than medical science. For example, in financial market, people observe stock indices change constantly over time, and they are interested in predicting the future strength of the economy based on those changing factors. Then the driving force of the economical change can be seen as a latent process and we can make similar predictions as described in this work.

Though the method proposed is designated in the univariate survival setting, the extension to multivariate survival setting is possible. A preliminary thought about Laplace functional for a bivariate survival setting would be $E_{U[0,t]}(e^{-\int\int U_s dH_s^1 dH_s^2})$, where H_s^1 and H_s^2 correspond to the baseline cumulative hazard functions of the bivariate survival.

In Chapter IV, we propose a method that can mechanistically model the complex effect of the radiation. The advantage of our modeling approach is that we are able to reproduce the diversity of hormesis effect, while retaining the power of rigorous theoretical justifications for such models. The method developed in this chapter will be applicable to many situations when exposure is time changing, effects are complicated, and the disease has a period of latent development before diagnosis.

Extensions to discretized exposures can be easily made. A good example is Hodgkin's

disease. In modern medical practice, patients with chronic disease who experience relapse usually received multiple treatments during their lifetimes. As the quality of therapy improves, the interval for patients to receive next treatment during relapse is getting longer. This results in discretized exposure (treatment). The incidence model that takes into consideration of the effect of continuous and accumulated radiation, could be modified to account for those situations.

Some other examples where the model could be used are: in clinical trials, patients receive drugs on continuous basis, however, when drug cures the disease, its side effect also can be counteracting the treatment effect; bone marrow transplant patients experiencing abnormal platelets, effects of air pollution on pulmonary disorders and other environmental exposure problems.

APPENDICES

APPENDIX A

Appendices for Chapter II

A.1 EM algorithm in restricted mean MI procedure

We assume the mean structure as in 2.1, but keep the distribution of Y otherwise unspecified. The goal is to predict each censored individual's mean restricted lifetime based on the converged $\widehat{\beta}^{(k)T} Z_i$ values.

We calculate $\widehat{\beta}^{(0)T} Z_i$ as the initial expected value of Y_i based on covariate profile Z_i. We use $\widehat{\beta}^{(0)}$ to estimate the initial expected value of Y_i based on $\widehat{\beta}^{(0)T} Z_i$ for each censored individual i.

To get the next iterated estimate, $\widehat{\beta}^{(1)}$, we sample from $N\{\widehat{\beta}^{(0)T} Z_i, Var(\widehat{\beta}^{(0)T} Z_i)\}$ for each patient censored at $X_i = C_i$. We retain sample values that are greater than $\log C_i$, and then use the sample average to estimate $E(Y_i|C_i, Z_i)$, our E-step of the pseudo EM algorithm. After the 'E' step is completed for each censored individual, this completed dataset is used to calculate $\widehat{\beta}^{(1)}$ under (2.1), our M-step. We repeat 'E' and 'M' steps until the convergence of $\widehat{\beta}^{(k)}$.

A.2 Residual sampling procedure in restricted mean MI algorithm

In order to generate an augmented dataset, we append an appropriate residual to predicted mean, $\widehat{E}(Y_i|C_i, Z_i)$, for each censored patient i. We achieve this goal by creating

an imputing pool with patients whose categorical covariates match Z_i, and whose $\widehat{E}(Y|Z)$ fall within some small $b-$margin of $\widehat{E}(Y_i|Z_i)$ [to account for similarity in continuous co-variates]. Once we form the imputing pool, we estimate the pool's survival curve using the Kaplan Meier method. Then we randomly select a probability, s_i, from a Uniform(a,1) distribution, where a is the minimum survival probability in this pool. Next we select a failure time where the estimated survival probability from the KM curve equals to s_i. The corresponding residual is taken from the patient with the observed failure time that was selected. This is then the error term to be appended to $\widehat{E}(Y_i|C_i, Z_i)$. This approach is similar to the nearest neighborhood approach in Hsu et al. (2006) modified for the restricted mean setting.

A.3 Nonparametric Estimate for $E\left\{\log min(\tau, T)\right\}$

Suppose random samples $T_1, ..., T_n$ come from a non negative random variable, T, with survival function, S_T, and $C_1, ..., C_n$ come from a random variable, C, that is independent of T. Let $X_i = min(T_i, C_i), i = 1, ..., n$ be the observed times to event. Suppose $Y = g\{min(\tau, T)\}$ with τ a fixed positive constant, and g some strictly monotone function with associated inverse $g^{(-1)}$ that satisfies:

$$g \quad : \quad [0, +\infty) \rightarrow (-\infty, +\infty)$$

$$g^{-1} \quad : \quad (-\infty, +\infty) \rightarrow [0, +\infty)$$

Without loss of generality, assume g to be monotone increasing. Then

$$\widehat{\mu} = \int_{X_{(1)}}^{g^{(-1)}(0)} \{\widehat{P}(T > t) - 1\} dg(t) + \int_{g^{(-1)}(0)}^{\tau} \widehat{P}(T > t) dg(t) + \widehat{P}(T > \tau) g(\tau)$$

is a consistent estimator of $E(Y)$, where $\widehat{P}(T > t)$ is Kaplan Meier estimator for $P(T > t)$.

Proof.

For $Y = g\{min(\tau, T)\}$, the distribution can be characterized as:

$$
\begin{aligned}
F_Y(y) &= P[g\{min(\tau, T)\} \leq y]I\{y < g(\tau)\} + P[g\{min(\tau, T)\} \leq y]I\{y \geq g(\tau)\} \\
&= F_T\{g^{(-1)}(y)\}I\{y < g(\tau)\} + I\{y \geq g(\tau)\}.
\end{aligned}
$$

The random variable, Y, has a range of $(-\infty, g(\tau)]$, with a discontinuity point of jump size $dF_Y\{g(\tau)\} = F_Y\{g(\tau)\} - F_Y\{g(\tau^-)\} = S_T(\tau)$, at $g(\tau)$.

$E(Y)$ can be written as:

$$
\mu = \int_0^{g(\infty)} \{P(Y > -y) - 1\}dy + \int_0^{g(\tau)} P(Y > y)dy + dF_Y\{g(\tau)\} \times g(\tau).
$$

where $dF_Y\{g(\tau)\} = F_Y\{g(\tau)\} - F_Y\{g(\tau)^-\} = S_T(\tau)$. Denote the three terms on the right hand side of the equation and corresponding estimates as μ_i and $\widehat{\mu}_i, i = 1, 2, 3$ respectively.

For μ_1, it can be shown that:

$$
\begin{aligned}
\mu_1 &= \int_0^{g(\infty)} \{P(Y > -y) - 1\}dy \\
&= \int_0^{g(\infty)} [P\{g(T) > -y)\} - 1]dy \\
&= \int_0^{g(\infty)} [P(T > g^{-1}(-y)) - 1]dy \\
&= \int_{g^{(-1)}\{-g(\infty)\}}^{g^{(-1)}(0)} \{P(T > t) - 1\}dg(t),
\end{aligned}
$$

where the last line is obtained by the transformation $t = g^{-1}(-y), y = -g(t)$. To obtain a consistent estimate of μ_1, we replace $P(T > t)$ with its corresponding Kaplan-Meier estimates, and change the lower bound of the integral to $X_{(1)}$, the first observed failure time. This is because $\widehat{P}(T > t)$ remains 1 until $t = X_{(1)}$ and monotone transformation preserves the ranks of observed event times. Hence $\widehat{\mu}_1$ is: $\int_{X_{(1)}}^{g^{(-1)}(0)} \{\widehat{P}(T > t) - 1\}dg(t)$.

Similarly, for μ_2, we have

$$
\begin{aligned}
\mu_2 &= \int_0^{g(\tau)} P(Y > y) dy \\
&= \int_0^{g(\tau)} [P\{g(T) > y)\}] dy \\
&= \int_0^{g(\tau)} P\{T > g^{-1}(y)\} dy \\
&= \int_{g^{(-1)}(0)}^{\tau} P(T > t) dg(t),
\end{aligned}
$$

where the last line is obtained via the transformation $t = g^{-1}(y), y = g(t)$. The corresponding consistent estimate $\widehat{\mu}_2$ becomes $\int_{g^{(-1)}(0)}^{\tau} \widehat{P}(T > t) dg(t)$.

For the third term $\mu_3 = dF_Y\{g(\tau)\} \times g(\tau)$, with $dF_Y\{g(\tau)\} = S_T(\tau)$, the consistent estimate is $\widehat{P}(T > \tau)g(\tau)$.

Combining $\widehat{\mu}_i, i = 1, 2, 3$, we obtain a consistent estimate of μ. Particularly, if g is the log function, we obtain:

$$
\widehat{E}(Y) = \int_{X_{(1)}}^1 \{\widehat{P}(T > t) - 1\} d\log t + \int_1^{\tau} \widehat{P}(T > t) d\log t + \widehat{P}(T > \tau) \times \log \tau,
$$

APPENDIX B

Appendices for Chapter III

B.1 Expectaion $E(U_x|X, \Delta, z[0,t])$

Based on the survival function in (3.2) and the data structure described in Section 3.3, we derive that $E(U_x|X, \Delta, , z[0,t])$, that is, the conditional expectation of U_x given full knowledge of the individual's failure and censoring information.

First, let $\Delta = 0$. Then we have

$$
\begin{aligned}
E(U_x|X = t, \Delta = 0, z[0,t]) &= \int U_x f(U_x|X = t, \Delta = 0, z[0,t]) dU_x \\
&= \frac{\int U_x f(C = t|U_x, z[0,t]) P(T > t|U_x, z[0,t]) f(U_x) dU_x}{f(C = t|z[0,t]) S(t|z[0,t])} \\
&= \frac{\int U_x P(T > t|U_x = u, z[0,t]) f(U_x) dU_x}{S(t|z[0,t])}
\end{aligned}
$$

which becomes, based on (3.2)

$$
\frac{\int U_x e^{-\int_0^t U_s dH_s} f(U_x) dU_x}{\int e^{-\int_0^t U_s dH_s} f(U_x) dU_x}.
$$

Finally, we have

$$
E(U_x|X = t, \Delta = 0, z[0,t]) = \frac{E_{U[0,t]}\left\{ U_x \exp(-\int_0^t U_s dH_s) \right\}}{E_{U[0,t]}\left\{ \exp(-\int_0^t U_s dH_s) \right\}}.
$$

The imputed frailty, $\Theta^0(t, H[0,t]|z[0,t])$ as in (3.4), is a special case of the above quantity when $x = t$.

Now let $\Delta = 1$, then we have

$$
\begin{aligned}
E(U_x|X = t, \Delta = 1, z[0,t]) &= \int U_x f(U_x|X = t, \Delta = 1, z[0,t])dU_x \\
&= \frac{\int U_x P(C > t|U_x, z[0,t])f(T = t|U_x, z[0,t])f(U_x)dU_x}{P(C > t|z[0,t])f(T = t|z[0,t])} \\
&= \frac{\int U_x f(T = t|U_x, z[0,t])f(U_x)du}{f(T = t|z[0,t])}
\end{aligned}
$$

which becomes, based on (3.2) and the relationship $-\frac{d}{dt}S(t|z[0,t]) = f(T = t|z[0,t])$

$$
\frac{\int U_x U_t e^{-\int_0^t U_s dH_s} f(U_x)dU_x}{\int U_t e^{-\int_0^t U_s dH_s} f(U_x)dU_x}.
$$

Finally, we have

$$
E(U_x|X = t, \Delta = 1, z[0,t]) = \frac{E_{U[0,t]}\left\{U_x U_t \exp(-\int_0^t U_s dH_s)\right\}}{E_{U[0,t]}\left\{U_t \exp(-\int_0^t U_s dH_s)\right\}}.
$$

This reduces to the impute for Kosorok et al. (2004) hazard model, as we described in Section 3.1, in the special case where $U_s = U e^{\beta^T z_s}$.

B.2 Regularity conditions

(a) The true function H is strictly increasing and absolutely continuous, with differentials exist everywhere.

(b) True values of parameter vector β_0 are in the interior of a compact Euclidean space.

(c) The covariate process $Z(t)$ is left continuous with total bounded variation within $[0, \tau]$.

(d) For at risk process $Y(t)$, $P(Y(\tau) = 1|Z_t) > 0$.

(e) The at risk set will not shrink to empty, i.e., when $T = \tau$, $P(censoring|Z_t) > 0$.

(f) The posterior frailty Θ and its derivatives with respect to dH and β are bounded within $[0, \tau]$.

(g) Hessian matrix I evaluated at true values of H and β is positive definite.

B.3 Property of martingale transform $\int_0^\tau \xi(u,t)dM(u)$

We denote process $\int_0^\tau \xi(u,t)dM(u)$ as $V(t)$. To see when $V(t)$ is a martingale, we only need to check when $E\{dV(t)|\mathcal{F}_{t-}\} = 0$, where $\mathcal{F}_{t-} = \sigma\{N(s), Y(s), z(s) : s \in [0,t)\}$ is the same filtration of martingale M,.

First, we notice that:

$$
\begin{aligned}
dV(t) &= \int_0^\tau \{\xi(u, t + dt)dM(u) - \xi(u,t)dM(u)\} \\
&= \int_0^\tau \{\xi'(u,t)dt\}\, dM(u),
\end{aligned}
$$

where $\xi'(u,t)$ denotes the partial derivative of ξ with respect to t.

Next, if we assume the boundness and smoothness of deterministc function ξ, we have:

$$
\begin{aligned}
E\{dV(t)|\mathcal{F}_{t-}\} &= \int_0^\tau E\{\xi'(u,t)dtdM(u)|\mathcal{F}_{t-}\} \\
&= dt \int_0^\tau \xi'(u,t)E\{dM(u)|\mathcal{F}_{t-}\} \\
&= \begin{cases} 0 & \text{if } u \geq t \\ dt \int_0^\tau \xi'(u,t)dM(u) & \text{if } u < t \end{cases}
\end{aligned}
$$

Furthermore, $\int_0^\tau \xi'(u,t)dtdM(u)$ will be zero only if $\xi'(u,t) = 0$. That is, when $u < t$ and $\xi(u,t)$ does not depend on t (giving $\xi'(u,t) = 0$), $V(t)$ will be a martingale. This establishes the martingale property of the score.

Similarly, for a process with the form $V^*(t) = \int_t^\tau \xi(u,t)dM(u)$, u always lies in the future of t, hence $E\{dV^*(t)|\mathcal{F}_{t-}\} = 0$.

B.4 Covariance for martingale $\int_0^\tau \xi(u,t)dM(u)$

Suppose M is a $\mathcal{F}$-martingale and the corresponding counting process has compensator A. Let $V(t) = \int_0^\tau \xi(u,t)dM(u)$. Let $\xi(u)$ be a funciton of u only, and in the special

case of this paper, $\xi(u) = 1 + \int_0^u \frac{\dot{\Theta}(u,H[0,u],x|z_i[0,u])}{\Theta^0(u,H[0,u]|z_i[0,u])} dH_x$. So $V(t) = \int_0^t \xi(u)dM(u) + \int_t^\tau \xi(u,t)dM(u)$. Both $\xi(u)$ and $\xi(u,t)$ are predictable given filtration with respect to either u or t, giving the variance seen in (3.17) for (3.16).

B.5   Second order Derivatives of estimating equations

The estimating equations are given in (3.14). The second order derivatives of equations include the following four terms:

$$
\begin{aligned}
U_H^H(H[0,\tau],t,y,\beta) &= \frac{\partial}{\partial dH_y} U_H(H[0,\tau],t,\beta) \\
&= \sum_{i=1}^n -Y_i(y)\Theta^0(y,H[0,y]|z_i[0,y]) \\
&\quad - \int_{y+}^t Y_i(u)\dot{\Theta}(u,H[0,u],y|z_i[0,u])dH_u \\
&\quad + \int_{y+}^t \left\{ \frac{\dot{\Theta}(u,H[0,u],y|z_i[0,u])}{\Theta^0(u,H[0,u]|z_i[0,u])} dN_i(u) \right. \\
&\qquad \left. -Y_i(u)\dot{\Theta}(u,H[0,u],y|z_i[0,u])dH_u \right\} \\
&\quad + dH_y \int_{y+}^\tau \left(\left[\frac{\ddot{\Theta}(u,H[0,u],y,y|z_i[0,u])}{\Theta^0(u,H[0,u]|z_i[0,u])} \right. \right. \\
&\qquad \left. -\left\{ \frac{\dot{\Theta}(u,H[0,u],y|z_i[0,u])}{\Theta^0(u,H[0,u]|z_i[0,u])} \right\}^2 \right] dN_i(u) \\
&\qquad \left. -Y_i(u)\ddot{\Theta}(u,H[0,u],y,y|z_i[0,u])dH_u \right) \\
&\quad + \int_{y+}^t \int_{x+}^\tau \left[\left\{ \frac{\ddot{\Theta}(u,H[0,u],x,y|z_i[0,u])}{\Theta^0(u,H[0,u]|z_i[0,u])} \right. \right. \\
&\qquad \left. - \frac{\dot{\Theta}(u,H[0,u],x|z_i[0,u])\dot{\Theta}(u,H[0,u],y|z_i[0,u])}{\{\Theta^0(u,H[0,u]|z_i[0,u])\}^2} \right\} dN_i(u) \\
&\qquad \left. -Y_i(u)\ddot{\Theta}(u,H[0,u],x,y|z_i[0,u])dH_u \right] dH_x \\
&\quad - \int_0^t Y_i(y)\dot{\Theta}(y,H[0,y],x|z_y^i)dH_x,
\end{aligned}
$$

$$
\begin{aligned}
U_H^\beta(H[0,\tau],t,\beta) &= \frac{\partial}{\partial\beta}U_H(H[0,\tau],t,\beta) \\
&= \sum_{i=1}^n \int_0^t -Y_i(x)\Theta^\beta(y,H[0,y]|z_y^i)dH_x \\
&\quad +dH_x\int_{x+}^\tau \left(\left[\frac{\dot\Theta^\beta(u,H[0,u],x|z_i[0,u])}{\Theta^0(u,H[0,u]|z_i[0,u])}\right.\right. \\
&\quad \left.\left. -\frac{\dot\Theta(u,H[0,u],x|z_i[0,u])\Theta^\beta(u,H[0,u]|z_i[0,u])}{\{\Theta^0(u,H[0,u]|z_i[0,u])\}^2}\right]dN_i(u)\right. \\
&\quad \left. -Y_i(u)\dot\Theta^\beta(u,H[0,u],t_j|z_i[0,u])dH_u\right) \\[2mm]
U_\beta^H(H[0,\tau],\beta) &= \frac{\partial}{\partial dH_y}U_\beta(H[0,\tau],\beta) \\
&= \sum_{i=1}^n \int_{y+}^\tau \left(\left[\frac{\Theta^{\beta,y}(x,H[0,x]|z_i[0,x])}{\Theta^0(x,H[0,x]|z_i[0,x])}\right.\right. \\
&\quad \left.\left. -\frac{\Theta^\beta(x,H[0,x]|z_i[0,x])\dot\Theta(x,H[0,x],y|z_i[0,x])}{\{\Theta^0(x,H[0,x]|z_i[0,x])\}^2}\right]dN_i(x)\right. \\
&\quad \left. -Y_i(x)\Theta^{\beta,y}(x,H[0,x]|z_i[0,x])dH_x\right) \\
&\quad -\sum_{i=1}^n Y_i(y)\Theta^\beta(y,H[0,y]|z_y^i) \\[2mm]
U_\beta^\beta(H[0,\tau],\beta) &= \frac{\partial}{\partial\beta}U_\beta(H[0,\tau],\beta) \\
&= \sum_{i=1}^n \int_0^\tau \left[\left\{\frac{\Theta^{\beta,\beta}(x,H[0,x]|z_i[0,x])}{\Theta^0(x,H[0,x]|z_i[0,x])}\right.\right. \\
&\quad \left.\left. -\left\{\frac{\Theta^\beta(x,H[0,x]|z_i[0,x])}{\Theta^0(x,H[0,x]|z_i[0,x])}\right\}^2\right\}dN_i(x)\right. \\
&\quad \left. -Y_i(x)\Theta^{\beta,\beta}(x,H[0,x]|z_i[0,x])dH_x\right],
\end{aligned}
$$

where

$$\ddot{\Theta}(u, H[0,u], x, y|z_i[0,u]) = E(U_u U_x U_y|e^{\int_0^u U_s dH_s})$$

$$-E(U_u U_x|e^{\int_0^u U_s dH_s})\Theta(y, H[0,u]|z_i[0,u]))$$

$$+\dot{\Theta}(u, H[0,u], x|z_i[0,u])\Theta(y, H[0,u]|z_i[0,u])$$

$$+\Theta(u, H[0,u]|z_i[0,u])\dot{\Theta}(y, H[0,u], t_k|z_i[0,u])$$

$$\dot{\Theta}^{\beta}(u, H[0,u], x|z_i[0,u]) = \frac{\partial}{\partial\beta}\dot{\Theta}(u, H[0,u], x|z_i[0,u])$$

$$\Theta^{\beta,y}(u, H[0,u]|z_i[0,u]) = \frac{\partial}{\partial dH_y}\Theta^{\beta}(u, H[0,u]|z_i[0,u])$$

$$\Theta^{\beta,\beta}(u, H[0,u]|z_i[0,u]) = \frac{\partial^2}{\partial\beta\partial\beta^T}\Theta(u, H[0,u]|z_i[0,u]).$$

When scaled by n, those derivatives will converge in probability to some deterministic functions that can be expressed as follows:

$$n^{-1}U_H^H(H[0,\tau], t, y, \beta) \xrightarrow{P} \phi_H^H(H[0,\tau], t, y, \beta)$$

$$n^{-1}U_H^{\beta}(H[0,\tau], y, \beta) \xrightarrow{P} \phi_H^{\beta}(H[0,\tau], y, \beta)$$

$$n^{-1}U_{\beta}^H(H[0,\tau], t, \beta) \xrightarrow{P} \phi_{\beta}^H(H[0,\tau], \beta)$$

$$n^{-1}U_{\beta}^{\beta}(H[0,\tau], \beta) \xrightarrow{P} \phi_H^H(H[0,\tau], t, y, \beta).$$

B.6   Covariance matrix for $\{dW_1(t), W_2\}$

Based on Section 3.5.2, $W_1(t) = \int_0^t \phi_H^H(H[0,\tau], t, y, \beta)dV(y) + \phi_H^{\beta}(H[0,\tau], t, \beta)(\widehat{\beta} - \beta)$. Therefore, we have

$$dW_1(t) = \phi_H^H(H[0,\tau], t, t, \beta)dV(t) + \int_0^t \dot{\phi}_H^H(H[0,\tau], t, y, \beta)dV(y)$$

$$+\dot{\phi}_H^{\beta}(H[0,\tau], t, \beta)(\widehat{\beta} - \beta),$$

where $\dot{\phi}_H^H(H[0,\tau], t, y, \beta) = \frac{\partial}{\partial t}\phi_H^H(H[0,\tau], t, y, \beta)$ and $\dot{\phi}_H^{\beta}(H[0,\tau], t, \beta) = \frac{\partial}{\partial t}\phi_H^{\beta}(H[0,\tau], t, \beta)$.

Then for $s, t \in [0, \tau]$, let $\sigma_{W_1}^{*,2}(s, t) = cov\left\{dW_1(s), dW_1(t)\right\}$

$$
\begin{aligned}
\sigma_{W_1}^{*,2}(s, t) = {} & \phi_H^H(H[0, \tau], s, s, \beta)\phi_H^H(H[0, \tau], t, t, \beta)E\left\{dV(s)dV(t)\right\} \\
& + \phi_H^H(H[0, \tau], s, s, \beta)\int_0^t \dot{\phi}_H^H(H[0, \tau], s, y, \beta)E\left\{dV(y)dV(s)\right\} \\
& + \phi_H^H(H[0, \tau], s, s, \beta)\dot{\phi}_H^\beta(H[0, \tau], t, \beta)E\left\{dV(s)(\widehat{\beta} - \beta)\right\}, \\
& + \phi_H^H(H[0, \tau], t, t, \beta)\int_0^s \dot{\phi}_H^H(H[0, \tau], s, y, \beta)E\left\{dV(y)dV(t)\right\} \\
& + \int_0^s\int_0^t \dot{\phi}_H^H(H[0, \tau], s, y, \beta)\dot{\phi}_H^H(H[0, \tau], s, y^*, \beta)E\left\{dV(y)dV(y^*)\right\} \\
& + \dot{\phi}_H^\beta(H[0, \tau], t, \beta)\int_0^s \dot{\phi}_H^H(H[0, \tau], s, y, \beta)E\left\{dV(y)(\widehat{\beta} - \beta)\right\}, \\
& + \dot{\phi}_H^\beta(H[0, \tau], s, \beta)\phi_H^H(H[0, \tau], t, t, \beta)E\left\{dV(t)(\widehat{\beta} - \beta)\right\}, \\
& + \dot{\phi}_H^\beta(H[0, \tau], s, \beta)\int_0^t \dot{\phi}_H^H(H[0, \tau], s, y, \beta)E\left\{dV(y)(\widehat{\beta} - \beta)\right\}, \\
& + \dot{\phi}_H^\beta(H[0, \tau], s, \beta)\dot{\phi}_H^\beta(H[0, \tau], t, \beta)E\left\{(\widehat{\beta} - \beta)^2\right\}.
\end{aligned}
$$

Similarly, for $W_2 = \int_0^\tau \phi_\beta^H(H[0, \tau], y, \beta)dV(y) + \phi_\beta^\beta(H[0, \tau], \beta)(\widehat{\beta} - \beta)$, let $\sigma_{W_2}^{*,2}(\beta) = var(W_2)$. Then we have

$$
\begin{aligned}
\sigma_{W_2}^{*,2}(\beta) = {} & \int_0^\tau\int_0^\tau \phi_\beta^H(H[0, \tau], y, \beta)\phi_\beta^H(H[0, \tau], y^*, \beta)E\left\{dV(y)dV(y^*)\right\} \\
& + 2\phi_\beta^\beta(H[0, \tau], \beta)\int_0^\tau \phi_\beta^H(H[0, \tau], y, \beta)E\left\{dV(y)(\widehat{\beta} - \beta)\right\}, \\
& + \left\{\phi_\beta^\beta(H[0, \tau], \beta)\right\}^2 E\left\{(\widehat{\beta} - \beta)^2\right\}.
\end{aligned}
$$

The covariance between $dW_1(t)$ and W_2, denoted as $\sigma^{*,2}_{W_1,W_2}(t,\beta)$, becomes

$$
\begin{aligned}
\sigma^{*,2}_{W_1,W_2}(t,\beta) &= \phi^H_H(H[0,\tau],t,t,\beta)\int_0^\tau \phi^H_\beta(H[0,\tau],y,\beta)E\left\{dV(y)dV(t)\right\} \\
&+ \int_0^t\int_0^\tau \dot{\phi}^H_H(H[0,\tau],t,y,\beta)\phi^\beta_H(H[0,\tau],y^*,\beta)E\left\{dV(y)dV(y^*)\right\} \\
&+ \dot{\phi}^\beta_H(H[0,\tau],t,\beta)\int_0^\tau \phi^H_\beta(H[0,\tau],y,\beta)E\left\{dV(y)(\widehat{\beta}-\beta)\right\}, \\
&+ \phi^H_H(H[0,\tau],t,t,\beta)\phi^\beta_\beta(H[0,\tau],\beta)E\left\{dV(t)(\widehat{\beta}-\beta)\right\} \\
&+ \phi^\beta_\beta(H[0,\tau],\beta)\int_0^t \dot{\phi}^H_H(H[0,\tau],t,y,\beta)E\left\{dV(y)(\widehat{\beta}-\beta)\right\} \\
&+ \dot{\phi}^\beta_H(H[0,\tau],t,\beta)\phi^\beta_\beta(H[0,\tau],\beta)E\left\{(\widehat{\beta}-\beta)^2\right\}.
\end{aligned}
$$

B.7   Hessian matrix *I*

Along the lines of Chen (2009), Chen (2010), Hessian matrix *I* involves four different components: $I_{dH_jdH_j}$, $I_{dH_kdH_j}$, $I_{dH_j\beta}$, and $I_{\beta\beta}$ for $j,k=1,...,p$. Those terms can be expressed as:

$$
\begin{aligned}
I_{dH_jdH_j} &= -\frac{\partial^2 l}{\partial dH_j\partial dH_j} \\
&= \left\{\sum_{i=1}^n dN_i(t_j)\right\}^2 (dH_j)^{-2} \\
&+ \int_{t_j}^\tau \left[\frac{\ddot{\Theta}(u,H[0,u],t_j,t_j|z_i[0,u])}{\Theta^0(u,H[0,u]|z_i[0,u])}\right. \\
&\left.+\left\{\frac{\dot{\Theta}(u,H[0,u],t_j|z_i[0,u])}{\Theta^0(u,H[0,u]|z_i[0,u])}\right\}^2\right]dN_i(u), \\
I_{dH_kdH_j} &= -\frac{\partial^2 l}{\partial dH_k\partial dH_j} \\
&= \sum_{i=1}^n\int_{t_k\vee t_j}^\tau \left\{\frac{\ddot{\Theta}(u,H[0,u],t_k,t_j|z_i[0,u])}{\Theta^0(u,H[0,u]|z_i[0,u])}\right. \\
&\left.+\frac{\dot{\Theta}^2(u,H[0,u],t_k|z_i[0,u])\dot{\Theta}(u,H[0,u],t_j|z_i[0,u])}{\{\Theta^0(u,H[0,u]|z_i[0,u])\}^2}\right\}dN_i(u),
\end{aligned}
$$

$$
\begin{aligned}
I_{dH_j\beta} &= -\frac{\partial^2 l}{\partial dH_j \partial \beta} \\
&= \sum_{i=1}^{n} \int_{t_j}^{\tau} \left\{ \frac{\dot{\Theta}^\beta(u, H[0,u], t_j | z_i[0,u])}{\Theta^0(u, H[0,u] | z_i[0,u])} \right. \\
&\quad \left. - \frac{\dot{\Theta}(u, H[0,u], t_j | z_i[0,u]) \Theta^\beta(u, H[0,u] | z_i[0,u])}{\{\Theta^0(u, H[0,u] | z_i[0,u])\}^2} \right\} dN_i(u) \\
&\quad - Y_i(t_j) \Theta^\beta(t_j, H[0,t_j] | z_{t_j}^i) \\
&\quad - \int_{t_j}^{\tau} Y_i(u) \dot{\Theta}^\beta(u, H[0,u], t_j | z_i[0,u]) dH_u, \\
I_{\beta\beta} &= -\frac{\partial^2 l}{\partial \beta \partial \beta^T} \\
&= \sum_{i=1}^{n} \int_0^{\tau} \left[\frac{\Theta^{\beta,\beta}(u, H[0,u] | z_i[0,u])}{\Theta^0(u, H[0,u] | z_i[0,u])} - \left\{ \frac{\Theta^\beta(u, H[0,u] | z_i[0,u])}{\Theta^0(u, H[0,u] | z_i[0,u])} \right\}^2 \right] dN_i(u) \\
&\quad - Y_i(u) \Theta^\beta(u, H[0,u] | z_i[0,u]) dH_u
\end{aligned}
$$

where

$$
\begin{aligned}
\ddot{\Theta}(u, H[0,u], t_k, t_j | z_i[0,u]) &= E(U_u U_j U_k | e^{\int_0^u U_s dH_s}) \\
&\quad - E(U_u U_j | e^{\int_0^u U_s dH_s}) \Theta(t_k, H[0,u] | z_i[0,u])) \\
&\quad + \dot{\Theta}(u, H[0,u], t_k | z_i[0,u]) \Theta(t_j, H[0,u] | z_i[0,u]) \\
&\quad + \Theta(u, H[0,u] | z_i[0,u]) \dot{\Theta}(t_j, H[0,u], t_k | z_i[0,u]) \\
\dot{\Theta}^\beta(u, H[0,u], t_j | z_i[0,u]) &= \frac{\partial}{\partial \beta} \dot{\Theta}(u, H[0,u], t_j | z_i[0,u])
\end{aligned}
$$

B.8 Properties of estimates using profile likelihood

Suppose that $\sqrt{n}\left\{\widehat{H}_t - H_t\right\}$ has limiting process $V(t)$, then we have the following equation:

$$
\begin{aligned}
\widehat{H}_t = \sum_{i=1}^{n} \int_0^t dN_i(x) \times \\
\Theta_i(\widehat{H}_x, \beta) Y_i(x) dx - \sum_{i=1}^{n} \int_x^\tau \left\{ \frac{\dot{\Theta}_i(\widehat{H}_u, \beta)}{\Theta_i(\widehat{H}_u, \beta)} dN_i(u) - \dot{\Theta}_i(\widehat{H}_u, \beta) Y_i(u) \widehat{d}H_u \right\}.
\end{aligned}
$$

We can then look at convergence behavior of $\left\{ \widehat{H}_t(\widehat{H}_t, \beta_0) - H_t \right\}$ since it is necessary to derive property of $\widehat{\beta}$ under profile likelihood. The derivation is as follows:

$$
\sqrt{n} \left\{ \widehat{H}_t(\widehat{H}_t, \beta_0) - H_t \right\}
$$

$$
= \; n^{-1/2} \sum_{i=1}^n \int_0^t \frac{dN_i(x) - Y_i(x)\Theta_i(H_x, \beta_0)dH_x}{n^{-1}\sum_{i=1}^n Y_i(x)\Theta_i(\widehat{H}_x, \beta_0)\hat{w}_i(x)}
$$

$$
+ \; n^{-1/2} \sum_{i=1}^n \int_0^t \frac{Y_i(x)\Theta_i(H_x, \beta_0)dH_x - Y_i(x)\Theta_i(\widehat{H}_x, \beta_0)dH_x}{n^{-1}\sum_{i=1}^n Y_i(x)\Theta_i(\widehat{H}_x, \beta_0)\hat{w}_i(x)}
$$

$$
+ \; n^{-1/2} \sum_{i=1}^n \int_0^t \frac{dH_x}{n^{-1}\sum_{i=1}^n Y_i(x)\Theta_i(\widehat{H}_x, \beta_0)\hat{w}_i(x)} \times
$$

$$
\left(\int_x^\tau \frac{\dot\Theta_i(\widehat{H}_u, \beta_0)}{\Theta_i(\widehat{H}_u, \beta_0)} \left\{ dN_i(u) - \Theta_i(H_u, \beta_0)Y_i(u)dH_u \right\} \right)
$$

$$
+ \; n^{-1/2} \sum_{i=1}^n \int_0^t \frac{dH_x}{n^{-1}\sum_{i=1}^n Y_i(x)\Theta_i(\widehat{H}_x, \beta_0)\hat{w}_i(x)} \times
$$

$$
\left(\int_x^\tau \frac{\dot\Theta_i(\widehat{H}_u, \beta_0)}{\Theta_i(\widehat{H}_u, \beta_0)} \left\{ \Theta_i(H_u, \beta_0) - \Theta_i(\widehat{H}_u, \beta_0) \right\} Y_i(u)dH_u \right)
$$

$$
- \; n^{-1/2} \sum_{i=1}^n \int_0^t \frac{dH_x}{n^{-1}\sum_{i=1}^n Y_i(x)\Theta_i(\widehat{H}_x, \beta_0)\hat{w}_i(x)} \times
$$

$$
\left(\int_x^\tau \dot\Theta_i(\widehat{H}_u, \beta_0)Y_i(u)(d\widehat{H}_u - dH_u) \right)
$$

$$
= \; n^{-1/2} \sum_{i=1}^n \int_0^t \frac{\{1 + dH_x\dot\Theta_i(\widehat{H}_x, \beta_0)/\Theta_i(\widehat{H}_x, \beta_0)\}}{n^{-1}\sum_{i=1}^n Y_i(x)\Theta_i(\widehat{H}_x, \beta_0)\hat{w}_i(x)} dM_i(x)
$$

$$
- \; n^{1/2} \int_0^t \frac{n^{-1}\sum_{i=1}^n Y_i(x)\dot\Theta_i(H_x, \beta_0)(\widehat{H}_x - H_x)dH_x}{n^{-1}\sum_{i=1}^n Y_i(x)\Theta_i(\widehat{H}_x, \beta_0)\hat{w}_i(x)}
$$

$$
+ \; n^{-1/2} \sum_{i=1}^n \int_0^t \frac{dH_x}{n^{-1}\sum_{i=1}^n Y_i(x)\Theta_i(\widehat{H}_x, \beta_0)\hat{w}_i(x)} \left\{ \int_{x+}^\tau \frac{\dot\Theta_i(\widehat{H}_u, \beta_0)}{\Theta_i(\widehat{H}_u, \beta_0)} dM_i(u) \right\}
$$

$$
- \; n^{-1/2} \sum_{i=1}^n \int_0^t \frac{dH_x}{n^{-1}\sum_{i=1}^n Y_i(x)\Theta_i(\widehat{H}_x, \beta_0)\hat{w}_i(x)} \times
$$

$$
\left(\int_x^\tau \frac{\dot\Theta_i(\widehat{H}_u, \beta_0)\dot\Theta_i(H_u, \beta_0)}{\Theta_i(\widehat{H}_u, \beta_0)} (\widehat{H}_u - H_u)Y_i(u)dH_u \right)
$$

$$
- \; n^{-1/2} \sum_{i=1}^n \int_0^t \frac{dH_x}{n^{-1}\sum_{i=1}^n Y_i(x)\Theta_i(\widehat{H}_x, \beta_0)\hat{w}_i(x)} \times
$$

$$
\left(\int_x^\tau \dot\Theta_i(\widehat{H}_u, \beta_0)Y_i(u)(d\widehat{H}_u - dH_u) \right)
$$

The first term can be further decomposed as follows by consistency of $\widehat{H}$:

$$n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{\{1 + dH_x \dot{\Theta}_i(H_x, \beta_0)/\Theta_i(H_x, \beta_0)\}}{n^{-1} \sum_{i=1}^{n} Y_i(x)\Theta_i(H_x, \beta_0)w_i(x)} dM_i(x)\{1 + o_p(1)\}$$

Furthermore, we know that

$$n^{-1} \sum_{i=1}^{n} Y_i(x)\Theta_i(H_x, \beta_0)w_i(x)$$

$$\text{(B.1)} \qquad = \quad n^{-1} \left\{ \sum_{i=1}^{n} Y_i(x)\Theta_i(H_x, \beta_0) - \sum_{i=1}^{n} \int_x^\tau \frac{\dot{\Theta}_i(H_u, \beta_0)}{\Theta_i(H_u, \beta_0)} dM_i(u) \right\}$$

As $n \to \infty$, by WLLN, $n^{-1} \sum_{i=1}^{n} Y_i(x)\Theta_i(H_x, \beta_0)$ converges in probability to a deterministic function $c_x = E\{\Theta(H_x, \beta_0); X \geq x\}$. On the other hand, the second term,

$$n^{-1} \sum_{i=1}^{n} \int_x^\tau \frac{\dot{\Theta}_i(H_u, \beta_0)}{\Theta_i(H_u, \beta_0)} dM_i(u),$$

is a Martingale with predictable variation process (PVP) as follows:

$$n^{-2} \sum_{i=1}^{n} \int_x^\tau \frac{\dot{\Theta}_i^2(H_u, \beta_0)}{\Theta_i(H_u, \beta_0)} Y(u) dH_u.$$

The PVP converges to zero as $n \to \infty$. That is, (B.1) converges uniformly to c_x as $n \to \infty$.

By Martingale Central Limit Theorum (MCLT),

$$n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{\{1 + dH_x \dot{\Theta}_i(H_x, \beta_0)/\Theta_i(H_x, \beta_0)\}}{n^{-1} \sum_{i=1}^{n} Y_i(x)\Theta_i(H_x, \beta_0)w_i(x)} dM_i(x) \quad \Rightarrow \quad W_1(t)$$

where $W_1(t)$ is a mean-zero Gaussian process with variance

There are two implications for subsequent derivations for (B.1). First, we can substitute c_x as denominator when establishing asymptotics; Second, the $o_p(1)$ term can be dropped since the product of $o_p(1)$ and a converged Gaussian process is still an $o_p(1)$ term.

Similarly, the third term, independent of first term, becomes (up to $o_p(1)$):

$$n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{dH_x}{n^{-1} \sum_{i=1}^{n} Y_i(x)\Theta_i(H_x, \beta_0)} \int_{x+}^\tau \frac{\dot{\Theta}_i(H_u, \beta_0)}{\Theta_i(H_u, \beta_0)} dM_i(u).$$

The integrand is in fact a Martingale that converges to a mean-zero Gaussian process $W_2(x)$ with variance $\sigma_2(x) = (dH_x/c_x)^2 E\left[\int_{x+}^{\tau} \left\{\frac{\dot{\Theta}(H_u,\beta_0)}{\Theta(H_u,\beta_0)}\right\}^2 Y(u)\Theta(H_u,\beta_0)dH_u\right]$. Hence the whole term converges to process $\int_0^t W_2(x)dx$. Furthermore, W_1 and W_2 are independent, we can combine them as one zero-mean Wiener process W with variance function $\sigma(t) = I(x \in [0,t])\sigma_1(x) + I(x \in (t,\tau])\sigma_2(x)$.

Since the second term in (B.1) includes the process $\sqrt{n}\left\{\widehat{H}_x(\widehat{H}_x,\beta_0) - H_x\right\}$, this term will converge to the recursive process $V_1(t) = \int_0^t \frac{E\left\{\dot{\Theta}(H_x,\beta_0);X\geq x\right\}}{E\left\{\Theta(H_x,\beta_0);X\geq x\right\}}V(x)dH_x$. Similarly, the fourth and fifth terms will converge to:

$$
\begin{aligned}
V_2(t) &= \int_0^t \frac{dH_x}{c_x}\int_x^{\tau} E\left\{\frac{\dot{\Theta}^2(H_x,\beta_0)}{\Theta(H_x,\beta_0)}; X \geq x\right\} V(u)dH_u \\
V_3(t) &= \int_0^t \frac{dH_x}{c_x}\int_x^{\tau} E\left\{\dot{\Theta}(H_x,\beta_0); X \geq x\right\} dV(u)
\end{aligned}
$$

In summary, $\sqrt{n}\{\widehat{H}_t - H_t\}$ converges to $V(t)$ that satisfies the following equation:

(B.2)
$$
V(t) = W(t) - V_1(t) - V_2(t) - V_3(t).
$$

This is a non homogeneous second order differential equations with time changing coefficients. In general there is no explicit solution.

We can rewrite $\sqrt{n}(\widehat{\beta} - \beta_0)$ as:

$$
\sqrt{n}(\widehat{\beta} - \beta_0) = \{n^{-1}I_{pr}(\beta_0)\}^{-1}n^{-1/2}U_{pr}(\beta_0) + o_p(1).
$$

We have already shown that $n^{-1}I_{pr}(\beta_0) \xrightarrow{p} I_1(\beta_0)$.

For $U_{pr}(\beta_0)$, it can be written as:

$$
\begin{aligned}
& n^{-1/2}U_{pr}(\beta_0) \\
= \; & n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\frac{\Theta_i^{\beta_0}(\widehat{H}_x,\beta_0)}{\Theta_i(\widehat{H}_x,\beta_0)}\left\{dN_i(x)-Y_i(x)\Theta_i(\widehat{H}_x,\beta_0)\widehat{dH_x}\right\} \\
= \; & n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\frac{\Theta_i^{\beta_0}(\widehat{H}_x,\beta_0)}{\Theta_i(\widehat{H}_x,\beta_0)}\left\{dN_i(x)-Y_i(x)\Theta_i(H_x,\beta_0)dH_x\right\} \\
+ \; & n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\frac{\Theta_i^{\beta_0}(\widehat{H}_x,\beta_0)}{\Theta_i(\widehat{H}_x,\beta_0)}\left\{Y_i(x)\Theta_i(H_x,\beta_0)dH_x-Y_i(x)\Theta_i(\widehat{H}_x,\beta_0)dH_x\right\} \\
+ \; & n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\frac{\Theta_i^{\beta_0}(\widehat{H}_x,\beta_0)}{\Theta_i(\widehat{H}_x,\beta_0)}\left\{Y_i(x)\Theta_i(\widehat{H}_x,\beta_0)dH_x-Y_i(x)\Theta_i(\widehat{H}_x,\beta_0)\widehat{dH_x}\right\} \\
= \; & n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\frac{\Theta_i^{\beta_0}(\widehat{H}_x,\beta_0)}{\Theta_i(\widehat{H}_x,\beta_0)}dM_i(x) \\
- \; & n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\frac{\Theta_i^{\beta_0}(\widehat{H}_x,\beta_0)}{\Theta_i(\widehat{H}_x,\beta_0)}Y_i(x)\dot{\Theta}_i(H_x,\beta_0)dH_x(\widehat{H}_x-H_x) \\
\text{(B.3)} \quad - \; & n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\Theta_i^{\beta_0}(\widehat{H}_x,\beta_0)Y_i(x)(\widehat{dH_x}-dH_x)
\end{aligned}
$$

The first term is (up to $o_p(1)$) a Martingale and its asymptotic behavior is equivalent to that of $n^{-1/2}U(\beta_0)$.

According to (3.16), $n^{-1/2}U(\beta_0)$ is a Martingale with PVP:

$$
n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\frac{\{\Theta_i^{\beta_0}(H_x,\beta_0)\}^2}{\Theta_i(H_x,\beta_0)}Y_i(x)dH(x),
$$

that converges to $I_1(\beta_0)$. By Martingale CLT, $n^{-1/2}U(\beta_0)$, a realization of a martingale process at $t=\tau$, converges in distribution to zero-mean Normal with variance $I_1(\beta_0)$. The arguments are similar to Fleming and Harrington (2005).

The second term in (B.3) will converge to $\int_0^{\tau}E\left\{\frac{\Theta^{\beta_0}(H_x,\beta_0)\dot{\Theta}(H_x,\beta_0)}{\Theta(H_x,\beta_0)};X\geq x\right\}V(x)dH_x$

Last term in (B.3) equals (up to $o_p(1)$) $n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\Theta_i^{\beta_0}(H_x,\beta_0)Y_i(x)(\widehat{dH_x}-dH_x)$, and it converges to following process:

$$= \quad \sqrt{n} \sum_{i=1}^{n} \int_0^\tau \Theta_{i,x}^{0,\beta} Y_i(x) \frac{dN_i(x) - h_x Y_i(x) \Theta_i(\widehat{H}_x, \widehat{\beta}) w_i(x)}{\sum_{i=1}^{n} Y_i(x) \Theta_i(\widehat{H}_x, \widehat{\beta}) w_i(x)}$$

$$\text{(B.4)} n^{-1/2} \sum_{i=1}^{n} \int_0^\tau \Theta_{i,x}^{0,\beta} Y_i(x) \frac{\{1 + h_x (\log \Theta_i(\widehat{H}_x, \widehat{\beta}))'\} dM_i(x) + h_x \int_{x+}^\tau (\log \Theta_{i,s})' dM_i(s)}{n^{-1} \sum_{i=1}^{n} Y_i(x) \Theta_i(\widehat{H}_x, \widehat{\beta}) w_i(x)}.$$

Follow the similar arguments in Section 4.1, (B.4) is asymptotically equivalent to distribution of martingale $M_1^*(\tau)$ plus the integration of another indepedent martingale $M_2^*(x)$ over $[0, \tau]$ where:

$$M_1^*(\tau) \quad = \quad n^{-1/2} \sum_{i=1}^{n} \int_0^\tau \Theta_{i,x}^{0,\beta} Y_i(x) \frac{1 + h_x (\log \Theta_i(\widehat{H}_x, \widehat{\beta}))'}{C_x} dM_i(x);$$

$$M_2^*(x) \quad = \quad n^{-1/2} \Theta_{i,x}^{0,\beta} Y_i(x) h_x \sum_{i=1}^{n} \int_{x+}^\tau \frac{(\log \Theta_{i,s})'}{C_x} dM_i(s),$$

with PVPs:

$$n^{-1} \quad \sum_{i=1}^{n} \quad \int_0^\tau \left\{ \Theta_{i,x}^{0,\beta} \frac{1 + h_x (\log \Theta_i(\widehat{H}_x, \widehat{\beta}))'}{C_x} \right\}^2 Y_i^2(x) \Theta_i(\widehat{H}_x, \widehat{\beta}) h_x;$$

$$n^{-1} \{\Theta_{i,x}^{0,\beta} Y_i(x) h_x\}^2 \quad \sum_{i=1}^{n} \quad \int_{x+}^\tau \left\{ \frac{(\log \Theta_{i,s})'}{C_x} \right\}^2 Y_i(s) \Theta_{i,s} h_s,$$

that converge to

$$\sigma_1^*(\tau) \quad = \quad E \left[\int_0^\tau \left\{ \Theta_{i,x}^{0,\beta} \frac{1 + h_x (\log \Theta_i(\widehat{H}_x, \widehat{\beta}))'}{C_x} \right\}^2 Y_i^2(x) \Theta_i(\widehat{H}_x, \widehat{\beta}) h_x \right];$$

$$\sigma_2^*(x) \quad = \quad \{\Theta_{i,x}^{0,\beta} Y_i(x) h_x\}^2 E \left[\int_{x+}^\tau \left\{ \frac{(\log \Theta_{i,s})'}{C_x} \right\}^2 Y_i(s) \Theta_{i,s} h_s \right],$$

as $n \to \infty$. Hence by Martingale Central Limit Theorum and Slutsky Theorum, (B.4) converges to a mean zero Normal $W_1^*(\tau)$ with variance $\sigma_1^*(\tau)$ plus integration of a Gaussian process $W_2^*(x)$ with variance $\sigma_2^*(x)$ over $[0, \tau]$, respectively. That is,

$$\sqrt{n} \sum_{i=1}^{n} \int_0^\tau \Theta_{i,x}^{0,\beta} Y_i(x) (h_x - \widehat{h_x}) \xrightarrow{D} W_1^*(\tau) + \int_0^\tau W_2^*(x) dx$$

Hence, by Slutsky Theorum, the following holds:

$$n^{-1/2} \sum_{i=1}^{n} \int_0^\tau \Theta_{i,x}^{0,\beta} Y_i(x)(h_x - \widehat{h_x}) \Rightarrow 0.$$

Combining these results, we have :

$$\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{D} N(0, I_1^{-1}(\beta_0)).$$

APPENDIX C

Appendices for Chapter IV

C.1 Covariance matrix for (4.7)

Based on Section 4.3.3, for $s, t \in [0, \tau]$, the components in (4.8) are

$$
\begin{aligned}
\sigma_H^2(s,t) &= \int_{s^+ \vee t^+}^{\tau} \psi(x,s)\psi(x,t)P(T \geq x)\lambda(x|z[0,x])dx \\
\sigma_{H,A}^2(s,t) &= \int_{s \vee t}^{\tau} \psi(x, s \vee t)\mu(x)P(T \geq x)\lambda(x|z[0,x])dx \\
\sigma_{H,\beta}^2(s,\beta) &= \int_{s}^{\tau} \psi(x,s)\rho(x)P(T \geq x)\lambda(x|z[0,x])dx \\
\sigma_A^2(s,t) &= \int_{0}^{s \vee t} \mu^2(x)P(T \geq x)\lambda(x|z[0,x])dx \\
\sigma_{A,\beta}^2(s,\beta) &= \int_{s}^{\tau} \mu(x)\rho(x)P(T \geq x)\lambda(x|z[0,x])dx \\
\sigma_\beta^2(\beta) &= \int_{0}^{\tau} \rho^2(x)P(T \geq x)\lambda(x|z[0,x])dx
\end{aligned}
$$

C.2   Covariance matrix for $\{dW_1(t), dW_2(t), W_3\}$

Based on $W_1(t)$, $W_2(t)$ and W_3 given in Section 4.3.3, we have

$$
\begin{aligned}
dW_1(t) \;=\; & \phi_H^H(H[0,\tau], A[0,\tau], t, t, \beta)dV^H(t) \\
& + \int_0^t \dot\phi_H^H(H[0,\tau], A[0,\tau], t, y, \beta)dV^H(y) \\
& + \phi_H^A(H[0,\tau], A[0,\tau], t, t, \beta)dV^A(t) \\
& + \int_0^t \dot\phi_H^A(H[0,\tau], A[0,\tau], t, y, \beta)dV^A(y) \\
& + \dot\phi_H^\beta(H[0,\tau], A[0,\tau], t, \beta)(\widehat\beta - \beta),
\end{aligned}
$$

where

$$
\begin{aligned}
\dot\phi_H^H(H[0,\tau], A[0,\tau], t, y, \beta) \;&=\; \frac{\partial}{\partial t}\phi_H^H(H[0,\tau], A[0,\tau], t, y, \beta) \\
\dot\phi_H^A(H[0,\tau], A[0,\tau], t, y, \beta) \;&=\; \frac{\partial}{\partial t}\phi_H^A(H[0,\tau], A[0,\tau], t, y, \beta) \\
\dot\phi_H^\beta(H[0,\tau], t, \beta) \;&=\; \frac{\partial}{\partial t}\phi_H^\beta(H[0,\tau], A[0,\tau], t, \beta),
\end{aligned}
$$

$$
\begin{aligned}
dW_2(t) \;=\; & \phi_A^H(H[0,\tau], A[0,\tau], t, t, \beta)dV_H(t) \\
& + \int_0^t \dot\phi_A^H(H[0,\tau], A[0,\tau], t, y, \beta)dV^H(y) \\
& + \phi_A^A(H[0,\tau], A[0,\tau], t, t, \beta)dV_A(t) \\
& + \int_0^t \dot\phi_A^A(H[0,\tau], A[0,\tau], t, y, \beta)dV^A(y) \\
& + \dot\phi_A^\beta(H[0,\tau], A[0,\tau], t, \beta)(\widehat\beta - \beta)
\end{aligned}
$$

where

$$
\begin{aligned}
\dot\phi_A^H(H[0,\tau], A[0,\tau], t, y, \beta) \;&=\; \frac{\partial}{\partial t}\phi_A^H(H[0,\tau], A[0,\tau], t, y, \beta) \\
\dot\phi_A^A(H[0,\tau], A[0,\tau], t, y, \beta) \;&=\; \frac{\partial}{\partial t}\phi_A^A(H[0,\tau], A[0,\tau], t, y, \beta) \\
\dot\phi_A^\beta(H[0,\tau], t, \beta) \;&=\; \frac{\partial}{\partial t}\phi_A^\beta(H[0,\tau], A[0,\tau], t, \beta),
\end{aligned}
$$

Then for $s, t \in [0, \tau]$, let $\sigma_{W_1}^{*,2}(s, t) = cov\left\{dW_1(s)dW_1(t)\right\}$

$\sigma_{W_1}^{*,2}(s, t) =$

$\phi_H^H(H[0,\tau], A[0,\tau], s, s, \beta)\phi_H^H(H[0,\tau], A[0,\tau], t, t, \beta)E\left\{dV_H(s)dV_H(t)\right\}$

$+\phi_H^H(H[0,\tau], A[0,\tau], s, s, \beta)\int_0^t \dot{\phi}_H^H(H[0,\tau], A[0,\tau], s, y, \beta)E\left\{dV_H(y)dV_H(s)\right\}$

$+\phi_H^H(H[0,\tau], A[0,\tau], s, s, \beta)\phi_H^A(H[0,\tau], A[0,\tau], t, t, \beta)E\left\{dV_H(s)dV_A(t)\right\}$

$+\phi_H^H(H[0,\tau], A[0,\tau], s, s, \beta)\int_0^t \dot{\phi}_H^A(H[0,\tau], A[0,\tau], s, y, \beta)E\left\{dV_A(y)dV_H(s)\right\}$

$+\phi_H^H(H[0,\tau], A[0,\tau], s, s, \beta)\dot{\phi}_H^\beta(H[0,\tau], A[0,\tau], t, \beta)E\left\{dV_H(s)(\widehat{\beta} - \beta)\right\}$

$+\phi_H^H(H[0,\tau], A[0,\tau], t, t, \beta)\int_0^s \dot{\phi}_H^H(H[0,\tau], A[0,\tau], s, y, \beta)E\left\{dV_H(y)dV_H(t)\right\}$

$+\int_0^s \int_0^t \dot{\phi}_H^H(H[0,\tau], A[0,\tau], s, y, \beta)$

$\dot{\phi}_H^H(H[0,\tau], A[0,\tau], s, y^*, \beta)E\left\{dV_H(y)dV_H(y^*)\right\}$

$+\phi_H^A(H[0,\tau], A[0,\tau], t, t, \beta)\int_0^s \dot{\phi}_H^H(H[0,\tau], A[0,\tau], s, y, \beta)E\left\{dV_H(y)dV_A(t)\right\}$

$+\int_0^s \int_0^t \dot{\phi}_H^H(H[0,\tau], A[0,\tau], s, y, \beta)$

$\dot{\phi}_H^A(H[0,\tau], A[0,\tau], s, y^*, \beta)E\left\{dV_H(y)dV_A(y^*)\right\}$

$+\dot{\phi}_H^\beta(H[0,\tau], A[0,\tau], t, \beta)\int_0^s \dot{\phi}_H^H(H[0,\tau], A[0,\tau], s, y, \beta)E\left\{dV_H(y)(\widehat{\beta} - \beta)\right\}$

$+\phi_H^A(H[0,\tau], A[0,\tau], s, s, \beta)\phi_H^H(H[0,\tau], A[0,\tau], t, t, \beta)E\left\{dV_A(s)dV_H(t)\right\}$

$+\phi_H^A(H[0,\tau], A[0,\tau], s, s, \beta)\int_0^t \dot{\phi}_H^H(H[0,\tau], A[0,\tau], s, y, \beta)E\left\{dV_H(y)dV_A(s)\right\}$

$+\phi_H^A(H[0,\tau], A[0,\tau], s, s, \beta)\phi_H^A(H[0,\tau], A[0,\tau], t, t, \beta)E\left\{dV_A(s)dV_A(t)\right\}$

$+\phi_H^A(H[0,\tau], A[0,\tau], s, s, \beta)\int_0^t \dot{\phi}_H^A(H[0,\tau], A[0,\tau], s, y, \beta)E\left\{dV_A(y)dV_A(s)\right\}$

$+\phi_H^A(H[0,\tau], A[0,\tau], s, s, \beta)\dot{\phi}_H^\beta(H[0,\tau], A[0,\tau], t, \beta)E\left\{dV_A(s)(\widehat{\beta} - \beta)\right\},$

$+\phi_H^H(H[0,\tau], A[0,\tau], t, t, \beta)\int_0^s \dot{\phi}_H^A(H[0,\tau], A[0,\tau], s, y, \beta)E\left\{dV_A(y)dV_H(t)\right\}$

$$+ \int_0^s \int_0^t \dot{\phi}_H^A(H[0,\tau], A[0,\tau], s, y, \beta) \dot{\phi}_H^H(H[0,\tau], A[0,\tau], s, y^*, \beta) E\{dV_A(y)dV_H(y^*)\}$$

$$+ \phi_H^A(H[0,\tau], A[0,\tau], t, t, \beta) \int_0^s \dot{\phi}_H^A(H[0,\tau], A[0,\tau], s, y, \beta) E\{dV_A(y)dV_A(t)\}$$

$$+ \int_0^s \int_0^t \dot{\phi}_H^H(H[0,\tau], A[0,\tau], s, y, \beta) \dot{\phi}_H^A(H[0,\tau], A[0,\tau], s, y^*, \beta) E\{dV_H(y)dV_A(y^*)\}$$

$$+ \dot{\phi}_H^\beta(H[0,\tau], A[0,\tau], t, \beta) \int_0^s \dot{\phi}_H^A(H[0,\tau], A[0,\tau], s, y, \beta) E\left\{dV_A(y)(\widehat{\beta} - \beta)\right\}$$

$$+ \dot{\phi}_H^\beta(H[0,\tau], A[0,\tau], s, \beta) \phi_H^H(H[0,\tau], A[0,\tau], t, t, \beta) E\left\{dV_H(t)(\widehat{\beta} - \beta)\right\},$$

$$+ \dot{\phi}_H^\beta(H[0,\tau], A[0,\tau], s, \beta) \int_0^t \dot{\phi}_H^H(H[0,\tau], A[0,\tau], s, y, \beta) E\left\{dV_H(y)(\widehat{\beta} - \beta)\right\},$$

$$+ \dot{\phi}_H^\beta(H[0,\tau], A[0,\tau], s, \beta) \phi_H^A(H[0,\tau], A[0,\tau], t, t, \beta) E\left\{dV_A(t)(\widehat{\beta} - \beta)\right\},$$

$$+ \dot{\phi}_H^\beta(H[0,\tau], A[0,\tau], s, \beta) \int_0^t \dot{\phi}_H^A(H[0,\tau], A[0,\tau], s, y, \beta) E\left\{dV_A(y)(\widehat{\beta} - \beta)\right\},$$

$$+ \dot{\phi}_H^\beta(H[0,\tau], A[0,\tau], s, \beta) \dot{\phi}_H^\beta(H[0,\tau], A[0,\tau], t, \beta) E\left\{(\widehat{\beta} - \beta)^2\right\}.$$

The covariance of $\{dW_1(s), dW_2(t)\}$, denoted as $\sigma_{W_1,W_2}^{*,2}(s,t)$ can be obtained in similar fashion.

Now let $\sigma_{W_3}^{*,2}(\beta) = var(W_3)$. Then we have

$$\sigma_{W_3}^{*,2}(\beta) =$$

$$\int_0^\tau \int_0^\tau \phi_\beta^H(H[0,\tau], A[0,\tau], y, \beta)$$

$$\phi_\beta^H(H[0,\tau], A[0,\tau], y^*, \beta) E\{dV_H(y)dV_H(y^*)\}$$

$$+ 2\phi_\beta^\beta(H[0,\tau], A[0,\tau], \beta) \int_0^\tau \phi_\beta^H(H[0,\tau], A[0,\tau], y, \beta) E\left\{dV_H(y)(\widehat{\beta} - \beta)\right\},$$

$$+ \int_0^\tau \int_0^\tau \phi_\beta^A(H[0,\tau], A[0,\tau], y, \beta)$$

$$\phi_\beta^A(H[0,\tau], A[0,\tau], y^*, \beta) E\{dV_A(y)dV_A(y^*)\}$$

$$+ 2\phi_\beta^\beta(H[0,\tau], A[0,\tau], \beta) \int_0^\tau \phi_\beta^A(H[0,\tau], A[0,\tau], y, \beta) E\left\{dV_A(y)(\widehat{\beta} - \beta)\right\},$$

$$+ \left\{\phi_\beta^\beta(H[0,\tau], A[0,\tau], \beta)\right\}^2 E\left\{(\widehat{\beta} - \beta)^2\right\}.$$

The covariance between $dW_1(t)$ and W_3, denoted as $\sigma^{*,2}_{W_1,W_3}(t,\beta)$, becomes

$$\sigma^{*,2}_{W_1,W_3}(t,\beta) =$$

$$\phi^H_H(H[0,\tau],A[0,\tau],t,t,\beta)\int_0^\tau \phi^H_\beta(H[0,\tau],A[0,\tau],y,\beta)E\{dV_H(y)dV_H(t)\}$$

$$+\int_0^t\int_0^\tau \dot{\phi}^H_H(H[0,\tau],A[0,\tau],t,y,\beta)$$

$$\phi^\beta_H(H[0,\tau],A[0,\tau],y^*,\beta)E\{dV_H(y)dV_H(y^*)\}$$

$$+\phi^A_H(H[0,\tau],A[0,\tau],t,t,\beta)\int_0^\tau \phi^H_\beta(H[0,\tau],A[0,\tau],y,\beta)E\{dV_H(y)dV_A(t)\}$$

$$+\int_0^t\int_0^\tau \dot{\phi}^A_H(H[0,\tau],A[0,\tau],t,y,\beta)$$

$$\phi^\beta_H(H[0,\tau],A[0,\tau],y^*,\beta)E\{dV_A(y)dV_H(y^*)\}$$

$$+\dot{\phi}^\beta_H(H[0,\tau],A[0,\tau],t,\beta)\int_0^\tau \phi^H_\beta(H[0,\tau],A[0,\tau],y,\beta)E\left\{dV_H(y)(\widehat{\beta}-\beta)\right\}$$

$$+\phi^H_H(H[0,\tau],A[0,\tau],t,t,\beta)\int_0^\tau \phi^A_\beta(H[0,\tau],A[0,\tau],y,\beta)E\{dV_A(y)dV_H(t)\}$$

$$+\int_0^t\int_0^\tau \dot{\phi}^H_H(H[0,\tau],A[0,\tau],t,y,\beta)$$

$$\phi^A_\beta(H[0,\tau],A[0,\tau],y^*,\beta)E\{dV_H(y)dV_A(y^*)\}$$

$$+\phi^A_H(H[0,\tau],A[0,\tau],t,t,\beta)\int_0^\tau \phi^A_\beta(H[0,\tau],A[0,\tau],y,\beta)E\{dV_A(y)dV_A(t)\}$$

$$+\int_0^t\int_0^\tau \dot{\phi}^A_H(H[0,\tau],A[0,\tau],t,y,\beta)$$

$$\phi^A_\beta(H[0,\tau],A[0,\tau],y^*,\beta)E\{dV_A(y)dV_A(y^*)\}$$

$$+\dot{\phi}^\beta_H(H[0,\tau],A[0,\tau],t,\beta)\int_0^\tau \phi^A_\beta(H[0,\tau],A[0,\tau],y,\beta)E\left\{dV_A(y)(\widehat{\beta}-\beta)\right\}$$

$$+\phi^H_H(H[0,\tau],A[0,\tau],t,t,\beta)\phi^\beta_\beta(H[0,\tau],A[0,\tau],\beta)E\left\{dV_H(t)(\widehat{\beta}-\beta)\right\}$$

$$+\int_0^t \dot{\phi}^H_H(H[0,\tau],A[0,\tau],t,y,\beta)\phi^\beta_\beta(H[0,\tau],A[0,\tau],\beta)E\left\{dV_H(t)(\widehat{\beta}-\beta)\right\}$$

$$+\phi^A_H(H[0,\tau],A[0,\tau],t,t,\beta)\phi^\beta_\beta(H[0,\tau],A[0,\tau],\beta)E\left\{dV_A(t)(\widehat{\beta}-\beta)\right\}$$

$$+\int_0^t \dot{\phi}^A_H(H[0,\tau],A[0,\tau],t,y,\beta)\phi^\beta_\beta(H[0,\tau],A[0,\tau],\beta)E\left\{dV_A(t)(\widehat{\beta}-\beta)\right\}$$

$$+\dot{\phi}^\beta_H(H[0,\tau],A[0,\tau],t,\beta)\phi^\beta_\beta(H[0,\tau],A[0,\tau],\beta)E\left\{(\widehat{\beta}-\beta)^2\right\}.$$

The covariance between $dW_2(t)$ and W_3, denoted as $\sigma^{*,2}_{W_2,W_3}(t,\beta)$, can be obtained in

similar way.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aalen, O. O. (1992). Modelling heterogeneity in survival analysis by the compound poisson distribution. *Annals of Applied Probability* **4,** 951–972.

Andersen, P. K., Hansen, M., and Klein, J. P. (2004). Regression analysis of restricted mean survival time based on pseudo observations. *Life Data Analysis* **10,** 335–350.

Andersen, P. K. and Perme, M. P. (2010). Pseudo-observations in survival analysis. *Statistical Methods in Medical Research* **19,**.

Bagdonnavicius, V. and Nikulin, M. S. (1999). Generalized proportional hazards model based on modified partial likelihood. *Lifetime Data Analysis* **5,** 329–350.

Buckley, J. D. and James, I. R. (1979). Linear regression with censored data. *Biometrika* **66,** 429–436.

Chen, P. Y. and Tsiatis, A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics* **57,** 1030–1038.

Chen, Y. H. (2009). Weighted breslow-type estimator and maximum likelihood estimation in semiparametric transformation models. *Biometrika* **96,** 591–600.

Chen, Y. H. (2010). Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *Journal of the Royal Statistical Society, Series B* **72,** 235–251.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B* **34,** 187–200.

Dabrowska, D. M. and Doksum, K. A. (1979). Partial likelihood in transformation models with censored data. *Scandanavian Journal of Statistics* **10,** 1–24.

Faucett, C. L., Schenker, N., and Taylor, J. M. (2002). Survival analysis using auxiliary variables via multiple imputation, with application to aids clinical trial data. *Biometrics* **58,** 37–47.

Fine, J. P., Ying, Z., and Wei, L. J. (1998). n a linear transformation model for censored data. *Biometrika* **85,** 980–986.

Finkelstein, D. M. and Schoenfeld, D. A. (1994). Analyzing survival in the presence of an auxiliary variable. *Statistics in Medicine* **13,** 1747–1754.

Fleming, T. and Harrington, D. (2005). *Counting process and survival analysis*. Wiley Series in Probability and Statistics.

Gjessing, H., Aalen, O. O., and Hjort, N. L. (2003). Frailty models based on levy processes. *Advanced Applied Probability* **35,** 532–550.

Gray, R. J. (1994). A kernel method for incorporating information on disease progression in the analysis of survival. *Biometrika* **81,** 527–539.

Heitjan, D. F. and Little, R. J. A. (1991). Multiple imputation for the fatal accident reporting system. *Applied Statistics* **40,**.

Hougaard, P. (1996). Frailty models for survival data. *Lifetime Data Analysis* **1,** 255–273.

Hougaard, P. (2000). *Analysis of multivariate survival data*. Springer.

Hsu, C. H., Taylor, J. M., Murray, S., and Commenges, D. (2006). Survival analysis using auxiliary variables via non-parametric multiple imputation. *Statistics in Medicine* **25,** 3503–3517.

Irwin, J. O. (1949). The standard error of an estimate of expectational life. *Journal of Hygiene* **47,** 188–189.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53,** 457–481.

Karrison, K. (1987). Restricted mean life with adjustment for covariates. *Journal of the American Statistical Association* **82,** 1169–1176.

Klein, J. P. (1992). Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics* **48,** 795–806.

Kosorok, M. R., Lee, B. L., and Fine, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. *the Annals of Statistics* **32,** 1448–1491.

Lachout, P. (2001). Linear transformations of wiener process that born wiener process, brownian bridge or ornstein-uhlenbeck process. *Kybernetika* **6,** 647–667.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Little, R. J. A. and Rubin, D. (2002). *Statistical Analysis with Missing Data, 2nd Edition*. New York: Wiley.

Liu, L. X., Murray, S., and Tsodikov, A. (2011). Multiple imputation based on restricted mean model for censored data. *Statistics in Medicine* **30,** 1339–1350.

Mackenzie, T. and Abrahamowicz, M. (2005). Using categorical markers as auxiliary variables in log-rank tests and hazard ratio estimation. *The Canadian Journal of Statistics* **33,** 201–219.

Malani, H. M. (1995). A modification of the redistribution to the right algorithm using disease markers. *Biometrika* **82,** 515–526.

Martinussen, T. and Scheike, T. H. (2006). *Dynamic Regression Models for Survival Data*. Springer Science+Business Media, Inc., 233 Spring Street, New York, NY 10013, USA.

McDonald, D. (1947). *Elements of applied probability for engineering, mathematics and systems science*. World Science Publishing Co. Pte. Ltd, Suite 202 1060 Main Street, River Edge, NJ 07661 USA.

Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *the Annals of Statistics* **22,** 712–731.

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

Murphy, S. A. and van der Vaart, A. (2000). On profile likelihood. *Journal of the American Statistical Association* **95,** 449–485.

Murray, S. and Tsiatis, A. A. (1996). Nonparametric survival estimation using prognostic longitudinal covariates. *Biometrics* **52,**.

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

Pepe, M. and Fleming, T. (1998). Weighted kaplan-meier statistics - a class of distance tests for censored survival data. *Biometrics* **45,**.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology - Methodological Issues* .

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Rubins, D. B. and Schenker, N. (1991). Multiple imputation in health-care databases: An overview and some applications. *Statistics in Medicine* **10,** 585–598.

Shiryaev, A. N. (1960). Some problems in the spectral theory of higher order moments i. *The Probability Application* **5,** 265–284.

Taylor, J. M., Hsu, C. H., and Murray, S. (2002). Survival estimation and testing via multiple imputation. *Statistics & Probability letters* **58,** 221–232.

Tsodikov, A. (2003). Semiparametric models: a generalized self-consistency approach. *Journal of Royal Statistical Society, Series B* **65,** 759–774.

Tsodikov, A. and Muller, W. A. (1998). Modeling carcinogenesis under a time-changing exposure. *Mathematical Biosciences* **152,** 179–191.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16,** 439–454.

Yakovlev, A. Y., Tsodikov, A., and Bass, L. (1992). A stochastic model of hormesis. *Mathematical Biosciences* **116,** 197–219.

Yashin, A. T. and Manton, K. G. (1997). Effects of unobserved and partially observed covariate processes on system failure: a review of models and estimation strategies. *Statistical Science* **12,** 20–34.

Yue, H. and Chan, K. S. (1997). A dynamic frailty models for multivariate survival data. *Biometrics* **53,** 785–793.

Zeng, D. L. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *Journal of Royal Statistical Society, Series B* **69,** 1–30.

Zeng, D. L. and Lin, D. Y. (2010). A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. *Statistica Sinica* **20,** 871–910.

Zucker, D. M. (1988). Restricted mean life with covariates: Modification and extension of a useful survival analysis method. *Journal of the American Statistical Association* **93,** 702–709.