Analysis of marked recurrent events in the presence of a terminating event

by

Yu Ma

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2012

Doctoral Committee:

Associate Professor Douglas E. Schaubel, Chair
Professor Richard A. Hirth
Associate Professor Ananda Sen
Assistant Professor Wen Ye

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

CHAPTER I

Introduction

In many biomedical studies where the event of interest is recurrent (e.g., hospital admission), marks are observed upon the occurrence of each event (e.g., medical costs, length of stay). In Chapter II, we propose novel methods which contrast group-specific cumulative means, influenced by the recurrent event rate and survival probability. Our proposed methods utilize a form of hierarchical modeling: a proportional hazards model for the terminating event; a proportional rates model for the conditional recurrent event rate given survival; and a generalized estimating equations approach for the marks, given an event has occurred. Group-specific cumulative means are estimated (as processes over time) by averaging fitted values from the afore-listed models, with the averaging being with respect to the marginal covariate distribution. Large sample properties are derived, while simulation studies are conducted to assess finite sample properties. We apply the proposed methods to data obtained from the CANADA-USA Peritoneal Dialysis Study (CANUSA).

Typically in observational studies, it is necessary to account for measured and unmeasured heterogeneity across study subjects. This is often accomplished through model covariates (for measured factors) and frailty variates (to account for unmeasured predictors). In Chapter III, we investigate a frequently occurring data structure where the event of interest is recurrent (e.g., hospital admission), marks are observed

upon the occurrence of each event (e.g., length of stay) and the recurrent event process may be permanently stopped by a terminating event (e.g., death). Similar to Chapter II, the methods proposed in Chapter III model the terminating event hazard, conditional recurrent event rate given survival and the mark process. However, in Chapter III all such models are fully parametric, and estimation is carried out simultaneously. Most importantly, residual correlation across the terminating event, recurrent event and mark process is captured by a frailty variate, assumed to follow a Normal distribution. Maximum likelihood based estimation is carried out via a Gaussian Quadrature technique for integration. Through simulation, the methods are shown to work well for practical sample sizes. In contrast, significant biases are detected when estimation is based on methods which fail to account for residual correlation among the event, death and mark processes, especially when heterogeneity is large within the population.

In Chapter IV, we develop inverse weighting methods to contrast group-specific cumulative means. Both the underlying data structure and the target estimands are the same as those from Chapter II. However, for the methods proposed in Chapter IV, we avoid constructing semi-parametric or parametric models for each process to achieve consistent estimates. We further take into account of treatment imbalance and unobserved censoring times by combining Inverse Probability of Treatment Weighting (IPTW) and Inverse Probability of Censoring Weighting (IPCW). Efficiency is compared with the procedure proposed in Chapter II.

CHAPTER II

Semi-parametric methods contrasting group-specific cumulative mean

2.1 Introduction

In many biomedical studies, subjects may experience the event of interest multiple times. Examples include repeated hospital admissions, epileptic seizures, repeated use of illegal drugs, and others. Although chief interest often lies in modeling the recurrent event process, often investigators are also interested in an outcome measure associated with each recurrent event; for example, the medical cost incurred during each hospitalization, or the length of treatment during each hospital visit. Such outcome measures are defined as marks, and measure quantitative or qualitative aspects of each event occurrence. The marks may also be influenced by either internal or external covariates in the study. In the setting where the study population is heterogeneous, it is often of interest to know how subjects from different demographic groups exhibit certain characteristics over the study period, or how subjects from different populations respond to different treatments over the course of follow-up. One also needs to acknowledge the fact that the recurrent event process can be terminated permanently by a terminating event (e.g., death). Since this type of data structure is very common in biomedical studies, it is of much interest to establish accurate

methods to appropriately model the marked recurrent event process in the presence of a terminating event, while accounting for group-specific risk profile difference.

Although there is a rich body of work based on the recurrent/terminal event setting (e.g., Cook and Lawless, 1997; Li and Lagakos, 1997; Ghosh and Lin, 2000, 2002; Huang and Wang, 2004; Liu, Wolfe, and Huang, 2004; Ye, Kalbfleisch, and Schaubel, 2007), very few methods have been proposed to deal with a marked recurrent process which is subject to a terminating event. Ghosh and Lin (2000, 2002) proposed models for the marginal mean number of events, while Cook and Lawless (1997) developed models for the conditional recurrent event rate given survival. A latent variable (frailty) was introduced by Liu et al. (2004) and Ye et al. (2007), with the recurrent and terminating events assumed to be independent conditional on the frailty. In comparing group-specific cumulative means, it is often beneficial to understand the component processes; in particular, the relationship between the covariates and the terminating event process, the recurrent event process and the marks associated with each recurrent event. For example, a group's higher medical costs might be due to their longer survival, higher hospitalization rate or greater cost per hospital admission. Interesting group-specific differences with respect to any of these processes could go undetected if one focuses only on the overall mean.

We propose semiparametric methods that compare group-specific means which are viewed as processes over time. The proposed methods assume a proportional hazards model for the terminating event; a proportional rates model for the conditional recurrent event process given survival; and a generalized estimating equations model (Liang and Zeger, 1986) for the marks given the occurrence of recurrence event. Our estimator combines the fitted values from the above-listed models by computing the estimated cumulative mean for each subject, then averaging over all subjects. The covariate adjustments are allowed to vary for the terminating event process, recurrent event process and model for marks and for different groups. The baseline hazard for

the terminating event model and baseline rate for the recurrent event model are also allowed to be group specific. If any of the component models are of interest, different modeling strategies could be adopted.

The motivating example for our method is the Canada-USA (CANUSA) Peritoneal Dialysis Study, where interest lies in comparing group-specific means between American and Canadian peritoneal dialysis (PD) patients. The CANUSA study was a prospective cohort study of patients commencing continuous peritoneal dialysis in 14 centers in Canada and the United States. Between September 1, 1990 and December 31, 1992, a total of 679 patients were enrolled. There were 90 deaths, 130 transplants and 1,340 hospitalizations. Patient's length of each hospital visit was recorded and the number of hospitalizations varies between patients. Patients may experience death during the course of the follow up period. Hence, a patient's hospitalization experience is terminated by death. Besides knowing a patient's survival time (subject to right censoring) and hospitalization information, we also record factors that might influence the length of stay (e.g., serum albumin, normalized protein catabolic rate, subjective global assessment, percent lean body mass, Kt/V, creatinine clearance rate, country, gender, race).

In many studies, investigators are interested in estimating medical costs associated with each hospitalization. In the above example, the length of stay during each hospitalization can be treated as the mark. Length of stay could also be used as a surrogate for costs. In fact, our proposed method provides an important framework for analyzing cost data which is rarely discussed in the cost analysis literature. In terms of cost analysis, mean cost, cumulative cost, lifetime cost or restricted lifetime cost are of interest. Depending on the setting and objective of the study, nonparametric, marginal, conditional, or joint cost models have been developed (e.g., Lin, 2000; Huang, 2002, 2009; Liu, Conaway, Knaus and Bergin, 2008; Cai, Zeng, and Pan, 2010). Lin (2000) developed a proportional means regression model for cumu-

lative medical costs. The proposed inference procedures were semi-parametric and the method targets "study-duration" medical cost. Since study duration is a somewhat artificial time limit, and since the covariates might impact survival as well, the interpretation of such results may be inappropriate in terms of inference regarding true lifetime medical cost. Huang (2002) developed calibration regression methods to model lifetime medical cost and survival time jointly. The method postulates linear covariate effects on both lifetime medical cost and survival time, which can be measured on certain transformed cost and time scales, respectively. Liu et al. (2008) developed a four-part random effects model to analyze correlated medical cost data. The model of Liu et al. (2008) targets longitudinal costs, (e.g., the cost accumulation process), instead of total costs. The model also takes account of both the mean structure and inter-temporal correlation among longitudinal medical costs. In addition, the model of Liu et al. (2008) takes into consideration the presence of zero costs. In recent work of Cai, Zeng, and Pan (2010), costs are treated as marks, with each mark being associated with the occurrence of each recurrent event. Therefore, the method of Cai et al. (2010) could be classified as a conditional model for cost data.

We structure the remainder of the chapter as follows. In Section 2.2, we introduce the proposed model, followed by the estimation method. Asymptotic properties are listed in Section 2.3, with proofs of the theorems given in the Appendix. Simulation studies are reported in Section 2.4. An application of our approach is given in Section 2.5. In Section 2.6, we provide discussion and explore some future research areas.

2.2 Proposed Models and Estimating Methods

2.2.1 Notation and setup

We first establish the required notation. Let D_i denote the time of the terminating event for subject i, while C_i is the censoring time for subject i. We let

$X_i = \min(C_i, D_i)$ and let $Y_i(t) = I(X_i \geq t)$ be the at-risk indicator. We let the true terminating event process be represented by $N_i^{*D}(t)$ and we let $N_i^{*R}(t)$ equal the total number of recurrent events for subject i up to time t. The observed terminating event process and recurrent event process are $N_i^D(t) = I(D_i \leq t, D_i < C_i)$ and $N_i^R(t) = N_i^{*R}(t \wedge X_i)$, respectively. Notice that recurrent events do not occur after death. We define $G_i(t)$ as the mark for subject i at time t. It is assumed that marks only occur at time of a recurrent event. In addition, subjects are divided into groups (e.g., by treatment type, gender, diagnosis) and A_i is used to denote group for subject i. For ease of presentation, we consider the case where A_i is binary ($A_i = 0, 1$), although the proposed methods can accommodate more than two groups. We then set up the group indicator $A_{ij} = I(A_i = j)$. Each subject is characterized by a vector of covariates, Z_i.

Of interest for each subject is the cumulative mark

$$B_i^*(t) = \int_0^t G_i(u) dN_i^{*R}(u),$$

with mean $\mu_{ij}(t)$, defined as

$$\mu_{ij}(t) = E[B_i^*(t)|Z_i, A_i = j].$$

That is, $\mu_{ij}(t)$ represents the mean of $B_i^*(t)$ for a subject with covariate Z_i, under the hypothetical scenario where subject i is a member of group j. We wish to estimate $\mu_{ij}(t)$ using the observed data. Using properties of conditional expectations, we can

write:

$$\begin{aligned}\mu_{ij}(t) &= E\left[\int_0^t I(D_i > u)G_i(u)dN_i^{*R}(u)|Z_i, A_i = j\right] \\ &= \int_0^t S_{ij}(u)g_{ij}(u)dR_{ij}(u), \end{aligned} \tag{2.1}$$

where $S_{ij}(u) = P(D_i > u|Z_i, A_i = j)$, $g_{ij}(u) = E(G_i(u)|dN_i^{*R}(u) = 1, Z_i, A_i = j)$ and $dR_{ij}(u) = E(dN_i^{*R}(u)|Z_i, D_i > u, A_i = j)$.

To contrast group-specific means, we first define $\mu_j(t)$ as the cumulative mean for group j averaging across the marginal distribution of Z_i. That is, $\mu_j(t) = E[\mu_{ij}(t)] = E[E[B_i^*(t)|A_i = j, Z_i]]$ for $j = 0, 1$, where the outer expectation is taken with respect to the marginal distribution of Z_i. The quantities $\mu_0(t)$ and $\mu_1(t)$ are averaged with respect to the same covariate distribution. The difference in cumulative means between the two groups can be expressed as

$$\delta(t) = \mu_1(t) - \mu_0(t). \tag{2.2}$$

The goal here is to estimate $\delta(t)$, methods for which are proposed in the next subsection.

2.2.2 Estimation

We estimate $\mu_j(t)$ through $\mu_{ij}(t)$, and semiparametric models are assumed for each process in (3.1). The terminating event hazard is assumed to follow a proportional hazards model,

$$\lambda_{ij}(t) \equiv \lambda(t|Z_i, A_i = j) = \lambda_{0j}(t)\exp\{\beta_D^{'}Z_{iD}\}, \tag{2.3}$$

where β_D is a parameter vector, $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(s)ds$ is the group-specific cumulative baseline hazard, and Z_{iD} is a covariate vector comprised of appropriate elements of Z_i or functions thereof. Model (3.3) implies group-specific baselines and a regression parameter that is shared across groups. Note that group-specific regression coefficients are obtained through inclusion of group $\times$ covariate interactions. The estimator of β_D, denoted by $\hat{\beta}_D$, can be computed through the partial likelihood score function,

$$U^D(\beta_D) = \sum_{i=1}^{n}\sum_{j=0}^{1}\int_0^{\tau} A_{ij}\{Z_{iD} - \bar{Z}_j(t;\beta_D)\}dN_i^D(t), \tag{2.4}$$

where $\bar{Z}_j(t;\beta_D) = S_j^{(1)}(t;\beta_D)/S_j^{(0)}(t;\beta_D)$, with $S_j^{(k)}(t;\beta_D) = n^{-1}\sum_{i=1}^n A_{ij}Y_i(t)e^{\beta_D' Z_{iD}} Z_{iD}^{\otimes k}$ for k= 0, 1, 2, with $Z_{iD}^{\otimes 0} = 1$, $Z_{iD}^{\otimes 1} = Z_{iD}$ and $Z_{iD}^{\otimes 2} = Z_{iD}Z_{iD}'$. The baseline hazard is estimated through the Breslow-Aalen method,

$$\hat{\Lambda}_{0j}(t;\hat{\beta}_D) = n^{-1}\sum_{i=1}^{n}\int_0^t A_{ij}S_j^{(0)}(s;\hat{\beta}_D)^{-1}dN_i^D(s). \tag{2.5}$$

Having computed $\hat{\beta}_D$ and $\hat{\Lambda}_{0j}(t,\hat{\beta}_D)$, subject-specific survival can be estimated for each group via $\hat{S}_{ij}(t) = \exp\{-\hat{\Lambda}_{0j}(t;\hat{\beta}_D)\exp\{\hat{\beta}_D' Z_{iD}\}\}$.

Next, we assume a proportional rates model for the conditional recurrent event rate given survival, as

$$r_{ij}(t) \equiv E[dN_i^{*R}(t)|D_i > t, Z_i, A_i = j] = r_{0j}(t)\exp\{\beta_R' Z_{iR}\}, \tag{2.6}$$

where $R_{0j}(t) = \int_0^t r_{0j}(s)ds$ is the cumulative baseline rate, Z_{iR} is a covariate vector derived from elements of Z_i, and β_R is the regression parameter. The right hand side of equation (2.6) is the familiar proportional rates model (Lin et al., 2000) applied to the conditional event rate given survival (e.g., as in Liu et al., 2004; Schaubel and Cai, 2005; Ye at al., 2007). Then β_R and $R_{0j}(t)$ can be estimated through a process-based

analog of GEE, using the fact that the following equations

$$\sum_{i=1}^{n}\sum_{j=0}^{1}\int_0^{\tau} A_{ij}Z_{iR}\, dM_{ij}^R(s;\beta_R) \;=\; 0_{p\times 1} \tag{2.7}$$

$$\sum_{i=1}^{n}\int_0^{\tau} A_{ij} dM_{ij}^R(s;\beta_R) \;=\; 0 \tag{2.8}$$

have mean zero, where $M_{ij}^R(t;\beta_R) = N_i^R(t) - \int_0^t Y_i(s)\exp\{\beta_R' Z_{iR}\}dR_{0j}(s)$. The upper limit τ satisfies $P(X_i \geq \tau) > 0$ and is typically set to $\max\{X_i\}$. Solving equation (2.7) and (2.8), then reorganizing, $\hat{\beta}_R$ can be computed as the root of

$$U^R(\beta_R) = \sum_{i=1}^{n}\sum_{j=0}^{1}\int_0^{\tau} A_{ij}[Z_{iR} - \bar{Z}_j(s;\beta_R)]dN_i^R(s), \tag{2.9}$$

where $S_j^{(k)}(s;\beta_R) = n^{-1}\sum_{i=1}^n A_{ij}Y_i(s)Z_{iR}^{\otimes k}\exp\{\beta_R' Z_{iR}\}$ for $k = 0, 1, 2$, while $R_{0j}(t;\beta_R)$ can be estimated by its Breslow-Aalen analog,

$$\hat{R}_{0j}(t;\hat{\beta}_R) = n^{-1}\sum_{i=1}^{n}\int_0^{t} A_{ij}S_j^{(0)}(s;\hat{\beta}_R)^{-1}dN_i^R(s). \tag{2.10}$$

We assume the following model for the mark process,

$$g_{ij}(t) \equiv E[G_i(t)|dN_i^{*R}(t) = 1, Z_i, A_i = j] = g(t;\beta_G, Z_{iG}, A_i = j), \tag{2.11}$$

where Z_{iG} is a covariate vector consisting of elements of Z_i and an intecept term, and β_G is the corresponding parameter vector. The function g is a monotonic differentiable function of t, β_G, Z_{iG} and A_i.

In order to avoid distributional assumptions on $G_i(t)$, we estimate model (2.11) using GEE. We let $G_i = [G_i(T_{i1}), G_i(T_{i2}), ..., G_i(T_{iN_i^R})]' = [G_{i1}, G_{i2}, ..., G_{iN_i^R}]'$ and the corresponding mean vector given Z_{iG}, A_i and dN_i^{*R} by $g_{ij} = [g_{ij1}, g_{ij2}, ...g_{ijN_i^R}]'$, where

$N_i^R \equiv N_i^R(X_i)$. The covariance matrix of G_i is modeled as $V_i = \phi H_i^{1/2} R_i(\alpha) H_i^{1/2}$ where H_i is a $N_i^R \times N_i^R$ diagonal matrix with $v(g_{ijl})$ as the lth diagonal element with $l = 1, \ldots, N_i^R$. The variance function v is determined by the specific working probability distribution used for the marks, and ϕ is a dispersion parameter. The quantity $R_i(\alpha)$ is the correlation matrix of G_i; for instance, an exchangeable structure, $corr(G_{il}, G_{ik}) = \alpha$, for $l \neq k$. The working correlation matrix can be estimated through an iterative fitting process by using the current value of the parameter vector to compute the appropriate functions of the Pearson residual $\hat{e}_{ijl} = (G_{il} - \hat{g}_{ijl})[v(\hat{g}_{ijl})]^{-1/2}$. The generalized estimating equation is given by:

$$U^G(\beta_G) = \sum_{i=1}^{n} \sum_{j=0}^{1} \frac{\partial g_{ij}^{'}}{\partial \beta_G^{'}} A_{ij} V_i^{-1} [G_i - g_{ij}]. \tag{2.12}$$

We can obtain an estimate of β_G and V_i by the following iterative algorithm:

1. Compute an initial estimate of β_G assuming independence among (G_{il}, G_{ik}), for $l \neq k$

2. Compute the working correlation $R_i(\alpha)$ based on the standard residuals, the current β_G and the assumed structure of $R_i(\alpha)$

3. Get an estimate of the covariance V_i with $\hat{V}_i = \phi H_i^{1/2} \hat{R}_i(\alpha) H_i^{1/2}$

4. Update $\hat{\beta}_G$:

$$\hat{\beta}_G^{s+1} = \hat{\beta}_G^s - \left[\sum_{i=1}^{n} \sum_{j=0}^{1} \frac{\partial g_{ij}'}{\partial \beta_G'^s} A_{ij} \hat{V}_i^{-1} \frac{\partial g_{ij}}{\partial \beta_G^s} \right]^{-1} \left[\sum_{i=1}^{n} \sum_{j=0}^{1} \frac{\partial g_{ij}'}{\partial \beta_G'^s} A_{ij} \hat{V}_i^{-1} (G_i - g_{ij}) \right]$$

5. Iterate between steps 2-4 until convergence.

For example, for subject i, we specify $v(g_{ijl}) = g_{ijl}$ with j=0,1; $l = 1, \ldots N_i^R$; N_i^R=2;

ϕ=1. We specify an exchangeable working correlation matrix,

$$R_i(\alpha) = \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}$$

$$V_i = \begin{bmatrix} g_{ijl} & 0 \\ 0 & g_{ijl} \end{bmatrix}^{1/2} \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} g_{ijl} & 0 \\ 0 & g_{ijl} \end{bmatrix}^{1/2},$$

where α is estimated as $\hat{\alpha} = [(N^* - p)\phi]^{-1} \sum_{i=1}^{n} \sum_{l<k} \hat{e}_{ijl}\hat{e}_{ijk}$ with $N^* = 1/2 \sum_{i=1}^{n} N_i^R(N_i^R - 1)$, p being the number of regression parameters. Note that, under a working independence assumption, $R_i(\alpha)$ is replaced by the identity matrix. If ϕ is not known , it can be estimated as $\hat{\phi} = [N-p]^{-1} \sum_{i=1}^{n} \sum_{l=1}^{N_i^R} \hat{e}_{ijl}^2$ with $\hat{e}_{ijl} = (G_{il} - \hat{g}_{ijl})[v(\hat{g}_{ijl})]^{-1/2}$ and $N = \sum_{i=1}^{n} N_i^R$.

If the working correlation matrix is misspecified, that is $Cov(G_i) \neq V_i$, we still get a consistent estimator of β_G, and we can consistently estimate $Cov(\hat{\beta}_G)$ through the robust estimator

$$\begin{aligned} \Sigma &= \left[\sum_{i=1}^{n}\sum_{j=0}^{1} \frac{\partial g_{ij}'}{\partial \beta_G'} A_{ij} V_i^{-1} \frac{\partial g_{ij}}{\partial \beta_G}\right]^{-1} \left[\sum_{i=1}^{n}\sum_{j=0}^{1} \frac{\partial g_{ij}'}{\partial \beta_G'} A_{ij} V_i^{-1} Cov(G_i) V_i^{-1} \frac{\partial g_{ij}}{\partial \beta_G}\right] \\ &\quad \times \left[\sum_{i=1}^{n}\sum_{j=0}^{1} \frac{\partial g_{ij}'}{\partial \beta_G'} A_{ij} V_i^{-1} \frac{\partial g_{ij}}{\partial \beta_G}\right]^{-1}. \end{aligned} \quad (2.13)$$

In computing $\hat{\Sigma}$, β and α are replaced by their estimated values, while $Cov(G_i)$ is replaced by $[G_i - g_{ij}(\hat{\beta}_G)][G_i - g_{ij}(\hat{\beta}_G)]'$.

Finally, after estimating $\beta_D, \beta_R, \beta_G$, $R_{0j}(t)$ and $\Lambda_{0j}(t)$, we can estimate $\mu_j(t)$ by replacing the true parameters with their estimated counterparts, then averaging over

the empirical covariate distribution,

$$\hat{\mu}_j(t) = \frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_{ij}(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \hat{S}_{ij}(u)\hat{g}_{ij}(u)d\hat{R}_{ij}(u). \tag{2.14}$$

The group-specific difference, $\delta(t)$, is then estimated by

$$\hat{\delta}(t) = \hat{\mu}_1(t) - \hat{\mu}_0(t). \tag{2.15}$$

In the above estimation of $\delta(t)$, we allow β_D, β_R and β_G to be shared among different groups. In the event that the β_D, β_R and β_G are group-specific, we can develop similar estimation procedures from the following models:

$$\lambda_{ij}(t) = \lambda_{0j}(t)\exp\{\beta_{Dj}^{'}Z_{iD}\} \tag{2.16}$$

$$r_{ij}(t) = r_{0j}(t)\exp\{\beta_{Rj}^{'}Z_{iR}\} \tag{2.17}$$

$$g_{ij}(t) = g(t;\beta_{Gj}, Z_{iG}, A_i = j). \tag{2.18}$$

2.3 Asymptotic Properties

In this section, we summarize the essential asymptotic behavior of the proposed estimators by first listing the necessary conditions, for $i = 1, \ldots, n$ and j= 0, 1.

(a) $\{N_i^{*R}(.), D_i, C_i, Z_i, A_i, G_i(.)\}$ are independent and identically distributed

(b) $E[dN_i^{*R}(t)|D_i > t, Z_i, A_i, C_i > t]$=E$[dN_i^{*R}(t)|D_i > t, Z_i, A_i]$

(c) $E[G_i(t)|dN_i^{*R}(t) = 1, Z_i, A_i, C_i > t] = E[G_i(t)|dN_i^{*R}(t) = 1, Z_i, A_i]$

(d) $\lim_{\delta\to 0} Pr\{t \le D_i < t+\delta|D_i > t, Z_i, C_i > t\} = \lim_{\delta\to 0} Pr\{t \le D_i < t+\delta|D_i > t, Z_i\}$

(e) $Pr(Y_i(\tau) = 1) > 0$

(f) $\int_0^\tau d\Lambda_{0j}(t) < \infty$, $\int_0^\tau dR_{0j}(t) < \infty$ and $N_i^R(\tau) < \infty$

(g) Elements of Z_i are bounded almost surely.

(h) Postive-definiteness of the matrices, $A_j^D(\beta_D)$ and $A_j^R(\beta_R)$,

where

$$\begin{aligned}
A_j^D(\beta_D) &= E\left[\int_0^\tau \{Z_{iD} - \bar{z}_j(t;\beta_D)\}^{\otimes 2} Y_i(t) e^{\beta_D' Z_{iD}} d\Lambda_{0j}(t)\right] \\
A_j^R(\beta_R) &= E\left[\int_0^\tau \{s_j^{(2)}(t;\beta_R) s_j^{(0)}(t;\beta_R)^{-1} - \bar{z}_j(t;\beta_R)^{\otimes 2}\} s_j^{(0)}(t;\beta_R) dR_{0j}(t)\right] \\
\bar{z}_j(t;\beta) &= s_j^{(1)}(t;\beta) s_j^{(0)}(t;\beta)^{-1} \\
s_j^{(d)}(t;\beta) &= E[A_{ij} Y_i(t) Z_i^{\otimes d} \exp\{\beta' Z_i\}], d = 0, 1, 2.
\end{aligned}$$

Now, we describe the main asymptotic results from the proposed procedures. Proofs of the theorems are provided in the Appendix.

Theorem II.1. *Under conditions (a) to (h) and models (3.3),(3.6) and (3.11),* $\hat{\mu}_j$ *is a uniformly consistent estimator of* μ_j*. That is,* $\hat{\mu}_j(t)$ *converges to* $\mu_j(t)$ *almost surely for j=0,1 and* $t \in (0, \tau]$*. In addition,* $n^{1/2}\{\hat{\mu}_j(t) - \mu_j(t)\}$ *converges weakly to a zero-mean Gaussian process for j=0,1 and* $t \in (0, \tau]$ *with covariance function* $\sigma_j(s,t) = E[\psi_{ij}(s)\psi_{ij}(t)]$*, where*

$$\psi_{ij}(t) = \sum_{k=1}^{6} \psi_{ijk}(t)$$

$$
\begin{aligned}
\psi_{ij1}(t) &= -A_j^D(\beta_D)^{-1}U_{ij}^D(\beta_D) \times E\left[e^{\beta_D' Z_{iD}} \int_0^t S_{ij}(u^-|Z_{iD}) \int_0^u \{Z_{iD} - \bar{z}_j(r;\beta_D)\}' \right. \\
&\quad \left. \times d\Lambda_{0j}(r) g_{ij}(u|Z_{iG}) dR_{ij}(u|Z_{iR})\right] \\
\psi_{ij2}(t) &= -A_j^R(\beta_R)^{-1}U_{ij}^R(\beta_R) \times E\left[e^{\beta_R' Z_{iR}} \int_0^t S_{ij}(r^-|Z_{iD}) g_{ij}(r|Z_{iG}) \{Z_{iR} - \bar{z}_j(r;\beta_R)\}' dR_{0j}(r)\right] \\
\psi_{ij3}(t) &= E\left[\int_0^t S_{ij}(r^-|Z_{iD}) \frac{\partial g_{ij}'}{\partial \beta_G'} g_{ij}(r|Z_{iG}) dR_{ij}(r|Z_{iR})\right] E\left[\frac{\partial U_{ij}^G(\beta_G)}{\partial \beta_G'}\right]^{-1} U_{ij}^G(\beta_G) \\
\psi_{ij4}(t) &= -\int_0^t E[e^{\beta_D Z_{iD}} S_{ij}(r|Z_{iD}) g_{ij}(r|Z_{iG}) dR_{ij}(r|Z_{iR})] \int_0^r \frac{dM_{ij}^D(u;\beta_D)}{s_j^{(0)}(u;\beta_D)} \\
\psi_{ij5}(t) &= \int_0^t E[e^{\beta_R Z_{iR}} S_{ij}(r^-|Z_{iD}) g_{ij}(r|Z_{iG})] \frac{dM_{ij}^R(r;\beta_R)}{s_j^{(0)}(r;\beta_R)} \\
\psi_{ij6}(t) &= \mu_j(t|Z_i) - \mu_j(t)
\end{aligned}
$$

where we define

$$
\begin{aligned}
U_{ij}^D(\beta_D) &= \int_0^\tau A_{ij}\{Z_{iD} - \bar{z}_j(t;\beta_D)\} dM_{ij}^D(t;\beta_D) \\
U_{ij}^R(\beta_R) &= \int_0^\tau A_{ij}\{Z_{iR} - \bar{z}_j(t;\beta_R)\} dM_i^R(t;\beta_R) \\
U_{ij}^G(\beta_G) &= \partial g_{ij}'/\partial \beta_G' A_{ij} V_i^{-1}[G_i - g_{ij}].
\end{aligned}
$$

The proof of Theorem 1 unfolds through a series of Taylor expansions, several applications of the Strong Law of Large Numbers (Sen and Singer, 1993) and the Multivariate Central Limit Theorem (MCLT). A demonstration of tightness completes the proof of weak convergence using various results from empirical processes (Pollard 1990; van der Vaart and Wellner 1996; Bilias et al. 1997). In the next Theorem, we describe asymptotic results for $\hat{\delta}(t)$.

Theorem II.2. *Under conditions (a) to (h) and models (3.3),(3.6) and (3.11),* $\hat{\delta}(t)$ *is a uniformly consistent estimator of* $\delta(t)$*. That is,* $\hat{\delta}(t)$ *converges to* $\delta(t)$ *almost*

surely for j=0,1 and t∈ $(0, \tau]$. *Further,* $n^{1/2}\{\hat{\delta}(t) - \delta(t)\}$ *converges weakly to a zero-mean Gaussian process with covariance function* $\sigma_\delta(s,t) = E[\{\xi_{i1}(s) - \xi_{i0}(s)\}\{\xi_{i1}(t) - \xi_{i0}(t)\}]$.

It follows from Theorem 1 that $n^{1/2}\{\hat{\mu}_1(t) - \hat{\mu}_0(t)\}$ is asymptotically equivalent to $n^{-1/2}\sum_{i=1}^{n}\{\xi_{i1}(t) - \xi_{i0}(t)\}$, a scaled sum of zero-mean Normal variates. For fixed (s,t), convergence follows from the MCLT. Tightness follows from empirical processes theory. The covariance $\sigma_\delta(s,t)$ can be consistently estimated by replacing all limiting values with their empirical counterparts, then averaging across all subjects for $i = 1, \ldots, n$.

2.4 Simulation Study

The terminating event was generated from $\lambda_{ij}(t) = \lambda_{0j}\exp\{\beta_D' Z_i\}$, a proportional hazards model where $Z_i = (Z_{i1}, Z_{i2}, Z_{i3})'$. The Cox regression parameter was set to $\beta_D = (0.5, 0.3, 0.3)'$, with λ_{0j}=0.03 or 0.06. We let $Z_{i1} \sim Bernoulli(0.5)$, with Z_{i2} generated through the model $P(Z_{i2} = 1|Z_{i1})$=expit$\{h_1 Z_{i1}\}$ and $P(Z_{i3} = 1|Z_{i1}, Z_{i2})$=expit$\{h_2 Z_{i1} + h_3 Z_{i2}\}$, where $(h_1, h_2, h_3)' = (1,1,1)'$ and with expit(h)= $\exp\{h\}[1 + \exp\{h\}]^{-1}$. We generated recurrent events from the model $r_{ij}(t) = Q_i r_{0j}\exp\{\beta_R' Z_i\}$ by generating gap times between successive events as: $T_{i,k+1} = T_{i,k} - Q_i \log(U_{i,k})[r_{0j}\exp\{\beta_R' Z_i\}]^{-1}$ for $k = 1, \ldots, 50$, where each of the $U_{i,k}$ variates followed a Unif(0,1) distribution. The frailty, Q_i, could be considered an unmeasured predictor that is shared by all recurrent event times for the same subject. We let $Q_i \sim Gamma(\theta)$ where θ= 0.5 or 0.25 and represents Var(Q_i); note that $E(Q_i) = 1$. We set r_{0j}= 0.25 or 0.20, with the regression parameter set to $\beta_R = (0.5, 0.3, 0.3)'$. The model for the marks was as follows: $g_i(t) = \beta_{G_0} + \beta_{G_1} t + \beta_{G_2} Z_{i1} + \beta_{G_3} Z_{i2} + \beta_{G_4} Z_{i3}$, with $(\beta_{G_0}, \beta_{G_1}, \beta_{G_2}, \beta_{G_3}, \beta_{G_4})'$=$(1,1,1,1,1.5)'$. In the simulation, a working independence correlation structure was assumed.Censoring times were generated through

$\lambda_{ij}^C(t) = \lambda_{0j}^C \exp\{\beta_{D3} Z_{i3}\}$ and truncated at 60, which results in 30% $\sim$ 40% of censoring for different simulation configurations. The average observed number of recurrent events ranged from 2 to 4 per subject. Cumulative means were estimated at $t = 3, 6, 9, 12$, in order to give comparisons for early, middle and late follow up times. Sample sizes were set at $n = 400$ and 500 replicates were generated per configuration.

Table 2.1 provides results for the population average cumulative mean estimator for the various data configurations examined. For all data configurations, the estimated cumulative means are very close to the true values. The average asymptotic standard errors (ASE) agree well with the empirical standard deviations (ESD) and, correspondingly, the empirical coverage probabilities (CP) are close to the nominal value of 0.95.

In Table 2.2, we evaluated the proposed treatment effect estimator, $\hat{\delta}(t) = \hat{\mu}_1(t) - \hat{\mu}_0(t)$. The biases at different follow-up times are negligible compared to the true values. The ASEs also agree well with the ESDs, with the CP being generally quite close to the nominal value of 0.95.

2.5 Application

We applied our method to the CANUSA study to compare the estimated cumulative mean days hospitalized between U.S. versus Canadian peritoneal dialysis patients. All patients commencing continuous PD between September 1, 1990 and December 31, 1992 were eligible for the study. Demographic data recorded at enrollment included age, sex, race, functional status according to the Karnofsky score, underlying renal disease, insulin-dependent diabetes mellitus (IDDM), and history of cardiovascular disease (CVD). Estimates of nutritional status included subjective global assessment (SGA), protein catabolic rate (PCR) and percentage of lean body mass (PCTLBM). Adequacy of dialysis was estimated by measurement of total weekly Kt/V for urea, total weekly creatinine clearance (CCr), and serum beta-2-microglobulin(β2M). Death

served as the terminating event, while the recurrent event was hospitalization. Length of stay (number of days hospitalized) served as the mark for each hospital visit. Then, the natural interpretation of cumulative mark would be the total number of days hospitalized, which is our main interest.

A total of $n = 679$ patients were enrolled in the study. There were many more males (393) than females (286) enrolled. The mean age was 54, with a range of 18 to 82 years. Follow-up was terminated on December 31, 1993. There were 90 deaths, and the average number of days hospitalized across all patients was 7.8 days per patient. Of primary interest was to compare American and Canadian patients. Adjustment covariates included serum albumin (SALB), normalized protein catabolic rate (NPCR), subjective global assessment (SGA), percent lean body mass (PCTLBM), age, Karnofsky score and cardiovascular disease (CVD).

We first investigated the impact of each factor on the terminating event (e.g., mortality), recurrent event (e.g., hospitalization) and the marks (e.g., number of days hospitalized) based on the afore-listed covariates adjustment by fitting a proportional hazards model, a proportional rates model and a GEE model for the mark process, respectively. Table 2.3 gives the parameter estimates and p-values for the proportional hazards model, proportional rates model and a GEE linear regression model for the mark process. In the Canadian group, the only covariate that was significant for the terminating event model include percent lean body mass. A 1% increase in lean body mass is associated with 6% decrease in death hazard adjusting for other variables. In the American group, the covariates that were significant for the terminating event model include PCTLBM, age, CVD and SALB. For those patients who have a history of CVD, the mortality hazard is 2.27 times that of patients without CVD, adjusting for other covariates. A 1g/L increase in serum albumin concentration is associated with 6% decrease in the death hazard, covariate adjusted. Factors that are significant predictors of the hospital admission rate for either group include PCTLBM, Karnofsky

score, SALB and CVD. For the American patients, PCTLBM and age are significant in the model for marks, in addition to follow-up time. For the Canadian patients, only follow-up time is significant. Lower percentage of lean body mass, increased age and early hospitalization are associated with longer length of stay. Overall, percentage lean body mass and CVD are either significant or borderline significant in either the terminating event model or the recurrent event model for patients in both countries. Follow-up time at hospital admission is significant for both countries.

As evident in Figure 2.1, the estimated cumulative mean days hospitalized is slightly higher for Canadian patients than for American PD patients. This holds for most of the observed follow-up period, with the difference slightly decreasing with the increase in follow-up time. In addition, we estimated the difference in cumulative mean for American and Canadian PD patients at every half year interval and found that the difference tends to become increasingly non-significant as follow-up time increases (Table 2.4). Figure 2.2 gives the differences in cumulative mean between the patients in the two countries, and shows no trend during the first year, then a mild decreasing trend thereafter as time increases. The estimated cumulative mean difference stayed slightly above zero within the two year follow-up time window.

2.6 Discussion

We developed semiparametric methods to compare group-specific cumulative means associated with marked recurrent events in the presence of a terminating event. Our methods combine a proportional hazards model (Cox, 1972) for the terminating event, a proportional rates model (Lin et al, 2000) for the conditional recurrent event rate given survival and a GEE model for the marks. We estimated the terminating event hazard, recurrent event rate given survival and the marks given each recurrent event separately, integrating to get the estimated group-specific means. In our modeling, the parameters of interest are estimated separately. This is different from methods

that involve joint estimation or conditional estimation of the parameters of interest. In addition, we do not assume the covariates are necessarily shared by different models. The group-specific means are each averaging over the marginal covariate distribution, such that covariate imbalances are factored out.

Schaubel and Zhang (2010) developed a method which targets the same $\delta(t)$ of interest in our work. In the work of Schaubel and Zhang (2010), imbalances in the group-specific covariate distributions are adjusted through Inverse Probability of Treatment Weighting. The models of Schaubel and Zhang (2010) require that the censoring time be conditionally independent of the adjustment covariates, given group. The consistency of their estimators also requires that the logistic model for group assignment be correct. These assumptions would not be necessary in our approach. A potential advantage of Schaubel and Zhang (2010) is that the method does not require models for $\mu_j(t)$ or its components. That said, our proposed method does not require a model for $\mu_j(t)$ per se, and the models for the components would often be of interest to investigators.

In the causal inference literature, our measure could be interpreted as an average causal effect estimator. Chen and Tsiatis (2001) developed methods in estimating the average causal treatment difference in restricted mean lifetime. The causal treatment effect is defined through counterfactual random variables. In our modeling, the causal issues are not our main focus. Note that the appropriate application of the term causal depends on other properties of the observed data.

One potentially very useful variation of our methods would be joint modeling techniques. Cai et al.(2010) developed a semi-parametric proportional means model for marker data contingent on recurrent events. The authors used a marginal rate model, while the marks were analyzed through a proportional means model. The parameters for the marginal rate model and those of the proportional means model are linked by estimating equations through which the parameters are estimated jointly.

A terminating event was not considered. Another useful extension of our work would be incorporating shared latent variables, such that the recurrent event process would be subject specific and could allow for residual correlation between the terminating and recurrent event processes.

Table 2.1: Simulation Results: Performance of Proposed Estimator

λ_{00}, λ_{01}	r_{00}, r_{01}	$V(Q_i)$	t	% at risk	$\mu_1(t)$	Bias	ESD	ASE	CP
0.03/0.03	0.20/0.25	0.5	3	71%	6.98	0.04	0.77	0.74	0.92
			6	53%	16.01	0.10	1.75	1.72	0.95
			9	40%	26.12	0.27	2.92	2.95	0.96
			12	29%	36.66	0.35	4.31	4.38	0.96
0.03/0.03	0.20/0.25	0.25	3	71%	6.98	0.04	0.56	0.65	0.98
			6	53%	16.01	0.18	1.34	1.51	0.97
			9	40%	26.12	0.02	2.26	2.59	0.97
			12	29%	36.66	0.00	3.34	3.90	0.96
0.03/0.06	0.25/0.25	0.5	3	59%	6.23	0.02	0.66	0.67	0.95
			6	40%	12.84	0.02	1.50	1.43	0.93
			9	23%	18.98	0.11	2.42	2.28	0.92
			12	16%	24.38	0.22	3.31	3.18	0.94
0.03/0.06	0.25/0.25	0.25	3	59%	6.23	0.04	0.57	0.60	0.96
			6	40%	12.84	0.04	1.22	1.29	0.95
			9	23%	18.98	0.11	1.92	2.07	0.96
			12	16%	24.38	0.04	2.64	2.87	0.96
0.06/0.06	0.25/0.20	0.5	3	59%	4.99	0.03	0.57	0.56	0.94
			6	40%	10.27	0.13	1.18	1.19	0.95
			9	23%	15.19	0.19	1.82	1.87	0.96
			12	16%	19.50	0.20	2.47	2.58	0.96
0.06/0.06	0.25/0.20	0.25	3	59%	4.99	0.00	0.45	0.49	0.97
			6	40%	10.27	0.05	0.98	1.06	0.95
			9	23%	15.19	0.06	1.53	1.67	0.96
			12	16%	19.50	0.10	2.09	2.35	0.96

Notes: $\beta_D = (0.5, 0.3, 0.3)'$, $\beta_R = (0.5, 0.3, 0.3)'$, $\beta_G = (1, 1, 1, 1, 1.5)'$. Number of repetitions: 500. Number of subjects: $n = 400$.

Table 2.2: Simulation Results: Treatment effects

λ_{00}, λ_{01}	r_{00}, r_{01}	$V(Q_i)$	t	% at risk	$\delta(t)$	Bias	ESD	ASE	CP
0.03/0.03	0.20/0.25	0.5	3	71%	1.40	0.01	0.86	0.80	0.94
			6	53%	3.20	0.07	1.88	1.92	0.96
			9	40%	5.22	0.14	3.16	3.41	0.96
			12	29%	7.33	0.07	4.79	5.20	0.96
0.03/0.03	0.20/0.25	0.25	3	71%	1.40	0.04	0.74	0.76	0.97
			6	53%	3.20	0.06	1.62	1.83	0.97
			9	40%	5.22	-0.05	2.73	3.23	0.97
			12	29%	7.33	-0.25	4.04	4.94	0.98
0.03/0.06	0.25/0.25	0.5	3	59%	0.75	-0.03	0.90	0.87	0.93
			6	40%	3.17	-0.07	1.94	2.02	0.95
			9	23%	7.14	-0.19	3.30	3.47	0.95
			12	16%	12.29	-0.59	4.78	5.15	0.96
0.03/0.06	0.25/0.25	0.25	3	59%	0.75	-0.08	0.82	0.82	0.95
			6	40%	3.17	-0.08	1.80	1.93	0.97
			9	23%	7.14	-0.16	2.97	3.32	0.97
			12	16%	12.29	-0.05	4.35	4.92	0.97
0.06/0.06	0.25/0.20	0.5	3	59%	1.25	-0.03	0.87	0.78	0.91
			6	40%	2.57	-0.13	1.74	1.73	0.93
			9	23%	3.80	-0.06	2.76	2.84	0.96
			12	16%	4.88	0.11	3.83	4.02	0.96
0.06/0.06	0.25/0.20	0.25	3	59%	1.25	0.09	0.74	0.75	0.94
			6	40%	2.57	0.16	1.59	1.66	0.96
			9	23%	3.80	0.25	2.53	2.71	0.96
			12	16%	4.88	0.29	3.52	3.84	0.96

.

Notes: $\beta_D = (0.5, 0.3, 0.3)'$, $\beta_R = (0.5, 0.3, 0.3)'$, $\beta_G = (1, 1, 1, 1, 1.5)'$.Number of repetitions: 500. Number of Subjects: $n = 400$

Table 2.3: Analysis of CANUSA data: Parameter estimates for terminating event model, recurrent event model and model for marks

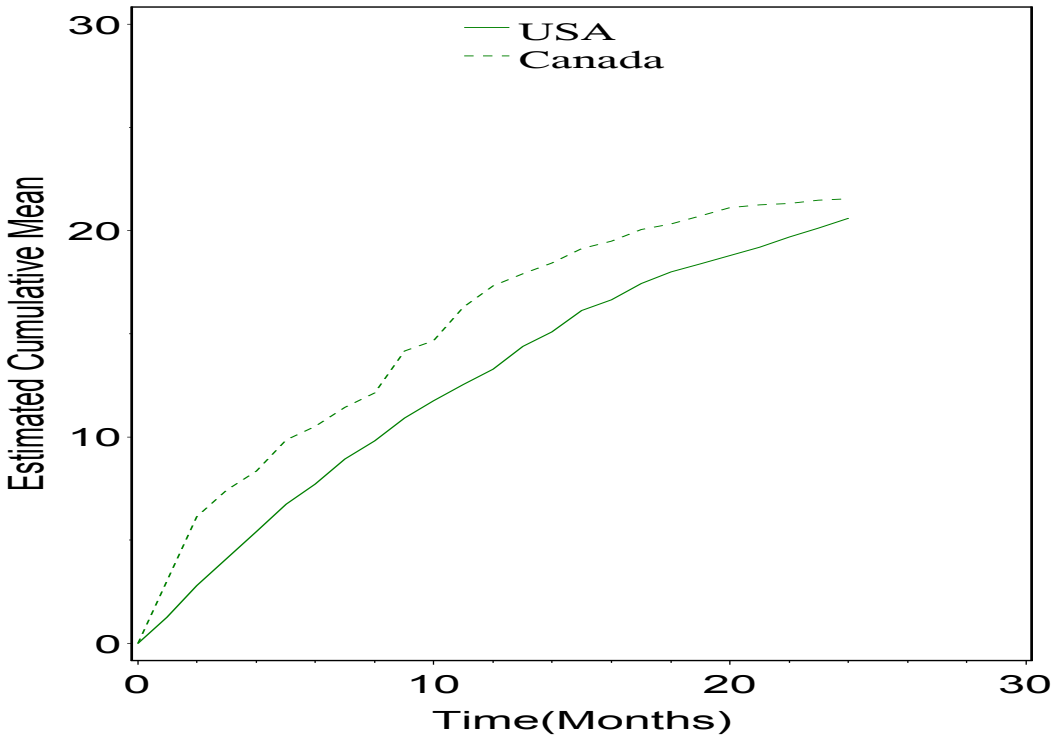
	$\lambda_{ij}(t)$			$r_{ij}(t)$			$g_{ij}(t)$	
Variable	$\hat{\beta}_D$	p	$\exp\{\hat{\beta}_D\}$	$\hat{\beta}_R$	p	$\exp\{\hat{\beta}_R\}$	$\hat{\beta}_G$	p
CANADA								
PCTLBM	-0.06	0.03	0.94	-0.02	0.02	0.98	-0.03	0.61
AGE	0.01	0.71	1.01	-0.01	0.33	0.99	0.00	0.98
KARNOF	-0.02	0.19	0.98	-0.02	0.01	0.98	-0.13	0.03
CVD	0.84	0.12	2.32	0.56	0.01	1.75	0.44	0.76
SALB	0.01	0.91	1.01	0.00	0.90	1.00	0.10	0.50
NPCR	-0.94	0.50	0.39	-0.09	0.83	0.92	1.17	0.68
SGA	0.08	0.62	1.09	-0.02	0.76	0.98	0.58	0.28
TIME	-	-	-	-	-	-	-0.30	<0.001
USA								
PCTLBM	-0.02	0.05	0.98	-0.01	0.16	0.99	-0.08	0.00
AGE	0.02	0.04	1.02	0.00	0.42	1.00	0.07	0.01
KARNOF	0.01	0.15	1.01	0.00	0.29	1.00	0.00	0.90
CVD	0.82	0.00	2.27	0.15	0.14	1.12	1.31	0.11
SALB	-0.07	0.01	0.94	-0.02	0.02	0.98	-0.13	0.15
NPCR	-0.29	0.55	0.75	0.22	0.25	1.24	1.51	0.37
SGA	0.01	0.93	1.01	-0.01	0.85	0.99	-0.18	0.58
TIME	-	-	-	-	-	-	-0.27	<0.001

.

Notes: $\lambda_{ij}(t)$: Survival model $r_{ij}(t)$: Recurrent event model $g_{ij}(t)$: marks model
p: p-value

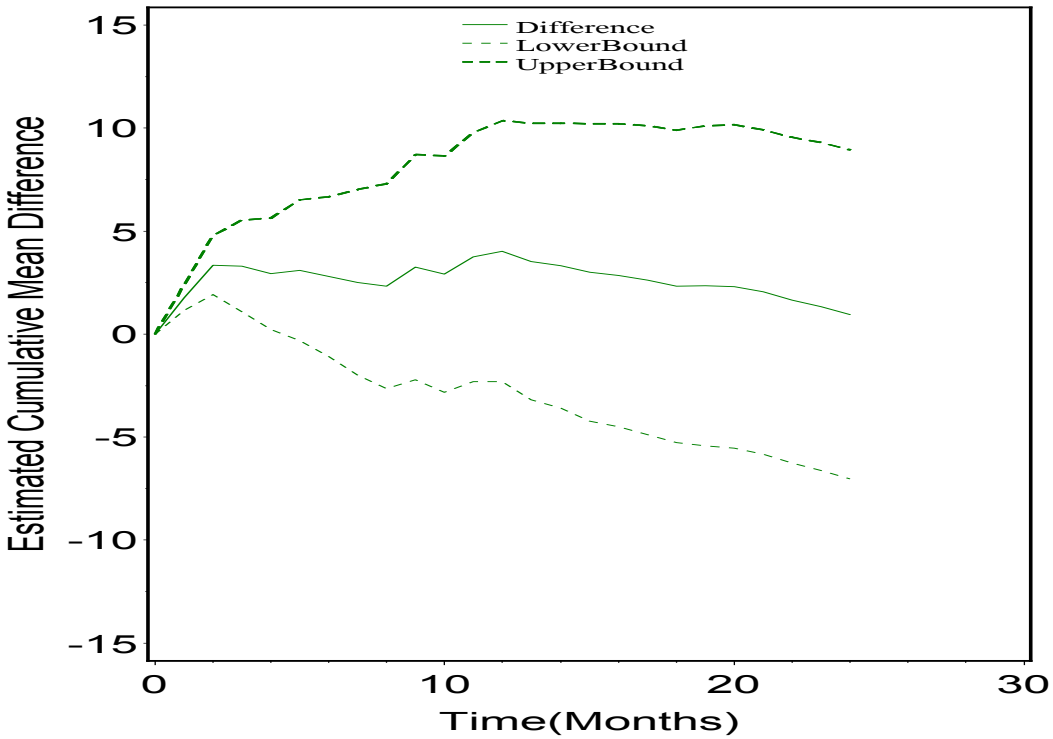
Table 2.4: Analysis of CANUSA data: Difference in length of hospitalization between American and Canadian patients over time (Canadian minus American)

Months	$\hat{\delta}(t)$	$\hat{SE}\{\hat{\delta}(t)\}$	P-value
6	2.79	1.98	0.17
12	4.02	3.23	0.22
18	2.32	3.87	0.56
24	0.95	4.07	0.82



The solid line represents American ($j = 0$) PD patients, and dashed line represents Canadian ($j = 1$) PD patients.

Figure 2.1: Analysis of the CANUSA data: Estimated cumulative mean number of days hospitalized for American and Canadian PD patients over time (measured in Months).



The solid line represents the estimated cumulative mean difference in days hospitalized (Canadian minus American). The dashed lines represent the corresponding 95% confidence intervals.

Figure 2.2: Analysis of the CANUSA data: point estimates and 95% confidence intervals for the estimated cumulative mean difference in days hospitalized $\hat{\delta}(t) = \hat{\mu}_1(t) - \hat{\mu}_0(t)$, between American ($j = 0$) and Canadian ($j = 1$) PD patients

CHAPTER III

Joint modeling of marked recurrent events in the presence of a terminating event

3.1 Introduction

In public health and medicine, the event of interest is often recurrent (e.g., repeated tumor occurrences, a series of hospitalizations). In recent decades, the study of such recurrent event process has drawn more and more attention. Often, the recurrent events have certain descriptive characteristics (e.g., size of tumor, or length of stay in each hospital visit). We refer to such these outcome measures associated with the recurrent events as "marks". In a sense, the mark characterizes each recurrent event. In clinical trials or epidemiologic studies, it is often the case that study subjects may die during the course of follow-up. Accurate analysis of recurrent events in this type of setting should allow for the fact that a terminating event may occur.

We further acknowledge the fact that both measured and unmeasured heterogeneity exists across study subjects, especially in observational studies. To account for measured heterogeneity, various regression models can be developed by including important adjustment factors as covariates. On the other hand, unmeasured het-

erogeneity could be explained via a "frailty" variate, an approach which has gained increasing popularity in recent decades, especially in the area of survival analysis. The purpose of a frailty (or random effect) is to represent unobserved heterogeneity in the model; which is appropriate if there is reason to believe that subjects with the same covariate value may have unequal risk of death or relapse of a particular disease. The term frailty was introduced by Vaupel, Manton and Stallard (1979) in the univariate survival analysis setting and was extended to the multivariate survival setting by Clayton (1978). In the context of the recurrent/terminal event setting, various methods have been developed to jointly model the dependence between the recurrent event and terminal event process using a frailty, e.g., Wang, Qin and Chiang, 2001; Huang and Wang, 2004; Liu, Wolfe and Huang, 2004; Ye, Kalbfleisch and Schaubel, 2007. By incorporating the frailty, we can potentially allow for residual correlation beyond what can be captured by the adjustment covariates. We investigate a type of data structure where the subjects may experience a sequence of recurrent events. These recurrent events are only observable if the study subjects are alive at each recurrent event occurrence. In other words, the recurrent event process may be stopped by the terminating event. The marks of these subjects are only observed if they experience each recurrent event. In Chapter III, interest lies not only in the individual processes (e.g., recurrent event, terminating event, mark), but also in the dependence structure among these processes; where the dependence is not only between the recurrent and terminal event but also between the recurrent/terminal event and the outcome measure associated with each recurrent event.

We propose joint modeling techniques targeted at the setting where, in addition to covariate effects, interest also lies in the heterogeneity within the study population, as well as the shape of the baseline death hazard and recurrent event rate functions. The proposed methods assume a proportional hazards model for the terminating event; a proportional rates model for the conditional recurrent event process given survival;

and a Poisson regression model for the marks given the occurrence of recurrence event. The baseline hazard is chosen to be piecewise constant in the assumed proportional hazards model. Similarly, the baseline rate for the recurrent event model is also assumed to be piecewise constant. Since the marks are generally non-negative, the Poisson distribution would appear to be a reasonable choice. In addition, we assume the frailty is at the subject level and applies to all three afore-listed models. The variance of the frailty variate reflects the degree of residual correlation among the death, recurrent event and mark process.

The application for our proposed models is the well-known Canada-USA (CANUSA) Peritoneal Dialysis Study(Canada-USA (CANUSA) Peritoneal Dialysis Study Group, 1996; Churchill, 1998; Ye, Kalbfleisch and Schaubel, 2007). The CANUSA Study Group (1996) focused on evaluating the relationship of adequacy of dialysis and nutritional status to mortality and morbidity. Churchill (1998) targeted the implication of adequacy of dialysis on peritoneal dialysis schedule. Ye, Kalbfleisch and Schaubel (2007) utilized the CANUSA study data to model the relationship between the hospitalization rates and failure of peritoneal analysis. One of our interests in this Chapter is the hospital admission or number of days hospitalized, which has rarely been presented before. Duration of hospital stay serves as the marks, which are only observed when a hospital admission occurs. The survival information (e.g., death, censoring) underlies the proportional hazards model and the PD patients' hospitalization times serves as the foundation of our recurrent event model, which is interpreted as being conditional on survival. Some of the clinical and non-clinical factors that are present in the above models are as follows: serum albumin, normalized protein catabolic rate, subjective global assessment, percent lean body mass, Kt/V, creatinine clearance rate, country, gender and race, Karnofsky score, cardiovascular disease and underlying cause of renal disease.

We structure the remainder of the chapter as follows. In Section 3.2, we introduce

the proposed proportional hazards model, proportional rates model and Poisson regression model, followed by the Gaussian quadrature estimation method. We conduct simulation studies in Section 3.3. In Section 3.4, we illustrate our proposed methods by applying them to the CANUSA study. In Section 3.5, we conclude this chapter with some discussion and future research areas.

3.2 Proposed Models and Estimating Methods

3.2.1 Notation and Setup

We first establish the required notation. Let D_i denote the time of the terminating event for subject i, while C_i is the censoring time. The quantity, $X_i = \min\{D_i, C_i\}$, is the observation time. Then, we denote $N_i^{*R}(t)$ as the total number of recurrent events for subject i up to time t. In addition to the terminating event, we introduce T_{ik}, a sequence of recurrent event times for suject i, $i = 1, \ldots, n$; $k = 1, \ldots, n_i$. Let $\delta_i = I(D_i \leq C_i)$ and $\Delta_{ik} = I(T_{ik} \leq X_i)$, indicators for subject i's terminating event and recurrent event time, respectively. Last, we introduce the event for the marks. Notice that recurrent events do not occur after death and marks are only observed at time of each uncensored recurrent event. Therefore it is implied that there is no information on the mark process available after the subject is dead or at times when the subject does not experience a recurrent event. Define $G_i(t)$ as the observed mark for subject i at time t. Often times, the marks are non-negative (e.g., number of days hospitalized, cell counts for each diagnosis) and, therefore, we assume $G_i(t) \geq 0$. The covariate vector is denoted by $Z_i = \{Z_{iD}, Z_{iR}, Z_{iG}\}$, where Z_{iD}, Z_{iR} and Z_{iG} represent the covariate vectors for the terminating event model, recurrent event model and the mark model respectively.

Collecting the previously-listed assumptions and conditions, the underlying data structure and proposed approach can be summarized as follows. The terminating

event process and the recurrent event process are correlated for the same subject through a mutual correlation with the covariate, Z_i, and also through a common random effect, w_i. The recurrent events only occur while the subject is alive. The recurrent event and the marks are also correlated for the same subject through the same random effect, w_i. The marks are only observed when the subject experiences a recurrent event. The marks are non-negative and follow a Poisson distribution. For both the terminating event hazard and recurrent event rate, the overall shape does not have to be imposed in advance; although constancy is assumed within pre-defined sub-intervals of time. If the follow-up time sub-intervals are chosen to be sufficiently small, one should obtain a fairly close approximation of the true underlying baseline rate and hazard functions.

3.2.2 Proposed Models

Based on the above assumptions, we establish the following piecewise proportional hazards model for the terminating event hazard. Let $a_0 = 0 < a_1 < ..... < a_{K-1} < a_K = \infty$ be a partition of the follow-up time. The hazard for the ith subject is given by

$$d\Lambda_i(t) = d\Lambda(t) \exp\{\beta_D^{'} Z_{iD} + w_i\}, \tag{3.1}$$

where $d\Lambda(t) = \sum_{k=1}^{K} \lambda_k I(a_{k-1} \leq t < a_k)$ is the baseline hazard function. The likelihood contribution for subject i with respect to the death process is given by

$$\begin{aligned} L_i^D &= d\Lambda_i(X_i)^{\delta_i} \exp\{-\Lambda_i(X_i)\} \\ &= \left[\sum_{k=1}^{K} \lambda_k I(a_{k-1} \leq X_i < a_k) \exp\{\beta_D^{'} Z_{iD} + w_i\}\right]^{\delta_i} \\ &\quad \times \exp\left\{-\int_0^{X_i} \exp\{\beta_D^{'} Z_{iD} + w_i\} \sum_{k=1}^{K} \lambda_k I(a_{k-1} \leq t < a_k) dt\right\}. \end{aligned} \tag{3.2}$$

Then, we establish the following proportional rates model for the conditional recurrent event rate given survival. Let $b_0 = 0 < b_1 < ..... < b_{Q-1} < b_Q = \infty$ be another partition of the follow-up of time axis, noting that it is not required to be the same partition given before (3.1). The recurrent event rate for the ith subject is then given by

$$\begin{aligned} dR_i(t) &= E[dN_i^{*R}(t)|D_i \geq t, Z_{iR}, w_i] \\ &= dR(t)\exp\{\beta_R^{'} Z_{iR} + w_i\}, \end{aligned} \tag{3.3}$$

where $dR(t) = \sum_{q=1}^{Q} r_q I(b_{q-1} \leq t < b_q)$ is the baseline recurrent event rate. Note that the recurrent event rate here is a conditional quantity; the conditional event rate, given survival. This expression resembles that of Ye, Kalbfleisch and Schaubel (2007) and Pan and Schaubel (2009). The likelihood contribution for subject i with respect to the recurrent event process is given as

$$\begin{aligned} L_i^R &= \prod_{k=1}^{n_i} \left[\sum_{q=1}^{Q} r_q I(b_{q-1} \leq T_{ik} < b_q) \exp\{\beta_R^{'} Z_{iR} + w_i\} \right]^{\Delta_{ik}} \\ &\quad \times \exp\left\{ - \int_0^{X_i} \exp\{\beta_R^{'} Z_{iR} + w_i\} \sum_{q=1}^{Q} r_q I(b_{q-1} \leq t < b_q) dt \right\}. \end{aligned} \tag{3.4}$$

Finally, we assume a Poisson regression model for the marks as follows,

$$\log[E\{G_i(t)|dN_i^{*R}(t) = 1, Z_{iG}(t), w_i\}] = \beta_G^{'} Z_{iG}(t) + w_i, \tag{3.5}$$

where $Z_{iG} = Z_{iG}(0)$, and $Z_i(t)$ is comprised of elements of $Z_i(0)$ and parametric functions of t. With respect to the mark distribution, the likelihood contribution for the ith subject equals

$$L_i^G = \exp\{G_i(t)(\beta_G^{'} Z_{iG}(t) + w_i)\} \exp\{-\exp\{\beta_G^{'} Z_{iG}(t) + w_i\}\}. \tag{3.6}$$

After combining the above three models, the likelihood for the ith subject is given by

$$L_i = \int L_i^D L_i^R L_i^G f_{w_i}(w_i) dw_i, \tag{3.7}$$

when the integral is over the support of w_i; or,

$$L_i = \int \exp\{\ell_i^D + \ell_i^R + \ell_i^G\} f_{w_i}(w_i) dw_i, \tag{3.8}$$

with

$$\begin{aligned} \ell_i^D &= \delta_i \left[\log \left\{ \sum_{k=1}^{K} \lambda_k I(a_{k-1} \leq X_i < a_k) \right\} + \beta_D^{'} Z_{iD} + w_i \right] \\ &\quad - \int_0^{X_i} \exp\{\beta_D^{'} Z_{iD} + w_i\} \sum_{k=1}^{K} \lambda_k I(a_{k-1} \leq t < a_k) dt, \end{aligned} \tag{3.9}$$

$$\begin{aligned} \ell_i^R &= \sum_{k=1}^{n_i} \left[\Delta_{ik} \log \left\{ \sum_{q=1}^{Q} r_q I(b_{q-1} \leq T_{ik} < b_q) \right\} + \beta_R^{'} Z_{iR} + w_i \right] \\ &\quad - \int_0^{X_i} \exp\{\beta_R^{'} Z_{iR} + w_i\} \sum_{q=1}^{Q} r_q I(b_{q-1} \leq t < b_q) dt, \end{aligned} \tag{3.10}$$

and

$$\ell_i^G = G_i(t)(\beta_G^{'} Z_{iG}(t) + w_i) - \exp\{\beta_G^{'} Z_{iG}(t) + w_i\}. \tag{3.11}$$

3.2.3 Proposed estimation methods

Several approaches have been adopted to estimate frailty models. Nielsen et al. (1992) adopted an EM algorithm by treating the frailties as unobserved quantities or missing values. However, the EM algorithm tends to converge slowly, and standard errors of the estimates can not be obtained directly. Another approach is the Penalized

Partial Likelihood (PPL), proposed by McGilchrist et al. (1991). The PPL algorithm is fast, but no standard error estimate of the frailty variance is given. Since our proposed method would involve evaluating complex integrals with joint frailties, the PPL method is not practical in our setting. To use the EM algorithm, conditional expectations of the random effects given the observed data need to be calculated and usually they do not have closed forms. In our setting, under the joint frailty models given above, the conditional expectations of the normal frailty do not have a closed form. Markov chain Monte Carlo (MCMC) methods (e.g., Metropolis-Hastings) could be adopted as in Liu, Wolfe and Huang (2004), but tend to be very computational expensive. Other options for fitting frailty models include Gaussian Quadrature, Laplace approximation (Tierney and Kadane, 1986) and the partial quasi-likelihood developed by Breslow and Clayton (1993). Since it is computationally fast and easy to implement using standard statistical software, we will use Gaussian Quadrature for our estimation.

Gaussian quadrature is well suited to numerically evaluate integrals against probability measures. Suppose we have the following integral of interest,

$$\int f(x)p(x)dx \approx \sum_{j=1}^{n} s_j f(x_j), \tag{3.12}$$

where $p(x)$ is the probability density function and $f(x)$ is a function of interest. Let s_j be the quadrature weights and x_j the integration points. The Gaussian quadrature chooses integration points in areas of high density.

For example, equation (4.8) can be approximated by weighted averages of U predetermined quadrature points s_u over random effect $w_i, u = 1, \ldots, U$ as follows,

$$L_i \approx \sum_{u=1}^{U} \exp\{\hat{\ell_i}^D + \hat{\ell_i}^R + \hat{\ell_i}^G\}\theta_u f_{w_i}(s_u), \tag{3.13}$$

where θ_u are the weights and

$$\begin{aligned}\hat{\ell}_i^{\,D} &= \delta_i \left[\log \left\{ \sum_{k=1}^{K} \hat{\lambda}_k I(a_{k-1} \leq X_i < a_k) \right\} + \hat{\beta}_D' Z_{iD} + s_u \right] \\ &\quad - \int_0^{X_i} \exp\{\hat{\beta}_D' Z_{iD} + s_u\} \sum_{k=1}^{K} \hat{\lambda}_k I(a_{k-1} \leq t < a_k) dt; \end{aligned} \tag{3.14}$$

$$\begin{aligned}\hat{\ell}_i^{\,R} &= \sum_{k=1}^{n_i} \left[\Delta_{ik} \log \left\{ \sum_{q=1}^{Q} \hat{r}_q I(b_{q-1} \leq T_{ik} < b_q) \right\} + \hat{\beta}_R' Z_{iR} + s_u \right] \\ &\quad - \int_0^{X_i} \exp\{\hat{\beta}_R' Z_{iR} + s_u\} \sum_{q=1}^{Q} \hat{r}_q I(b_{q-1} \leq t < b_q) dt; \end{aligned} \tag{3.15}$$

$$\hat{\ell}_i^{\,G} = G_i(t)(\hat{\beta}_G' Z_{iG}(t) + s_u) - \exp\{\hat{\beta}_G' Z_{iG}(t) + s_u\}. \tag{3.16}$$

The software used in our study is SAS (version 9.2). A NLMIXED procedure is used to carry out the intergral approximation and estimation. PROC NLMIXED selects the number of quadrature points adaptively by evaluating the log-likelihood function at the starting values of the parameters until two successive evaluations have a relative difference less than the value of the pre-defined tolerance number. The quadrature weights then will be determined based on the number of quadrature points. An empirical Bayes estimate of the random effect w_i, $\hat{w}_i$, is further computed such that the negative of the log-likelihood function is minimized based on the current vector of parameters. Then different optimizations techinques can be utilized to get an estimate of the parameter estimates (i.e., Dual quasi-newton, Newton-Raphson).

3.3 Simulation Study

To study the finite sample properties of the proposed estimator, we first assume frailty w_i to be normally distributed with mean 0 and variance σ^2, with σ^2=0.01, 0.0625 or 0.25 to examine the scenarios where there is relatively low correlation, moderate correlation and high correlation, respectively. The terminating event was generated from $d\Lambda_i(t) = d\Lambda(t)\exp\{\beta_D' Z_i + w_i\}$, a proportional hazards model where $Z_i = (Z_{i1}, Z_{i2}, Z_{i3})'$. The Cox regression parameter was set to $\beta_D = (0.5, 0.3, 0.3)'$, with $d\Lambda(t) = \sum_{k=1}^{2} \lambda_k I(a_{k-1} \leq t < a_k)$, where $\lambda_1 = 0.06, \lambda_2 = 0.03$, $a_0 = 0, a_1 = 10$ and $a_2 = \max\{X_i\}$. We let $Z_{i1} \sim Bernoulli(0.5)$, with Z_{i2} generated through the model $P(Z_{i2} = 1|Z_{i1})$=expit$\{h_1 Z_{i1}\}$ and $P(Z_{i3} = 1|Z_{i1}, Z_{i2})$=expit$\{h_2 Z_{i1} + h_3 Z_{i2}\}$, where $(h_1, h_2, h_3)' = (1, 1, 1)'$ and expit(h)=$\exp\{h\}[1 + \exp\{h\}]^{-1}$. We generated recurrent events from the model $dR_i(t) = dR(t)\exp\{\beta_R' Z_i + w_i\}$ by generating gap times between successive events as: $T_{i,k+1} = T_{i,k} - \log(U_{i,k})[dR(t)\exp\{\beta_R' Z_i + w_i\}]^{-1}$ for $k = 1, \ldots, 50$, where each of the $U_{i,k}$ variates followed a Unif(0,1) distribution. For simplicity, we let $dR(t) = \sum_{q=1}^{2} r_q I(b_{q-1} \leq t < b_q)$, with $r_0 = r_1 = 0.20$, $b_0 = 0, b_1$=median$\{T_{ik}\}$ and $b_2 = \max\{T_{ik}\}$ so that the piecewise constant assumption is still satisfied. The regression parameter was set to $\beta_R = (0.5, 0.3, 0.3)'$. The model for the marks was as follows: $\log[E\{G_i(t)|dN_i^{*R}(t) = 1, Z_i(t), w_i\}] = \beta_G' Z_i + w_i + 2t$, with $\beta_G = (0.5, 0.3, 0.3)'$. Censoring times were generated from a Unif(0,30) which resulted in about 30% of censoring. The average observed number of recurrent events is about three per subject. We used five quadrature points in our estimations. Sample sizes were set at $n = 2000$ and 2000 replicates were generated for each data configuration.

Table 3.1 provides performance of the parameters under very low correlation ($\sigma^2 = 0.01$) among the terminating events, recurrent events and marks. Since the correlation is very small (close to independence), estimation with or without considering the frailty w_i, both achieved fairly good performance in terms of bias and coverage probability. For all parameters, the estimated values are very close to the true values.

The average asymptotic standard errors (ASE) agree well with the empirical standard deviations (ESD) and, correspondingly, the empirical coverage probabilities (CP) are close to the nominal value of 0.95.

In Table 3.2, the correlation among the terminating events, recurrent events and marks was increased to be moderately large ($\sigma^2 = 0.0625$). In this setting, not considering the correlation resulted in increased bias and worse coverage. If we consider the frailty effect, the biases are negligible and the ASEs agree well with the ESDs with CPs generally being close to the nominal value of 0.95.

Last, we examine the results from Table 3.3 where correlation among the terminating events, recurrent events and marks was quite large ($\sigma^2 = 0.25$). The variance contributed by the frailty are over half of the total variance explained. Under this scenario, the estimation procedure including the frailty still outperforms the estimation procedure without frailty by a large extent. We can see from Table 3.3 that without considering the frailty resulted in highly elevated biases for most of the parameters and the variances of the parameter estimators are underestimated, yielding much lower CPs.

3.4 Application

We applied our methods to data obtained the previously-described CANUSA Peritoneal Dialysis Study. In this analysis, all patients commencing continuous PD between September 1, 1990 and December 31, 1992 were eligible for the study. Demographic data recorded at enrollment included age, sex, race, functional status (according to the Karnofsky score), underlying renal disease, insulin-dependent diabetes mellitus (IDDM), and history of cardiovascular disease (CVD). Estimates of nutritional status included subjective global assessment (SGA), and percentage of lean body mass (PCTLBM). Adequacy of dialysis was estimated by measurement of total weekly Kt/V for urea, total weekly creatinine clearance (CCr), and serum

beta-2-microglobulin(β2M). Death served as the terminating event, while the recurrent event was hospitalization. Length of stay (number of days hospitalized) served as the mark for each hospital visit.

A total of $n = 679$ patients were enrolled. There were many more males (393) than females (286) enrolled in the study. The mean age was 54, with a range of 18 to 82 years. Follow-up was terminated December 31, 1993. There were 90 deaths, and the average number of days hospitalized across all patients was 7.8 days per patient. Adjustment covariates included serum albumin (SALB), normalized protein catabolic rate (NPCR), subjective global assessment (SGA), percent lean body mass (PCTLBM), Kt/V, total weekly creatinine clearance (CCr), gender, race, age, Karnofsky score, underlying renal disease and cardiovascular disease (CVD).

As evident from Table 3.4, the covariates that are significant for the proportional hazards model include SALB, percent lean body mass and Age. A 1g/L increase in serum albumin concentration is associated with 11% decrease in the death hazard, adjusting for other covariates and conditional on the frailty. A 1% increase in lean body mass is associated with a 5% decrease in the hazard adjusting for other covariates. One year increase in age will result in 2% increase in hazard.

Factors that are significant predictors of the recurrent event rate include serum albumin, percentage lean body mass, Karnofsky score and Age. For example, a 1g/L increase in serum albumin concentration is associated with 6% decrease in hospitalization rate, adjusting for other variables. Age is the only significant variable in the Poisson regression model. Overall, serum albumin and percent lean body mass show up significant in two of three models where age is significant for all three models.

The heterogeneity effect are estimated through the variance of the frailty term. It appears to be highly significant(p-value<0.001). This shows much variation among the study subjects. The bottom half of Table 3.4 provides a comparison of parameter estimates based on a model which ignores the heterogeneity among the subjects. We

notice various degrees of bias in the parameter estimates, along with the changes in the significance levels for some of the clinical and demographic factors in the proportional hazards model, proportional rates model or the Poisson regression model.

3.5 Discussion

In this Chapter, we developed joint modelling methods to estimate the importance of certain parameters of interest. This is well suited to our hierarchical data setting where we have a sequence of recurrent events that could be stop by a terminating event. The marks that characterize each recurrent event are dependent upon the existence of each recurrent event. We specify separate models for each process (recurrent event process, terminating event process and mark process). The frailty term which is subject specific captures the dependence structure among the above three processes and serves as the bridge for our joint estimation. We assume a piecewise constant proportional hazards models for the terminating event while we assume a piecewise constant proportional rates model for the recurrent event rate given survival. We further assume a Poission regression type of model for the marks. One of our main contributions to the existing literature is the introduction of the marks. Here, we restrict the marks to be non-negative. In the conventional recurrent event literature, many papers focus on the recurrent event process itself (Cook and Lawless, 1997; Ghosh and Lin, 2000, 2002; Schaubel and Zhang, 2010). Their models essentially count the number of recurrent event occurrence, each event occurrence being binary (either 0 or 1). We built upon the notion that although marks are related to each recurrent event, they are not necessarily restricted to be binary. The marks can have their own processes depending upon the behavior of the marks. In other words, if we restrict the marks all to be identity, our models could be reduced to methods that only considering the recurrent event processes.

In this Chapter, we demonstrated the importance of capturing the heterogeneity

effect in modelling consideration. For example, a study population maybe not be homogeneous (some individuals will have higher hazards of death or recurrence of certain disease compared to others). In addition, it might be hard to measure all the factors related to the disease due to resources or other reasons. This makes models incorporating the frailty effect more reasonable. Our results show that when heterogeneity is present, building models without considering the effect of heterogeneity will cause different degree of biases depending on the scale of the heterogeneity effect.

Methods developed in this Chapter involve building parametric models for each of the underlying processes (i.e., recurrent event, terminating event, mark). In each of the processes, the frailty term was built in to represent the heterogeneity effect. They are assumed to be shared among all three processes. This assumption could be relaxed by introducing more than one frailty term. For example, we could quantify the dependence between the recurrent event process and the mark process by including one frailty term while using another frailty term for the mark process and terminating event process. The trade-off of introducing more than one frailty term is the added computation complexity since we need to estimated more parameters. Another point to mention is that the distribution of the the frailty term is assumed to be normal in the development of our models. Other distributions for the frailty term are also possible in building the models. (e.g., Gamma distribution).

We utilized Gaussian quadrature techniques in estimating the parameters of interest. The advantage of using Gaussian quadrature is the computation speed and relative accuracy. In order to use Gaussian quadrature, we would need to fully specify the likelihood function for the models described in this Chapter. Therefore, the models need to be fully parameterized. We typically do not know the true underlying models. Although, efficiency gains would result if the models being specified are accurate, substantial bias could arise if we have model misspecification.

Table 3.1: Simulation Results: Performance of the proposed parameter estimators under small heterogeneity

Parameters	Bias	ESD	ASE	CP	$BIAS^*$	ESD^*	ASE^*	CP^*
r_1	0.000	0.008	0.008	0.95	0.001	0.008	0.008	0.95
r_2	0.000	0.007	0.007	0.95	0.000	0.007	0.007	0.95
λ_1	0.000	0.004	0.004	0.96	0.000	0.004	0.004	0.96
λ_2	0.000	0.003	0.003	0.95	0.000	0.003	0.003	0.95
β_{G1}	-0.002	0.022	0.022	0.94	-0.003	0.022	0.021	0.93
β_{G2}	0.006	0.021	0.022	0.96	0.006	0.021	0.021	0.94
β_{G3}	-0.002	0.022	0.022	0.96	0.004	0.022	0.021	0.93
β_{D1}	0.001	0.057	0.058	0.95	0.000	0.057	0.058	0.95
β_{D2}	0.005	0.064	0.061	0.94	0.003	0.064	0.060	0.94
β_{D3}	-0.004	0.068	0.070	0.93	-0.005	0.068	0.069	0.93
β_{R1}	-0.002	0.029	0.031	0.95	-0.005	0.029	0.030	0.94
β_{R2}	0.003	0.031	0.032	0.96	0.001	0.030	0.031	0.96
β_{R3}	-0.002	0.035	0.036	0.97	-0.003	0.035	0.036	0.95
σ^2	0.000	0.004	0.004	0.92	-	-	-	-

Notes: $\beta_D = (0.5, 0.3, 0.3)'$, $\beta_R = (0.5, 0.3, 0.3)'$, $\beta_G = (0.5, 0.3, 0.3)'$,$\lambda_1 = 0.06$, $\lambda_2 = 0.03$, $r_1 = 0.2$, $r_2 = 0.2$, $\sigma^2 = 0.01$. Number of repetitions: 2000. Number of subjects: $n = 2000$. *Estimation assuming no correlation among terminating events, recurrent events and marks.

Table 3.2: Simulation Results: Performance of the proposed parameter estimators under moderate heterogeneity

Parameters	Bias	ESD	ASE	CP	$BIAS^*$	ESD^*	ASE^*	CP^*
r_1	0.000	0.008	0.008	0.96	0.008	0.008	0.008	0.82
r_2	0.000	0.007	0.008	0.94	-0.004	0.008	0.007	0.89
λ_1	0.000	0.004	0.004	0.96	0.001	0.004	0.004	0.95
λ_2	0.000	0.003	0.003	0.95	-0.001	0.003	0.003	0.93
β_{G1}	-0.004	0.024	0.026	0.97	0.003	0.024	0.020	0.90
β_{G2}	0.015	0.024	0.026	0.94	0.026	0.025	0.021	0.71
β_{G3}	0.000	0.025	0.026	0.95	0.021	0.026	0.021	0.77
β_{D1}	-0.006	0.060	0.059	0.94	-0.020	0.059	0.058	0.94
β_{D2}	0.015	0.064	0.062	0.95	0.006	0.062	0.061	0.94
β_{D3}	-0.009	0.068	0.071	0.95	-0.016	0.065	0.069	0.95
β_{R1}	-0.008	0.031	0.034	0.96	-0.024	0.031	0.030	0.86
β_{R2}	0.016	0.031	0.035	0.95	0.006	0.031	0.031	0.96
β_{R3}	-0.007	0.037	0.040	0.96	-0.016	0.038	0.036	0.90
σ^2	-0.001	0.006	0.006	0.94	-	-	-	-

Notes: $\beta_D = (0.5, 0.3, 0.3)'$, $\beta_R = (0.5, 0.3, 0.3)'$, $\beta_G = (0.5, 0.3, 0.3)'$,$\lambda_1 = 0.06$, $\lambda_2 = 0.03$, $r_1 = 0.2$, $r_2 = 0.2$, $\sigma^2 = 0.0625$. Number of repetitions: 2000. Number of subjects: $n = 2000$. *Estimation assuming no correlation among terminating events, recurrent events and marks.

Table 3.3: Simulation Results: Performance of the proposed parameter estimators under large heterogeneity

Parameters	Bias	ESD	ASE	CP	$BIAS^*$	ESD^*	ASE^*	CP^*
r_1	0.000	0.009	0.010	0.97	0.029	0.010	0.009	0.10
r_2	-0.001	0.009	0.009	0.96	-0.013	0.009	0.007	0.54
λ_1	0.010	0.004	0.004	0.97	0.004	0.004	0.004	0.88
λ_2	-0.001	0.003	0.003	0.96	-0.004	0.002	0.003	0.61
β_{G1}	-0.008	0.026	0.036	0.99	0.027	0.036	0.020	0.59
β_{G2}	0.032	0.029	0.036	0.90	0.065	0.041	0.020	0.26
β_{G3}	0.010	0.029	0.036	0.97	0.100	0.039	0.020	0.07
β_{D1}	-0.013	0.064	0.064	0.95	-0.060	0.058	0.058	0.84
β_{D2}	0.029	0.065	0.067	0.93	0.000	0.061	0.061	0.94
β_{D3}	-0.008	0.069	0.076	0.97	-0.032	0.064	0.069	0.94
β_{R1}	-0.014	0.036	0.041	0.98	-0.070	0.037	0.030	0.38
β_{R2}	0.032	0.035	0.043	0.93	-0.002	0.037	0.032	0.91
β_{R3}	-0.012	0.043	0.048	0.96	-0.039	0.045	0.036	0.75
σ^2	-0.010	0.012	0.013	0.85	-	-	-	-

Notes: $\beta_D = (0.5, 0.3, 0.3)'$, $\beta_R = (0.5, 0.3, 0.3)'$, $\beta_G = (0.5, 0.3, 0.3)'$,$\lambda_1 = 0.06$, $\lambda_2 = 0.03$, $r_1 = 0.2$, $r_2 = 0.2$, $\sigma^2 = 0.25$. Number of repetitions: 2000. Number of subjects: $n = 2000$. *Estimation assuming no correlation among terminating events, recurrent events and marks.

Table 3.4: Analysis of CANUSA study: Heterogeneity effect and parameter estimates under proportional hazards model, porpotional rates model and the Poisson regression model

	$d\Lambda_i(t)$			$dR_i(t)$			$G_i(t)$		
Variable	$\hat{\beta}_D$	SE	p	$\hat{\beta}_R$	SE	p	$\hat{\beta}_G$	SE	p
Frailty(Yes)									
SALB	-0.112	0.024	<0.001	-0.057	0.011	<0.001	-0.004	0.010	0.719
NPCR	-0.008	0.586	0.989	-0.007	0.276	0.981	0.049	0.241	0.840
SGA	-0.005	0.087	0.952	-0.113	0.041	0.006	-0.013	0.036	0.729
PCTLBM	-0.053	0.013	<0.001	-0.003	0.005	<0.001	-0.009	0.005	0.057
KTV	-0.030	0.219	0.891	-0.074	0.111	0.505	0.038	0.096	0.691
CCR	-0.003	0.006	0.669	0.002	0.003	0.509	0.001	0.002	0.926
CANADA	0.010	0.345	0.976	-0.009	0.164	0.955	-0.015	0.142	0.916
GENDER	-0.020	0.262	0.938	-0.044	0.126	0.725	0.091	0.110	0.406
RACE	-0.031	0.144	0.829	-0.028	0.070	0.686	0.076	0.062	0.220
AGE	0.021	0.009	0.027	-0.015	0.004	<0.001	0.033	0.003	<0.001
KARNOF	-0.002	0.009	0.866	-0.019	0.005	<0.001	-0.003	0.004	0.511
PRD	0.002	0.074	0.982	0.026	0.037	0.480	-0.048	0.032	0.130
CVD	0.041	0.261	0.876	0.006	0.128	0.965	0.089	0.011	0.425
$\hat{\sigma}^2$	1.325	0.083	<0.001	-	-	-	-	-	-
SALB	-0.054	0.021	0.012	-0.047	0.005	<0.001	0.006	0.002	0.000
NPCR	-0.010	0.544	0.986	-0.015	0.134	0.908	0.504	0.038	<0.001
SGA	0.008	0.077	0.922	-0.064	0.019	0.001	0.008	0.006	0.172
PCTLBM	-0.040	0.013	0.002	-0.027	0.003	<0.001	-0.006	0.001	<0.001
KTV	-0.047	0.198	0.814	-0.176	0.065	0.007	-0.122	0.019	<0.001
CCR	-0.001	0.005	0.898	0.004	0.002	0.003	0.001	0.000	0.001
CANADA	0.012	0.310	0.973	-0.010	0.077	0.895	-0.139	0.025	<0.001
GENDER	-0.026	0.232	0.911	-0.093	0.061	0.126	0.272	0.018	<0.001
RACE	-0.040	0.143	0.779	-0.089	0.033	0.007	0.081	0.009	<0.001
AGE	0.020	0.009	0.035	-0.023	0.002	<0.001	0.020	0.001	<0.001
KARNOF	-0.001	0.009	0.885	-0.018	0.002	<0.001	0.005	0.001	<0.001
PRD	0.010	0.066	0.886	0.008	0.017	0.642	-0.046	0.005	<0.001
CVD	0.044	0.230	0.849	0.037	0.059	0.531	0.266	0.018	<0.001
$\hat{\sigma}^2$	0	-	-	-	-	-	-	-	-

$d\Lambda_i(t)$: proportional hazards model;$dR_i(t)$: proportional rates model; $G_i(t)$: Poisson regression model.

Proposed method given in the top frame; parameter estimates in the bottom frame are based on a model with no frailty ($\sigma^2 \equiv 0$).

CHAPTER IV

Comparison of marginal means through inverse weighting methods

4.1 Introduction

There have been many methods developed for analysing recurrent event data in terms of various marginal means or rates models, largely due to the interpretation of the parameters and the fact that no particular dependency through event history need be assumed; e.g., Pepe and Cai, 1993; Lawless and Nadeau, 1995; Lin, Wei, Yang and Ying, 2000; Cai and Schaubel, 2004. Pepe and Cai (1993) developed partly marginal regression models that accommodate time dependent covariates. Later, Lawless and Nadeau (1995) presented robust semi-parametric techniques for estimating the cumulative mean function in discrete time. Lin et al. (2000) extended the approach of Lawless and Nadeau (1995) to continuous time and rigorously derived the asymptotic properties using empirical processes. Cai and Schaubel (2004) further proposed semi-parametric marginal means/rates model by allowing for more than one type of recurrent event. They also developed asymptotic properties for the parameter estimators. Investigators are often interested in estimating differences between groups (e.g., treatment versus control), and one approach is to make comparisons through the marginal means.

In this Chapter, we are interested in descriptive measures (i.e., "marks") that are associated with each recurrent event (e.g., repeated hospital admissions, epileptic seizures). Examples include entities such as the cost billed for each hospital visit, or the length of stay. In the recurrent event data setting, few methods have been proposed to analyze such type of data. One such method is that of Cai, Zeng and Pan (2010), which proposed a proportional means model for the marks, contingent upon the recurrent event occurrences, along with a marginal rate model for the recurrent event. Since in clinical trials or animal studies, study subjects may experience a terminating event (e.g., death), it is therefore necessary to acknowledge the fact that the recurrent event process could potentially be stopped permanently. The methods of Cai et al. (2010) would not apply to settings involving a terminating event.

A number of methods have been developed where both the recurrent event and terminating event are present. For example, Cook and Lawless (1997) developed models for the conditional recurrent event rate given survival. Ghosh and Lin (2000) proposed non-parametric inference procedures for the mean function, and later developed semi-parametric models for the marginal mean number of events (Ghosh and Lin, 2002). In such models, death is treated as a terminating event that prevents further recurrent event occurrence. Huang and Wang (2004) developed joint modelling techniques where a latent frailty variable is used to associate the recurrent event rate and terminating event hazard. The general data structure considered in this Chapter involves a sequence of recurrent events that could potentially be stopped by a terminating event. A mark is associated with each recurrent event, with marks only being observed when each recurrent event occurs.

The motivating example for our proposed methods is the Canada-USA (CANUSA) Peritoneal Dialysis (PD) Study (Canada-USA (CANUSA) Peritoneal Dialysis Study Group, 1996; Churchill 1998; Ye, Kalbfleisch and Schaubel, 2007). In this study, the sample consisted of PD patients starting dialysis in either Canada or USA. Such

patients experience repeated hospitalizations over the course of the study period. Whenever each hospitalization occurs, the number of days hospitalized are recorded. Demographic and some clinical information is collected at start of follow-up. Patients are removed from the study if either death or transplant occurs.

Of interest in Chapter IV is the integrated mark process, viewed as a function of time; as was the case in Chapter II. For example, in the motivating example (CANUSA Study), the recurrent event is hospital admission, the terminating event is death, and the mark is length of stay (measured in days). As a result, cumulative mark equals the total days hospitalized. The proposed methods take into account of treatment imbalance and censoring by incorporating Inverse Probability of Treatment Weighting (IPTW; Robins et al. 2000; Hernan et al. 2000; Anstrom and Tsiatis 2001) and Inverse Probability of Censoring Weighting (IPCW; Robins and Rotnitzky 1992; Robins 1993), respectively. We avoid building models for the recurrent event rate, terminating hazard and mark. The advantage of this less parametric approach is that we do not require models for each process to be correct (e.g., proportional hazards model, proportional rates model) in order to achieve consistent estimates of the marginal mean.

In Section 4.2, we introduce the proposed estimator and estimating methods. Asymptotic properties are given in Section 4.3, with proofs provided in the Appendix. Simulation studies are conducted in Section 4.4. An application of our proposed methods is given in Section 4.5, followed by discussion in Section 4.6.

4.2 Proposed Estimation Methods

4.2.1 Notation and Setup

We first establish the required notation. Let D_i denote the time of the terminating event for subject i, while C_i is the censoring time for subject i. We let

$X_i = \min(C_i, D_i)$ and let $Y_i(t) = I(X_i \geq t)$ be the at-risk indicator. We let the true terminating event counting process be represented by $N_i^{*D}(t) = I(D_i \leq t)$ and we let $N_i^{*R}(t)$ equal the total number of recurrent events for subject i up to time t. The observed terminating event counting process and recurrent event process are $N_i^D(t) = I(D_i \leq t, D_i < C_i)$ and $N_i^R(t) = N_i^{*R}(t \wedge X_i)$, respectively. Notice that recurrent events do not occur after death. We define $G_i(t)$ as the mark for subject i at time t. It is assumed that marks only occur at time of a recurrent event. In addition, subjects are divided into groups (e.g., by treatment type, gender, diagnosis; whatever is of chief interest to the investigator) and A_i is used to denote group for subject i. For ease of presentation, we consider the case where A_i is binary ($A_i = 0, 1$), although the proposed methods can easily accommodate more than two groups. We then set up the group indicator $A_{ij} = I(A_i = j)$. Each subject is characterized by a vector of possibly time dependent covariates, $Z_i(t)$. Typically, one would want to make the $j = 0$ and $j = 1$ groups comparable at time 0 (balance the baseline covariate distribution); would generally not want to adjust for events after $t = 0$. We set the baseline value of the covariate vector to $Z_i \equiv Z_i(0)$. Finally, we define $\mathcal{F}_i(t) = \{N_i(s), G_i(s), Z_i(s); s \in [0, t)\}$ as the recurrent event, mark and covariate history up to (but not including) time t.

Of interest for each subject is the cumulative mark

$$B_i^*(t) = \int_0^t G_i(u) dN_i^{*R}(u),$$

with mean $\mu_{ij}(t)$, defined as

$$\mu_{ij}(t) = E[B_i^*(t) | Z_i, A_i = j].$$

That is, $\mu_{ij}(t)$ represents the mean of $B_i^*(t)$ for a subject with covariate Z_i, under

the hypothetical scenario where subject i is a member of group j. To contrast group-specific means, we first define $\mu_j(t)$ as the cumulative mean for group j, averaging across the marginal distribution of Z_i. That is,

$$\mu_j(t) \equiv E[\mu_{ij}(t)] = E[E[B_i^*(t)|A_i = j, Z_i]], \tag{4.1}$$

for $j = 0, 1$, where the inner expectation conditions on Z_i and $A_i = j$, irrespective of the group to which subject i actually belongs (i.e., under the hypothetical scenario where, possibly contrary to fact that, $A_i = j$), and the outer expectation is taken with respect to the marginal distribution of Z_i. This way, even if the Z_i distributions are quite different across groups, $\mu_0(t)$ and $\mu_1(t)$ are comparable, in the sense that each is averaged with respect to the same covariate distribution. The quantity of chief interest is the difference in cumulative means between the two groups, which is expressed as

$$\delta(t) = \mu_1(t) - \mu_0(t). \tag{4.2}$$

The goal here is to estimate $\delta(t)$, methods for which are proposed in the next subsection.

4.2.2 Estimation

We develop semi-parametric procedures to estimate treatment-specific means. The proposed methods assume no functional form with respect to the relationship between the treatment-specific mean functions. If the treatment were randomized, a natural estimator of $\mu_j(t)$ could be given by

$$\hat{\mu}_j^*(t) = \int_0^t \frac{\sum_{i=1}^n I(A_i = j) G_{ij}(u) dN_{ij}^*(u)}{\sum_{i=1}^n I(A_i = j)}. \tag{4.3}$$

In observational studies, treatment is rarely randomized and, in the absence of randomization, the above estimator would be generally biased due to confounding. In addition to the aforementioned obstacle, censoring times are unknown for subjects observed to die. In order to tackle the above two complexities, we propose an inverse-weighted estimator. IPTW is utilized to account for the treatment imbalance, while IPCW is utilized to account for censoring.

4.2.3 Proposed Estimator: Inverse Weighting

The idea of using IPTW is to essentially create treatment-specific pseudo-populations with the same adjustment covariate distribution. An IPTW based nonparametric estimator of the mean of treatment group j has the interpretation of mean number of recurrent events which would result if the whole population had received treatment j. In order to derive the IPTW weight, we assume a logistic model for A_i given Z_i,

$$p_{ij}(\theta_0) \equiv P(A_i = j|Z_i;\theta_0) = \frac{\exp\{\theta_0' Z_i\}}{1+\exp\{\theta_0' Z_i\}}, \tag{4.4}$$

where $i = 1 \ldots n$ and $j = 0, 1$. Under maximum likelihood, we can get an estimate of θ_0 by solving the corresponding score equation. Then the IPTW weight function is given by

$$w_{ij}^A(\hat{\theta}) = \frac{A_{ij}}{p_{ij}(\hat{\theta})}. \tag{4.5}$$

Correspondingly, the purpose of using a IPCW weight is to replace the unobserved quantity, in this case $G_i(t)dN_i^{*R}(t)$, by an observed quantity which has the same expectation. We let the cause-specific hazard function for C_i be denoted by

$$\lambda_{ij}^C(t) = \lim_{dt\downarrow 0} \frac{1}{dt}\mathcal{P}\{t \le C_i < t+dt, C_i < D_i | X_i \ge t, \mathcal{F}_i(t), A_i = j\} \tag{4.6}$$

We set up the time-dependent covariate $Z_i^C(t)$ to satisfy $\lambda_{ij}^C\{t|Z_i^C(t), A_i = j\} = \lambda_{ij}^C\{t|\mathcal{F}_i(t), A_i = j\}$. That is, $Z_i^C(t)$ consists of time-dependent covariates, $Z_i^C(t)$, plus components of the event and/or mark history that predict censoring. We assume the following proportional hazards model for C_i,

$$\lambda_{ij}^C(t) = \lambda_{0j}^C(t) \exp\{\beta_C' Z_i^C(t)\}, \tag{4.7}$$

and let $\Lambda_{ij}^C(t) = \int_0^t \lambda_{ij}^C(s)ds$. Then the IPCW weight $w_{ij}^C(t)$ can be estimated by the following,

$$\hat{w}_{ij}^C(t) = \exp\{\hat{\Lambda}_{ij}^C(t)\}, \tag{4.8}$$

where $\hat{\Lambda}_{0j}^C(t) = n^{-1} \sum_{i=1}^n \int_0^t \hat{\pi}_j(s)^{-1} A_{ij} dN_i^C(s)$ and $\hat{\pi}_j(s) = n^{-1} \sum_{i=1}^n A_{ij} Y_i(s) \exp\{\hat{\beta}' Z_i^C(s)\}$. After the above definition, it is easy to show that the following estimator,

$$\hat{\mu}_j(t) = n^{-1} \sum_{i=1}^n \int_0^t w_{ij}^A(\hat{\theta}) \hat{w}_{ij}^C(s) G_i(s) dN_i^R(s), \tag{4.9}$$

which is IPTW and IPCW weighted, is a consistent estimator of $\mu_j(t)$. Therefore, the differences in cumulative means between the two groups can then be consistently estimated by,

$$\hat{\delta}(t) = \hat{\mu}_1(t) - \hat{\mu}_0(t). \tag{4.10}$$

4.3 Asymptotic Properties

In this section, we summarize the essential asymptotic behavior of the proposed estimators by first listing the necessary conditions, for $i = 1, \ldots, n$ and j= 0, 1.

(a) $\{N_i^R(.), X_i, Z_i(.), A_i, G_i(.)\}$ are independent and identically distributed

(b) $Pr(Y_i(t) = 1) > 0$ for all $t \in [0, \tau]$, where τ is a pre-specified constant

(c) $N_i^R(\tau) < \infty$ and $\Lambda_0^C(\tau) < \infty$

(d) $0 < Pr(A_i = j|Z_i) < 1$ for $j = 0, 1$.

(e) Covariates $Z_{il}(t)$ are bounded almost surely, where $t \in [0, \tau]$ and $Z_{il}(t)$ is the lth element of $Z_i(t)$

(f) Positive-definiteness of the matrices, $A_j^C(\beta_C)$ and $I(\theta)$, where

$$\begin{aligned} A_j^C(\beta_C) &= E\left[\int_0^\tau \{Z_i^C(t) - \bar{z}(t;\beta_C)\}^{\otimes 2} Y_i(t)\lambda_{ij}^C(t)dt\right] \\ \bar{z}_j(t;\beta_C) &= s_j^{(1)}(t;\beta_C)s_j^{(0)}(t;\beta_C)^{-1} \\ s_j^{(d)}(t;\beta_C) &= E[A_{ij}Y_i(t)Z_i(t)^{\otimes d}\exp\{\beta_C' Z_i\}], d = 0, 1, 2. \\ I(\theta) &= E[Z_i^{\otimes 2}(1 - p_{ij}(\theta))p_{ij}(\theta)] \end{aligned}$$

Now, we summarize the asymptotic properties from the proposed procedures. Proofs of the theorems are provided in the Appendix.

Theorem IV.1. *Under the above regularity conditions, $\hat{\mu}_j$ is a uniformly consistent estimator of μ_j. That is, $\hat{\mu}_j(t)$ converges to $\mu_j(t)$ almost surely for j=0,1 and t $\in (0, \tau]$. In addition, $n^{1/2}\{\hat{\mu}_j(t) - \mu_j(t)\}$ converges weakly to a zero-mean Gaussian process for j=0,1 and $t \in (0, \tau]$ with covariance function $\sigma_j(s,t) = E[\psi_{ij}(s)\psi_{ij}(t)]$,*

where

$$\psi_{ij}(t) = \sum_{i=1}^{3} \psi_{ijk}(t)$$

$$\begin{aligned}
\psi_{ij1}(t) &= H_j^{'}(t) I(\theta_0)^{-1} n^{-1/2} \sum_{i=1}^{n} U_i(\theta_0) \\
\psi_{ij2}(t) &= E[w_{ij}^A(\theta_0) \int_0^t w_{ij}^C(s) \exp\{\beta_C^{'} Z_i^C(s)\} \int_0^s \{Z_i(u) - \bar{z}(u;\beta_C)\} d\Lambda_{0j}(u) G_i(s) dN_i^R(s)] \\
&\quad \times A_j^C(\beta_C)^{-1} n^{-1/2} \sum_{i=1}^{n} U_{ij}^C(\beta_C) \\
&\quad + \int_0^t E[w_{ij}^A(\theta_0) \exp\{\beta_C Z_i^C(s)\} w_{ij}^C(s) G_i(s) dN_i(s)] n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^s \frac{dM_{ij}^C(u;\beta_C)}{s_j^{(0)}(u;\beta_C)} \\
\psi_{ij3}(t) &= n^{-1/2} \sum_{i=1}^{n} \int_0^t \{w_{ij}^A(\theta_0) w_{ij}^C(s) G_i(s) dN_i(s) - d\mu_j(s)\}
\end{aligned}$$

where we define

$$\begin{aligned}
U_{ij}^C(\beta_C) &= \int_0^\tau A_{ij} \{Z_i^C(t) - \bar{z}(t;\beta_C)\} dM_{ij}^C(t;\beta_C) \\
U_i(\theta) &= Z_i \{A_{i1} - p_{i1}(\theta)\} \\
H_j(t) &= E\left[\int_0^t (-1)^j A_{ij} \frac{1 - p_{ij}(\theta)}{p_{ij}(\theta)} Z_i w_{ij}^C(s) G_i(s) dN_i(s) \right] \\
dM_{ij}^C(s) &= A_{ij} \{dN_i^C(s) - Y_i(s) d\Lambda_{ij}^C(s)\}.
\end{aligned}$$

The proof of the above theorem involves the Weak Law of Large Numbers (WLLN), series of Tayor expansions, Central Limit Theorem (CLT) and various results from empirical processes (Pollard, 1990; van der Vaart and Wellner 1996).

Theorem IV.2. *Under the above regularity conditions, $\hat{\delta}(t)$ is a uniformly consistent estimator of $\delta(t)$. That is, $\hat{\delta}(t)$ converges to $\delta(t)$ almost surely for j=0,1 and $t \in (0, \tau]$.*

Furthermore, $n^{1/2}\{\hat{\delta}(t)-\delta(t)\}$ *converges weakly to a zero-mean Gaussian process with covariance function* $\Delta(s,t) = E[\{\psi_{i1}(s) - \psi_{i0}(s)\}\{\psi_{i1}(t) - \psi_{i0}(t)\}]$.

It follows from Theorem 1 that $n^{1/2}\{\hat{\mu}_1(t) - \hat{\mu}_0(t)\}$ is asymptotically equivalent to $n^{-1/2}\sum_{i=1}^{n}\{\psi_{i1}(t) - \psi_{i0}(t)\}$, a scaled sum of zero-mean Normal variates. For fixed (s,t), convergence follows from the MCLT. Tightness can be demonstrated using results from empirical processes theory, which completes the process aspect of the proof.

4.4 Simulation Study

We evaluated the finite-sample properties of the proposed estimator through simulation. The terminating event was generated from $\lambda_{ij}(t) = \lambda_{0j}\exp\{\beta_D' Z_i\}$, a proportional hazards model where $Z_i = (Z_{i1}, Z_{i2}, Z_{i3})'$. The Cox regression parameter was set to $\beta_D = (0.5, 0.3, 0.3)'$, with λ_{0j}=0.03 or 0.06. We let $Z_{i1} \sim Bernoulli(0.5)$, with Z_{i2} generated through the model $P(Z_{i2} = 1|Z_{i1})$=expit$\{h_1 Z_{i1}\}$, $P(Z_{i3} = 1|Z_{i1}, Z_{i2})$= expit$\{h_2 Z_{i1} + h_3 Z_{i2}\}$ and $P(A_i = 1|Z_{i1}, Z_{i2}, Z_{i3})$=expit$\{h_2 Z_{i1} + h_3 Z_{i2} + h_4 Z_{i3}\}$ where $(h_1, h_2, h_3, h_4)' = (1, 1, 1, 1)'$ and with expit(h)=exp$\{h\}[1 + \exp\{h\}]^{-1}$. We generated recurrent events from the model $r_{ij}(t) = Q_i r_{0j}\exp\{\beta_R' Z_i\}$ by generating gap times between successive events as: $T_{i,k+1} = T_{i,k} - Q_i \log(U_{i,k})[r_{0j}\exp\{\beta_R' Z_i\}]^{-1}$ for $k = 1, \ldots, 50$, where each of the $U_{i,k}$ variates followed a Unif(0,1) distribution. The frailty, Q_i, could be considered an unmeasured predictor that is shared by all recurrent event times for the same subject. We let $Q_i \sim Gamma(\theta)$ where θ= 0.5 or 0.25 and represents Var(Q_i); note that $E(Q_i) = 1$. We set r_{0j}= 0.25 or 0.20, with the regression parameter set to $\beta_R = (0.5, 0.3, 0.3)'$. The model for the marks was as follows: $g_i(t) = \beta_{G_0} + \beta_{G_1} t + \beta_{G_2} Z_{i1} + \beta_{G_3} Z_{i2} + \beta_{G_4} Z_{i3}$, with $(\beta_{G_0}, \beta_{G_1}, \beta_{G_2}, \beta_{G_3}, \beta_{G_4})'$=$(1, 1, 1, 1, 1.5)'$. Censoring times were generated through $\lambda_{ij}^C(t) = \lambda_{0j}^C\exp\{\beta_C' Z_i + \log(1.05)N_i(t^-)\}$ and truncated at 60, which resulted in

about 30% to 40% of censoring for different simulation configurations. In the above model, $\lambda_{0j}^C = 0.03$, $\beta_C = (0.2, 0.2, 0.2)'$ and $N_i(t^-) = N_i(t) - dN_i(t)$, the number of recurrent event up to (but not including) time t. The average observed number of recurrent events ranged from 2 to 4 per subject. Cumulative means were estimated at $t = 3, 6, 9, 12$, in order to give comparisons for early, middle and late follow up times. Sample sizes were set at $n = 400$ and 500 replicates were generated per configuration.

Table 4.1 provides results for the population average cumulative mean estimator for the various data configurations examined. For all data configurations, the estimated cumulative means are very close to the true values. The average asymptotic standard errors (ASE) agree well with the empirical standard deviations (ESD) and, correspondingly, the empirical coverage probabilities (CP) are close to the nominal value of 0.95.

In Table 4.2, we evaluated the proposed treatment effect estimator, $\hat{\delta}(t) = \hat{\mu}_1(t) - \hat{\mu}_0(t)$. The biases at different follow-up times are negligible compared to the true values. The ASEs also agree well with the ESDs, with the CP being generally quite close to the nominal value of 0.95.

To compare relative efficiencies to estimator proposed in Chapter II, we generated data under the same settings as those described in Chapter II. Table 4.3 provides results for $\hat{\mu}_1(t)$ for the various data configurations examined. For all data configurations, the estimated cumulative means are very close to the true values. Compared with the estimators presented in Table 2.1, we can see that we achieve substantial efficiency gains by utilizing the inverse weighted estimator. In Table 4.4, we evaluated the proposed treatment effect estimator, $\hat{\delta}(t) = \hat{\mu}_1(t) - \hat{\mu}_0(t)$. The biases at different follow-up times are negligible compared to the true values. Note that for the simulation presented in Table 4.3 and Table 4.4, the recurrent event process does not predict $\lambda_i^C(t)$. In addition, the efficiencies are comparable in terms of the treatment effect.

4.5 Application

Data were obtained from the Canada-USA (CANUSA) peritoneal dialysis (PD) study. More than three years of data were collected, including demographic, nutritional status, adequacy of dialysis. The final study population consists of $n = 679$ peritoneal dialysis patients from either US (577 patients) or Canada (102 patients). During the follow-up period, these PD patients may experience repeated hospitalization, death, transplant or loss of follow-up.

In our analysis, repeated hospitalizations were treated as recurrent event of interest while death were considered as the terminating event. The number of days hospitalized associated with each hospital admission was considered the mark. Chief interest in our analysis is to compare the marginal mean number of total days hospitalized between the PD patients from US and Canada. The CANUSA data being handled in this application has a hierarchical type of structure: patients are hospitalized conditioning on being alive up to their hospitalization visit; number of days hospitalized being observed are conditioned on each occurrence of hospital visit. In addition, we treat transplant or loss of follow-up as censoring event. The adjustment covariate included serum albumin (SALB), subjective global assessment (SGA), percent lean body mass (PCTLBM), Kt/V, CCR, gender, race, age, cardiovascular disease (CVD) and the event history.

The results given were based on the inverse weighting method where inverse probability of censoring weighting and inverse probability of treatment weighting were both considered. We first fitted a logistic regression model and a proportional hazards model to detect any significant covariates of interests. The logistic regression model, $p_{ij}(\theta_0)$, models the probability of being Canadian, given the covariates. The proportional hazards model, $\lambda_{ij}^{C}(t)$, was stratified by country. As shown in Table 4.5, SGA, PCTLBM, KTV, CCR, gender, race and age are all significant or borderline significant for the logistic regression model; while SALB, PCTLBM, age, CVD and

event history are significant predictors for the proportional hazards model. For example, a 1 % increase in lean body mass is associated with 4% decrease in the odds of being Canadian, adjusting for other covariates. Those having a history of CVD will have 2.11 times higher hazards of censoring compared with those without a history of CVD, adjusting for other covariates. Similarly, each additional hospital visit is associated with an 8% times higher hazard of censoring, adjusting for other covariates. Generally, higher percent of lean body mass, fewer hospitalizations, absence of cardio-vascular disease history and increase in SALB are each associated with lower risk of censoring.

Figure 4.1 presents the estimated cumulative mean number of days hospitalized for American and Canadian patients over a two year period. From the figure, we can notice that overall the estimated mean number of days hospitalized are similar between the two groups, with the solid line being close to the dashed line. Canadian PD patients have a slightly higher number of days hospitalized compared with the American patients throughout the two year period.

Figure 4.2 gave the mean difference estimator between the American and Canadian PD patients with 95 % pointwise confidence intervals plotted. As is evident from Figure 4.2, the mean difference estimator is very close to the reference ($\hat{\delta}(t) = 0$) over time. Throughout the two year period, the difference was slightly above zero. Correspondingly, we computed the estimated mean difference and associated asymptotic standard error at every half year interval as shown in Table 4.6. We notice that the estimated mean difference are close to zero and the differences are all non-significant. This means that American PD patients compared with their Canadian counterparts have an approximately equal mean number of days hospitalized over the entire follow-up period.

4.6 Discussion

The method developed in this Chapter combines an Inverse Probability of Treatment Weighting (IPTW) and Inverse Probability of Censoring Weighting (IPCW). In the development of IPTW, treatment assignment is dependent upon the baseline covariate distribution, such that treatment imbalance exists between any two comparison groups. For ease of interpretation, we demonstrated the performance under two comparison groups, but the proposed methods can easily be extended to settings where there exist more than two groups. In the development of IPCW, we allowed the censoring mechanism to depend on the covariate distribution, and the event history. Although we have strong dependence structure, our methods still perform very well under various simulation settings.

In this Chapter, we are mainly interested in the marginal cumulative differences between any two comparison groups. The estimator developed in this Chapter does not assume any underlying models for the recurrent event process, terminating event process or the mark process. On the contrary, the estimator proposed in Chapter II would need to allow the individual models to be specified correctly. Otherwise, we could see biased parameter estimates that could undermine the true outcome. Even though comparing with methods developed in Chapter II, we can see that under limited sample size, two methods achieve quite similar results. The estimator developed in this Chapter would achieve more robust results under model misspecification.

Schaubel and Zhang (2010) also proposed methods to estimate treatment effect on the marginal recurrent event mean via inverse weighting. Though there are similarities to our methods, Schaubel and Zhang (2010) did not consider the existence of marks. They used IPTW, and proposed two methods of dealing with C_i (e.g., imputing C_i when $D_i < C_i$; IPCW). In their IPCW model, they assume that neither covariates nor recurrent event history predicted censoring, which is quite stringent. We made the IPCW component richer. We did not consider an imputation approach, as it

would seem to be difficult to carry out in the sense that more complex modelling would need to be done than in Schaubel and Zhang (2010).

Table 4.1: Simulation Results: Performance of Proposed Inverse Weighted Estimator

λ_{00}, λ_{01}	r_{00}, r_{01}	$V(Q_i)$	t	% at risk	$\mu_1(t)$	Bias	ESD	ASE	CP
0.03, 0.03	0.20, 0.25	0.5	3	77%	6.98	-0.08	1.01	0.67	0.92
			6	61%	16.01	-0.29	2.32	1.53	0.93
			9	47%	26.12	-0.47	3.90	2.72	0.93
			12	36%	36.66	-0.71	5.78	4.22	0.94
0.03, 0.03	0.20, 0.25	0.25	3	77%	6.98	-0.10	0.74	0.62	0.93
			6	61%	16.01	-0.26	1.72	1.39	0.92
			9	47%	26.12	-0.31	3.02	2.43	0.93
			12	36%	36.66	-0.36	4.56	3.77	0.92
0.03, 0.06	0.25, 0.25	0.5	3	70%	6.23	0.02	0.64	0.63	0.93
			6	48%	12.84	0.06	1.41	1.33	0.92
			9	38%	18.98	0.13	2.33	2.15	0.92
			12	29%	24.38	0.25	3.44	3.06	0.91
0.03, 0.06	0.25, 0.25	0.25	3	70%	6.23	0.04	0.54	0.58	0.97
			6	48%	12.84	0.05	1.18	1.19	0.95
			9	38%	18.98	0.05	1.85	1.85	0.96
			12	29%	24.38	0.00	2.61	2.56	0.93
0.06, 0.06	0.25, 0.20	0.5	3	70%	4.99	0.00	0.46	0.52	0.96
			6	48%	10.27	0.03	1.08	1.07	0.94
			9	38%	15.19	0.03	1.75	1.68	0.92
			12	29%	19.50	-0.02	2.44	2.32	0.92
0.06, 0.06	0.25, 0.20	0.25	3	70%	4.99	-0.02	0.46	0.49	0.96
			6	48%	10.27	-0.04	0.95	0.97	0.95
			9	38%	15.19	-0.07	1.48	1.50	0.95
			12	29%	19.50	-0.08	2.07	2.08	0.95

Notes: $\beta_D = (0.5, 0.3, 0.3)'$, $\beta_R = (0.5, 0.3, 0.3)'$, $\beta_G = (1, 1, 1, 1, 1.5)'$. Number of repetitions: 500. Number of subjects: $n = 400$.

Table 4.2: Simulation Results: Treatment effects of Inverse Weighted Estimator

λ_{00}, λ_{01}	r_{00}, r_{01}	$V(Q_i)$	t	% at risk	$\delta(t)$	Bias	ESD	ASE	CP
0.03, 0.03	0.20, 0.25	0.5	3	77%	1.40	-0.15	1.39	1.12	0.94
			6	61%	3.20	-0.35	3.18	2.54	0.93
			9	47%	5.22	-0.68	5.29	4.43	0.94
			12	36%	7.33	-1.09	7.99	6.77	0.94
0.03, 0.03	0.20, 0.25	0.25	3	77%	1.40	-0.19	1.11	1.05	0.94
			6	61%	3.20	-0.48	2.54	2.34	0.95
			9	47%	5.22	-0.57	4.39	4.00	0.95
			12	36%	7.33	-0.73	6.50	5.99	0.95
0.03, 0.06	0.25, 0.25	0.5	3	70%	0.75	-0.03	1.40	1.22	0.94
			6	48%	3.17	-0.15	3.11	2.71	0.92
			9	38%	7.14	-0.48	5.13	4.52	0.95
			12	29%	12.29	-0.74	7.74	6.77	0.94
0.03, 0.06	0.25, 0.25	0.25	3	70%	0.75	-0.02	1.18	1.13	0.95
			6	48%	3.17	0.04	2.54	2.49	0.93
			9	38%	7.14	0.06	4.20	4.10	0.93
			12	29%	12.29	0.14	6.36	6.00	0.92
0.06, 0.06	0.25, 0.20	0.5	3	70%	1.25	0.00	1.02	1.08	0.96
			6	48%	2.57	-0.06	2.26	2.24	0.96
			9	38%	3.80	-0.11	3.58	3.52	0.94
			12	29%	4.88	-0.23	4.96	4.82	0.95
0.06, 0.06	0.25, 0.20	0.25	3	70%	1.25	0.15	0.98	1.02	0.96
			6	48%	2.57	0.36	2.01	2.07	0.95
			9	38%	3.80	0.51	3.09	3.20	0.95
			12	29%	4.88	0.68	4.24	4.36	0.95

.

Notes: $\beta_D = (0.5, 0.3, 0.3)'$, $\beta_R = (0.5, 0.3, 0.3)'$, $\beta_G = (1, 1, 1, 1, 1.5)'$.Number of repetitions: 500. Number of Subjects: $n = 400$.

Table 4.3: Simulation Results: Performance of Proposed Inverse Weighted Estimator

λ_{00}, λ_{01}	r_{00}, r_{01}	$V(Q_i)$	t	% at risk	$\mu_1(t)$	$Bias^*$	ESD^*	$Bias^\dagger$	$ESD^\dagger$	RE
0.03/0.03	0.20/0.25	0.5	3	71%	6.98	-0.06	0.57	0.04	0.77	1.82
			6	53%	16.01	-0.23	1.35	0.10	1.75	1.64
			9	40%	26.12	-0.50	2.25	0.27	2.92	1.61
			12	29%	36.66	-0.82	3.34	0.35	4.31	1.59
0.03/0.03	0.20/0.25	0.25	3	71%	6.98	-0.03	0.48	0.04	0.56	1.37
			6	53%	16.01	-0.21	1.12	0.18	1.34	1.41
			9	40%	26.12	-0.36	1.96	0.02	2.26	1.28
			12	29%	36.66	-0.70	2.96	0.00	3.34	1.20
0.03/0.06	0.25/0.25	0.5	3	59%	6.23	0.01	0.59	0.02	0.66	1.25
			6	40%	12.84	-0.15	1.19	0.02	1.50	1.56
			9	23%	18.98	-0.25	1.89	0.11	2.42	1.61
			12	16%	24.38	-0.33	2.61	0.22	3.31	1.59
0.03/0.06	0.25/0.25	0.25	3	59%	6.23	-0.08	0.46	0.04	0.57	1.49
			6	40%	12.84	-0.16	0.99	0.04	1.22	1.49
			9	23%	18.98	-0.27	1.61	0.11	1.92	1.39
			12	16%	24.38	-0.40	2.29	0.04	2.64	1.30
0.06/0.06	0.25/0.20	0.5	3	59%	4.99	0.00	0.45	0.03	0.57	1.61
			6	40%	10.27	-0.04	0.94	0.13	1.18	1.59
			9	23%	15.19	-0.07	1.48	0.19	1.82	1.54
			12	16%	19.50	-0.13	2.03	0.20	2.47	1.49
0.06/0.06	0.25/0.20	0.25	3	59%	4.99	-0.04	0.39	0.00	0.45	1.32
			6	40%	10.27	-0.11	0.86	0.05	0.98	1.28
			9	23%	15.19	-0.13	1.35	0.06	1.53	1.28
			12	16%	19.50	-0.19	1.88	0.10	2.09	1.23

Notes: $\beta_D = (0.5, 0.3, 0.3)'$, $\beta_R = (0.5, 0.3, 0.3)'$, $\beta_G = (1, 1, 1, 1, 1.5)'$. Number of repetitions: 500. Number of subjects: $n = 400$. *: Bias and ESD of estimator in Chapter IV. †: Bias and ESD of estimator in Chapter II. RE: relative efficiency, compared to the Chapter II estimator.

Table 4.4: Simulation Results: Treatment effects of Inverse Weighted Estimator

λ_{00}, λ_{01}	r_{00}, r_{01}	$V(Q_i)$	t	% at risk	$\delta(t)$	$Bias^*$	ESD^*	$Bias^\dagger$	$ESD^\dagger$	RE
0.03/0.03	0.20/0.25	0.5	3	71%	1.40	0.03	0.83	0.01	0.86	1.08
			6	53%	3.20	-0.12	1.84	0.07	1.88	1.04
			9	40%	5.22	-0.18	3.31	0.14	3.16	0.91
			12	29%	7.33	-0.18	4.89	0.07	4.79	0.96
0.03/0.03	0.20/0.25	0.25	3	71%	1.40	-0.02	0.76	0.04	0.74	0.95
			6	53%	3.20	-0.01	1.61	0.04	1.62	1.01
			9	40%	5.22	-0.09	2.78	-0.05	2.73	0.96
			12	29%	7.33	0.01	4.18	-0.25	4.04	0.94
0.03/0.06	0.25/0.25	0.5	3	59%	0.75	0.00	0.95	-0.03	0.90	0.90
			6	40%	3.17	0.05	2.07	-0.07	1.94	0.88
			9	23%	7.14	0.17	3.29	-0.19	3.30	1.01
			12	16%	12.29	0.13	4.67	-0.59	4.78	1.06
0.03/0.06	0.25/0.25	0.25	3	59%	0.75	0.04	0.83	-0.08	0.82	0.99
			6	40%	3.17	0.01	1.74	-0.08	1.80	1.08
			9	23%	7.14	-0.08	2.95	-0.16	2.97	1.02
			12	16%	12.29	-0.13	4.33	-0.05	4.35	1.01
0.06/0.06	0.25/0.20	0.5	3	59%	1.25	-0.03	0.86	-0.03	0.87	1.03
			6	40%	2.57	-0.09	1.77	-0.13	1.74	0.97
			9	23%	3.80	-0.19	2.76	-0.06	2.76	0.99
			12	16%	4.88	-0.27	3.84	0.11	3.83	0.99
0.06/0.06	0.25/0.20	0.25	3	59%	1.25	0.04	0.76	0.09	0.74	0.96
			6	40%	2.57	0.03	1.62	0.16	1.59	0.97
			9	23%	3.80	0.06	2.56	0.25	2.53	0.99
			12	16%	4.88	0.08	3.62	0.29	3.52	0.95

Notes: $\beta_D = (0.5, 0.3, 0.3)'$, $\beta_R = (0.5, 0.3, 0.3)'$, $\beta_G = (1, 1, 1, 1, 1.5)'$. Number of repetitions: 500. Number of subjects: $n = 400$. *: Bias and ESD of estimator in Chapter IV. †: Bias and ESD of estimator in Chapter II. RE: relative efficiency, compared to the Chapter II estimator.

Table 4.5: Analysis of CANUSA data: Parameter estimates for logistic regression model and proportional hazards model for censoring

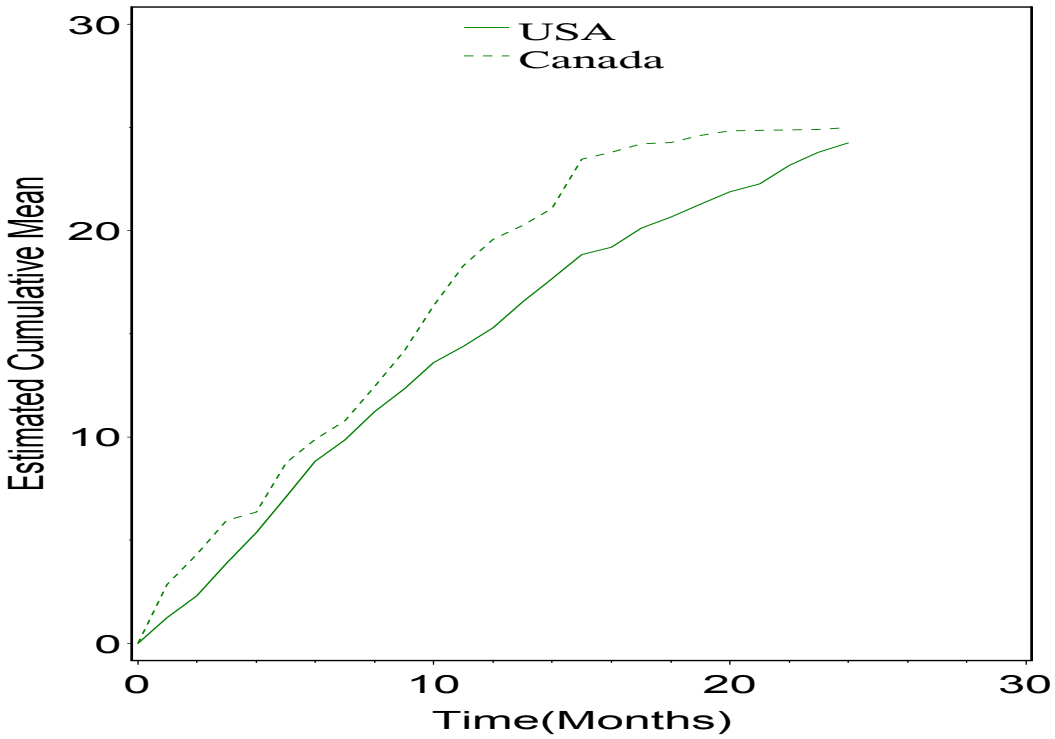
	$p_{ij}(\theta_0)$			$\lambda_{ij}^C(t)$		
Variable	$\hat{\theta}$	p	Odds Ratio	$\hat{\beta}_C$	p	$\exp\{\hat{\beta}_C\}$
SGA	0.34	0.00	1.41	-	-	-
KTV	-1.37	0.00	0.25	-	-	-
CCR	0.02	0.01	1.02	-	-	-
GENDER	0.47	0.10	1.59	-	-	-
RACE	0.24	0.06	1.28	-	-	-
AGE	-0.01	0.14	0.99	0.02	0.07	1.02
PCTLBM	-0.04	0.00	0.96	-0.03	0.00	0.97
SALB	-	-	-	-0.05	0.02	0.95
CVD	-	-	-	0.75	0.00	2.11
EVENT HISTORY	-	-	-	0.07	0.04	1.08

.

Notes: $p_{ij}(\theta_0) = P\{A_i = 1|Z_i\}$, with $j = 0$ for US patients and $j = 1$ for Canadian patients. $\lambda_{ij}^C(t)$: proportional hazards model for censoring p: p-value

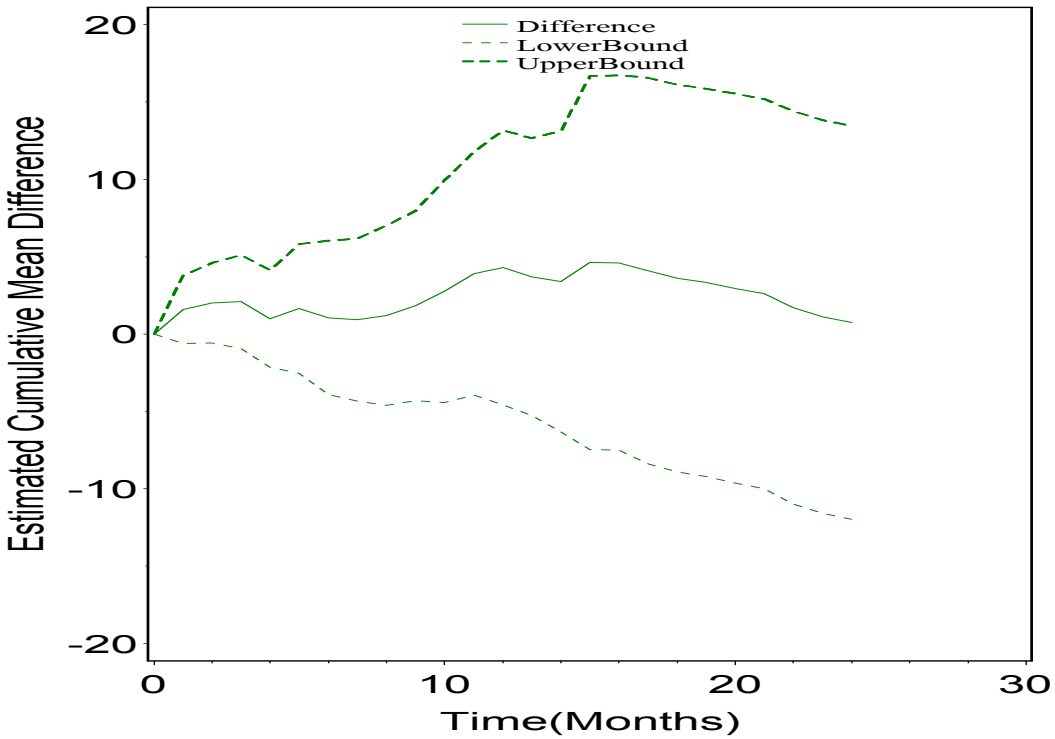
Table 4.6: Analysis of CANUSA data: Difference in length of hospitalization between American and Canadian patients over time (Canadian minus American)

Months	$\hat{\delta}(t)$	$\hat{SE}\{\hat{\delta}(t)\}$	P-value
6	1.06	2.54	0.68
12	4.29	4.53	0.35
18	3.61	6.38	0.58
24	0.75	6.49	0.91



The solid line represents American ($j = 0$) PD patients, and dashed line represents Canadian ($j = 1$) PD patients.

Figure 4.1: Analysis of the CANUSA data: Estimated cumulative mean number of days hospitalized for American and Canadian PD patients over time (measured in Months).



The solid line represents the estimated cumulative mean difference in days hospitalized (Canadian minus American). The dashed lines represent the corresponding 95% confidence intervals.

Figure 4.2: Analysis of the CANUSA data: point estimates and 95% confidence intervals for the estimated cumulative mean difference in days hospitalized $\hat{\delta}(t) = \hat{\mu}_1(t) - \hat{\mu}_0(t)$, between American ($j = 0$) and Canadian ($j = 1$) PD patients

CHAPTER V

Conclusion

5.1 Conclusion

In many biomedical and health related studies, the study subjects may experience a sequence of recurrent events (e.g., repeated hospital admissions, multiple tumor occurrences). Often times, there are outcome measures that describe either quantitative or qualitative aspects of the recurrent event, which we define as "marks". In addition, the sequence of recurrent events may potentially be stopped by a terminating event (e.g., death), especially during clinical trials or epidemiologic studies. In this dissertation, we developed three novel methodologies that take into account of the association between the recurrent event, marks and the terminating event.

In Chapter II, we developed semi-parametric methods to contrast group-specific means. We utilized a proportional hazards model for the terminating event hazards, a proportional rates model for the recurrent event given survival and a generalised estimating equation type model for the marks given each recurrent event. The marginal mean estimator is built from the above three component models. Our estimator takes into account potential treatment imbalances between groups by averaging over the marginal covariate distribution, analogous to average causal effect estimators.

Since study subjects often exhibit unmeasured heterogeneity (e.g., unequal risk of death or different rates of disease recurrence, even after conditioning on covariates),

we developed joint modelling techniques in Chapter III via a "frailty" term. We assumed a piecewise constant proportional hazards model for the terminating event hazards, a piecewise constant proportional rates model for the recurrent event rate given survival and a Poisson regression model for the marks given hospitalization. The variance of the frailty term quantifies the scale of the dependence among the recurrent events, terminating event and marks. Unlike Chapter II where the parameters are estimated separately for each of the individual models, parameters in Chapter III are estimated simultaneously via maximum likelihood.

In Chapter IV, we propose inverse weighting methods to contrast the marginal means. We employ an Inverse Probability of Censoring Weighting (IPCW) to account for censoring and Inverse Probability of Treatment Weighting (IPTW) to tackle treatment imbalances. The IPTW weight is built through a logistic regression model while the IPCW weight is constructed through a proportional hazards model for censoring. Our proposed estimator in this Chapter are considered solely non-parametric if the two weights mentioned above are absent. This is in contrast to the proposed estimator in Chapter II where we propose a regression model for each component process.

In summary, we developed three novel methods that accommodate the data structure of our interest. Each method addresses the problem from a different angle. There are also pros and cons in employing each method. For example, by building separate models for each process (as in Chapter II), ideally we can achieve efficiency gain if the models describe the data well. On the other hand, we could achieve robust results by implementing an estimator similar to the one in Chapter IV even with model misspecification. Heterogeneity effect is solely considered in Chapter III, which is not modelled in either Chapter II or Chapter IV. Although we considered all three processes (e.g., terminating event process, recurrent event process and mark process) in Chapter II, III and IV, our proposed methods could still be implemented if we only have two of the three processes. For example, if we only have terminating event

process and recurrent event process, we can just treat the marks associated with each recurrent event to be identity. Likewise, if we ignore the terminating event process, our proposed methods could still be applied.

APPENDICES

APPENDIX A

Proofs of Theorems in Chapter II

The proof of consistency and the derivation of the large-sample distribution are the same for $\hat{\mu}_1(t)$ and $\hat{\mu}_0(t)$, $j = 0$ or 1. Following the proofs for $\hat{\mu}_1(t)$ and $\hat{\mu}_0(t)$, the consistency and distribution of $\hat{\delta}(t)$ can be directly obtained, as is evident from the development that follows.

Proof of Theorem II.1:

Note that $\hat{\mu}_j(t) = n^{-1} \sum_{i=1}^{n} \hat{\mu}_j(t; \hat{\beta}_D, \hat{\beta}_R, \hat{\beta}_G, \hat{\Lambda}_{0j}, \hat{R}_{0j}|Z_i)$. $\tilde{\mu}_j(t) = n^{-1} \sum_{i=1}^{n}$ $\mu_j(t; \beta_D, \beta_R, \beta_G, \Lambda_{0j}, R_{0j}|Z_i)$ and $\mu_j(t) = E[\mu_j(t; \beta_D, \beta_R, \beta_G, \Lambda_{0j}, R_{0j}|Z_i)]$. Since $\hat{\beta}_D \xrightarrow{a.s.}$ β_D, $\hat{\beta}_R \xrightarrow{a.s.} \beta_R$, $\hat{\beta}_G \xrightarrow{a.s.} \beta_G$, $\hat{\Lambda}_{0j}(t) \xrightarrow{a.s.} \Lambda_{0j}(t)$ and $\hat{R}_{0j}(t) \xrightarrow{a.s.} R_{0j}(t)$, by Continuous Mapping Theorem, $\hat{\mu}_j(t) \xrightarrow{a.s.} \tilde{\mu}_j(t)$, for all t $\in (0, \tau]$. Then applying the uniform strong law of large numbers (USSL; Pollard, 1990), we have $\tilde{\mu}_j(t) \xrightarrow{a.s.} \mu_j(t)$.

Proof of Theorem II.2:

We can first decompose $n^{\frac{1}{2}}\{\hat{\mu}_j(t) - \mu_j(t)\}$ into six parts as follows

$$
\begin{aligned}
& n^{\frac{1}{2}}\{\hat{\mu}_j(t) - \mu_j(t)\} \\
= \; & n^{\frac{1}{2}}\{\hat{\mu}_j(t) - \tilde{\mu}_j(t)\} + n^{\frac{1}{2}}\{\tilde{\mu}_j(t) - \mu_j(t)\} \\
= \; & n^{-\frac{1}{2}} \sum_{i=1}^{n} \{\hat{\mu}_j(t; \hat{\beta}_D, \hat{\beta}_R, \hat{\beta}_G, \hat{\Lambda}_{0j}, \hat{R}_{0j}|Z_i) - \hat{\mu}_j(t; \beta_D, \hat{\beta}_R, \hat{\beta}_G, \hat{\Lambda}_{0j}, \hat{R}_{0j}|Z_i)\} \quad \text{(A.1)} \\
& + n^{-\frac{1}{2}} \sum_{i=1}^{n} \{\hat{\mu}_j(t; \beta_D, \hat{\beta}_R, \hat{\beta}_G, \hat{\Lambda}_{0j}, \hat{R}_{0j}|Z_i) - \hat{\mu}_j(t; \beta_D, \beta_R, \hat{\beta}_G, \hat{\Lambda}_{0j}, \hat{R}_{0j}|Z_i)\} \text{(A.2)} \\
& + n^{-\frac{1}{2}} \sum_{i=1}^{n} \{\hat{\mu}_j(t; \beta_D, \beta_R, \hat{\beta}_G, \hat{\Lambda}_{0j}, \hat{R}_{0j}|Z_i) - \hat{\mu}_j(t; \beta_D, \beta_R, \beta_G, \hat{\Lambda}_{0j}, \hat{R}_{0j}|Z_i)\} \text{(A.3)} \\
& + n^{-\frac{1}{2}} \sum_{i=1}^{n} \{\hat{\mu}_j(t; \beta_D, \beta_R, \beta_G, \hat{\Lambda}_{0j}, \hat{R}_{0j}|Z_i) - \hat{\mu}_j(t; \beta_D, \beta_R, \beta_G, \Lambda_{0j}, \hat{R}_{0j}|Z_i)\} \text{(A.4)} \\
& + n^{-\frac{1}{2}} \sum_{i=1}^{n} \{\hat{\mu}_j(t; \beta_D, \beta_R, \beta_G, \Lambda_{0j}, \hat{R}_{0j}|Z_i) - \mu_j(t; \beta_D, \beta_R, \beta_G, \Lambda_{0j}, R_{0j}|Z_i)\} \text{(A.5)} \\
& + n^{-\frac{1}{2}} \sum_{i=1}^{n} \{\mu_j(t; \beta_D, \beta_R, \beta_G, \Lambda_{0j}, R_{0j}|Z_i) - \mu_j(t)\} \quad \text{(A.6)}
\end{aligned}
$$

Next, we consider (A.1) to (A.6) in sequence. The first part can be expanded around β_D using a Taylor series as

$$
\begin{aligned}
(A.1) \quad = \quad & \frac{1}{n} \sum_{i=1}^{n} \int_0^t \left. \frac{\partial \hat{S}_{ij}(r^-|Z_{iD})}{\partial \beta_D'} \right|_{\beta_{D*}} n^{\frac{1}{2}}(\hat{\beta}_D - \beta_D) \hat{g}_{ij}(r|Z_{iG}) d\hat{R}_{ij}(r) \\
= \quad & -\frac{1}{n} \sum_{i=1}^{n} \int_0^t \Big[\exp\{-e^{\beta_{D*}' Z_{iD}} \hat{\Lambda}_0(r) + \beta_{D*}' Z_{iD}\} Z_{iD} \hat{\Lambda}_{0j}(r) \\
& - \exp\{-e^{\beta_{D*}' Z_{iD}} \hat{\Lambda}_0(r) + \beta_{D*}' Z_{iD}\} \int_0^r \frac{S_j^{(1)}(u; \beta_{D*})}{n S_j^{(0)}(u; \beta_{D*})^2} dN_i^D(u) \Big] \hat{g}_{ij}(r|Z_{iG}) d\hat{R}_{ij}(r) \\
& \times n^{\frac{1}{2}}(\hat{\beta}_D - \beta_D) \\
\xrightarrow{a.s.} \quad & -E[e^{\beta_D' Z_{iD}} \int_0^t S_{ij}(u^-|Z_{iD}) \int_0^u \{Z_{iD} - \bar{z}(r; \beta_D)\} d\Lambda_{0j}(r) g_{ij}(u|Z_{iG}) dR_{ij}(u)] \\
& \times n^{\frac{1}{2}}(\hat{\beta}_D - \beta_D).
\end{aligned}
$$

Through another Taylor expansion, we have $n^{\frac{1}{2}}(\hat{\beta}_D-\beta_D) = A_j^D(\beta_D)^{-1}n^{-\frac{1}{2}}\sum_{i=1}^n U_{ij}^D(\beta_D)$ as $n\to\infty$, such that

$$\begin{aligned}(A.1) &= -E[e^{\beta_D' Z_{iD}}\int_0^t S_{ij}(u^-|Z_{iD})\int_0^u\{Z_{iD}-\bar{z}(r;\beta_D)\}d\Lambda_{0j}(r)g_{ij}(u|Z_{iG})dR_{ij}(u)]\\ &\quad\times A_j^D(\beta_D)^{-1}n^{-\frac{1}{2}}\sum_{i=1}^n U_{ij}^D(\beta_D)\\ &= n^{-\frac{1}{2}}\sum_{i=1}^n\psi_{ij1}(t),\end{aligned}$$

as defined in Theorem II.2. Now consider the second part, as $n\to\infty$,

$$\begin{aligned}(A.2) &= \frac{1}{n}\sum_{i=1}^n\int_0^t\left\{\hat{S}_{ij}(r^-|Z_{iD})\hat{g}_{ij}(r|Z_{iG})\frac{\partial d\hat{R}_{ij}(r|Z_{iR})}{\partial\beta_R'}|_{\beta_R=\beta_{R*}}\right\}n^{\frac{1}{2}}(\hat{\beta}_R-\beta_R)\\ &= \frac{1}{n}\sum_{i=1}^n\int_0^t\hat{S}_{ij}(r^-|Z_{iD})\hat{g}_{ij}(r|Z_{iG})\\ &\quad\times\left\{Z_{iR}e^{\beta_{R*}Z_{iR}}d\hat{R}_{0j}(r)+e^{\beta_{R*}Z_{iR}}\frac{-S_j^{(1)}(r;\beta_{R*})}{S_j^{(0)}(r;\beta_{R*})^2}dN_i^R(r)\right\}n^{\frac{1}{2}}(\hat{\beta}_R-\beta_R)\\ &\xrightarrow{a.s.} -E\left[e^{\beta_R' Z_{iR}}\int_0^t S_{ij}(r^-|Z_{iD})g_{ij}(r|Z_{iG})\{Z_{iR}-\bar{z}(r;\beta_R)\}dR_{0j}(r)\right]n^{\frac{1}{2}}(\hat{\beta}_R-\beta_R).\end{aligned}$$

Through another Taylor expansion, $n^{\frac{1}{2}}(\hat{\beta}_R-\beta_R) = A_j^R(\beta_R)^{-1}n^{-\frac{1}{2}}\sum_{i=1}^n U_{ij}^R(\beta_R)$.
Therefore, as $n\to\infty$,

$$\begin{aligned}(A.2) &= -E\left[e^{\beta_R' Z_{iR}}\int_0^t S_{ij}(r^-|Z_{iD})g_{ij}(r|Z_{iG})\{Z_{iR}-\bar{z}(r;\beta_R)\}dR_{0j}(r)\right]\\ &\quad\times A_j^R(\beta_R)^{-1}n^{-\frac{1}{2}}\sum_{i=1}^n U_{ij}^R(\beta_R)\\ &= n^{-\frac{1}{2}}\sum_{i=1}^n\psi_{ij2}(t).\end{aligned}$$

where $\psi_{ij2}(t)$ is as defined in Theorem II.2. For the third part, we have:

$$\begin{aligned}(A.3) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t \hat{S}_{ij}(r^-|Z_{iD})\frac{\partial \hat{g}_{ij}(r|Z_{iG})}{\partial \beta_G'} g_{ij}(r|Z_{iG}) d\hat{R}_{ij}(r) n^{\frac{1}{2}}(\hat{\beta}_G - \beta_G) \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t \hat{S}_{ij}(r^-|Z_{iD})\frac{\partial \hat{g}_{ij}(r|Z_{iG})}{\partial \beta_G'} g_{ij}(r|Z_{iG}) d\hat{R}_{ij}(r) n^{\frac{1}{2}}(\hat{\beta}_G - \beta_G)\end{aligned}$$

Through another Taylor expansion, $n^{\frac{1}{2}}(\hat{\beta}_G - \beta_G) = E\left[\frac{\partial U_{ij}^G(\beta_G)}{\partial \beta_G'}\right]^{-1} n^{-\frac{1}{2}} U_{ij}^G(\beta_G)$. Therefore,

$$\begin{aligned}(A.3) &\xrightarrow{a.s.} E\left[\int_0^t S_{ij}(r^-|Z_{iD})\frac{\partial g_{ij}(r|Z_{iG})}{\partial \beta_G'} g_{ij}(r|Z_{iG}) dR_{ij}(r)\right]\left[E\frac{\partial U_{ij}^G(\beta_G)}{\partial \beta_G'}\right]^{-1} n^{-\frac{1}{2}} U_{ij}^G(\beta_G) \\ &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\psi_{ij3}(t),\end{aligned}$$

with $\psi_{ij3}(t)$ defined as in Theorem II.2. For the fourth part, as $n \to \infty$,

$$\begin{aligned}(A.4) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t n^{\frac{1}{2}}\{\hat{S}_{ij}(r^-|Z_{iD}) - S_{ij}(r^-|Z_{iD})\}\hat{g}_{ij}(r|Z_{iG}) d\hat{R}_{ij}(r;\beta_R|Z_{iR}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t n^{\frac{1}{2}}\{\exp\{-\hat{\Lambda}_{0j}(r;\beta_D)e^{\beta_D' Z_{iD}}\} - \exp\{-\Lambda_{0j}(r)e^{\beta_D' Z_{iD}}\}\}\hat{g}_{ij}(r|Z_{iG}) \\ &\quad \times \hat{R}_{ij}(r;\beta_R|Z_{iR}) \\ &= -\int_0^t \frac{1}{n}\sum_{i=1}^{n}\{e^{-\Lambda_{0j}(r)e^{\beta_D' Z_{iD}}} e^{\beta_D' Z_{iD}}\}\hat{g}_{ij}(r|Z_{iG}) d\hat{R}_{ij}(r;\beta_R|Z_{iR}) \\ &\quad \times n^{\frac{1}{2}}\{\hat{\Lambda}_{0j}(r;\beta_D) - \Lambda_{0j}(r)\} \\ &\xrightarrow{a.s.} -\int_0^t E[e^{\beta_D Z_{iD}} S_{ij}(r|Z_{iD}) g_{ij}(r|Z_{iG}) dR_{ij}(r|Z_{iR})] n^{\frac{1}{2}}\{\hat{\Lambda}_{0j}(r;\beta_D) - \Lambda_{0j}(r)\}\end{aligned}$$

Using a zero-mean structure,

$$
\begin{aligned}
& n^{\frac{1}{2}}\{\hat{\Lambda}_{0j}(r;\beta_D) - \Lambda_{0j}(r)\} \\
= \;& n^{-\frac{1}{2}} \left\{ \int_0^r \frac{\sum_{i=1}^n dN_i^D(u;\beta_D)}{\sum_{i=1}^n Y_i(u) e^{\beta_D' Z_{iD}}} - \Lambda_{0j}(r) \right\} \\
= \;& n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^r \frac{dM_{ij}^D(u;\beta_D)}{S_j^{(0)}(u;\beta_D)} \\
= \;& n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^r \frac{dM_{ij}^D(u;\beta_D)}{s_j^{(0)}(u;\beta_D)} + n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^r [S_j^{(0)}(u;\beta_D)^{-1} - s_j^{(0)}(u;\beta_D)^{-1}] dM_{ij}^D(u;\beta_D)
\end{aligned}
$$

The second term goes to zero by strong convergence, the Continuous Mapping Theorem and USLLN. Therefore, we have

$$
n^{\frac{1}{2}}\{\hat{\Lambda}_{0j}(r;\beta_D) - \Lambda_{0j}(r)\} \xrightarrow{a.s.} n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^r \frac{dM_{ij}^D(u;\beta_D)}{s_j^{(0)}(u;\beta_D)}. \tag{A.7}
$$

Combining the above argument related to the forth part,

$$
\begin{aligned}
(A.4) \quad \xrightarrow{a.s.} \quad & -\int_0^t E[e^{\beta_D Z_{iD}} S_{ij}(r|Z_{iD}) g_{ij}(r|Z_{iG}) dR_{ij}(r|Z_{iR})] n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^r \frac{dM_{ij}^D(u;\beta_D)}{s_j^{(0)}(u;\beta_D)} \\
= \quad & n^{-\frac{1}{2}} \sum_{i=1}^n \psi_{ij4}(t).
\end{aligned}
$$

For the fifth part, as $n \to \infty$,

$$
n^{\frac{1}{2}}\{\hat{R}_{0j}(r;\beta_R) - R_{0j}(r)\} \xrightarrow{a.s.} n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^r \frac{dM_{ij}^R(u;\beta_R)}{s_j^{(0)}(u;\beta_R)},
$$

Therefore, based on argument which parallel those in the derivation of (A.7)

$$
\begin{aligned}
(A.5) \quad &= \int_0^t \frac{1}{n}\sum_{i=1}^{n} S_{ij}(r^-|Z_{iD}) g_{ij}(r|Z_{iG}) e^{\beta_R Z_{iR}} n^{\frac{1}{2}} \{d\hat{R}_{0j}(r;\beta_R) - dR_{0j}(r)\} \\
&\xrightarrow{a.s.} \int_0^t \frac{1}{n}\sum_{i=1}^{n} S_{ij}(r^-|Z_{iD}) g_{ij}(r|Z_{iG}) e^{\beta_R Z_{iR}} n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{dM_{ij}^R(r;\beta_R)}{s_j^{(0)}(r;\beta_R)} \\
&\xrightarrow{a.s.} \int_0^t E\{S_{ij}(r^-|Z_{iD}) g_{ij}(r|Z_{iG}) e^{\beta_R Z_{iR}}\} n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{dM_{ij}^R(r;\beta_R)}{s_j^{(0)}(r;\beta_R)} \\
&= n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi_{ij5}(t).
\end{aligned}
$$

For the last part, it is straightforward to show that $(A.6) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi_{ij6}(t)$, as $n \to \infty$.

At last, combining the above six terms, we have $n^{\frac{1}{2}}\{\hat{\mu}_j(t) - \mu_j(t)\} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi_{ij}(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{6} \psi_{ijk}(t)$, completing the proof.

APPENDIX B

Proofs of Theorems in Chapter IV

Proof of Theorem IV.1: $\hat{\mu}_j(t) = n^{-1}\sum_{i=1}^{n}\int_0^t w_{ij}^A(\hat{\theta})\hat{w}_{ij}^C(s)G_i(s)dN_i^R(s)$. Acknowledging the fact that $\hat{w}_{ij}^C(s) \xrightarrow{p} w_{ij}^C(s)$ and $w_{ij}^A(\hat{\theta}) \xrightarrow{p} w_{ij}^A(\theta_0)$, then applying WLLN and using continuity,

$$\hat{\mu}_j(t) \xrightarrow{p} \int_0^t E[w_{ij}^A(\theta_0)w_{ij}^C(s)G_i(s)dN_i^R(s)] = \mu_j(t).$$

$$\begin{aligned} n^{1/2}\{\hat{\mu}_j(t) - \mu_j(t)\} &= n^{1/2}\{\hat{\mu}_j(t;\hat{w}^A,\hat{w}^C) - \hat{\mu}_j(t;w^A,\hat{w}^C)\} && \text{(B.1)} \\ &\quad + n^{1/2}\{\hat{\mu}_j(t;w^A,\hat{w}^C) - \hat{\mu}_j(t;w^A,w^C)\} && \text{(B.2)} \\ &\quad + n^{1/2}\{\hat{\mu}_j(t;w^A,w^C) - \mu_j(t)\} && \text{(B.3)} \end{aligned}$$

For the first term in the decomposition, we can write

$$
\begin{aligned}
(B.1) &= n^{-1/2}\sum_{i=1}^{n}\int_0^t \{w_{ij}^A(\hat{\theta}) - w_{ij}^A(\theta_0)\}\hat{w}_{ij}^C(s)G_i(s)dN_i^R(s) \\
&= n^{-1}\sum_{i=1}^{n}\int_0^t (-1)^j A_{ij}\frac{1-p_{ij}(\theta_0)}{p_{ij}(\theta_0)} I(\theta_0)^{-1}n^{-1/2}\sum_{l=1}^{n} U_l(\theta_0)\hat{w}_{ij}^C(s)G_i(s)dN_i(s) \\
&= H_j^{'}(t)I(\theta_0)^{-1}n^{-1/2}\sum_{i=1}^{n} U_i(\theta_0) \\
&= n^{-1/2}\sum_{i=1}^{n}\psi_{ij1}(t),
\end{aligned}
$$

where $H_j^{'}(t)$ and $I(\theta_0)$ are defined in Theorem IV.1. For the second term, we can further decompose as follows

$$
\begin{aligned}
(B.2) &= n^{1/2}\{\hat{\mu}_j(t; w^A, \hat{w}^C(t;\hat{\beta}_C, \hat{\Lambda}_{0j}^C)) - \hat{\mu}_j(t; w^A, \hat{w}^C(t;\beta_C, \hat{\Lambda}_{0j}^C))\} \qquad \text{(B.4)} \\
&\quad + n^{1/2}\{\hat{\mu}_j(t; w^A, \hat{w}^C(t;\beta_C, \hat{\Lambda}_{0j}^C)) - \hat{\mu}_j(t; w^A, \hat{w}^C(t;\beta_C, \Lambda_{0j}^C))\} \qquad \text{(B.5)}
\end{aligned}
$$

For the first piece in the above decomposition, we have

$$
\begin{aligned}
(B.4) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t w_{ij}^A(\hat{\theta})\frac{\partial w_{ij}^C(s)}{\partial \beta_C^{'}}\bigg|_{\beta_C^*} G_i(s)dN_i^R(s) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^t [\exp\left\{\int_0^s \exp\{\beta_C^{'}Z_i(u)\}d\hat{\Lambda}_{0j}^C(u) + \beta_C^{'}Z_i(s)\right\} Z_i(s)\hat{\Lambda}_{0j}^C(s) \\
&\quad - \exp\left\{\int_0^s \exp\{\beta_C^{'}Z_i(u)\}d\hat{\Lambda}_{0j}^C(u) + \beta_C^{'}Z_i(s)\right\}\int_0^s \frac{S_j^{(1)}(u;\beta_C)}{nS_j^{(0)}(u;\beta_C)^2}dN_i^C(u)] \\
&\quad \times G_i(s)dN_i^R(s)n^{\frac{1}{2}}(\hat{\beta}_C - \beta_C) \\
&\quad \xrightarrow{a.s} E\left[w_{ij}^A(\hat{\theta})\int_0^t w_{ij}^C(s)\exp\{\beta_C^{'}Z_i(s)\}\int_0^s Z_i(u) - \bar{z}(u;\beta_C)d\Lambda_{0j}(u)G_i(s)dN_i^R(s)\right] \\
&\quad \times n^{1/2}(\hat{\beta}_C - \beta_C)
\end{aligned}
$$

Through a Taylor series expansion, we have $n^{1/2}(\hat{\beta}_C - \beta_C) = A_j^C(\beta_C)^{-1} n^{-1/2} \sum_{i=1}^n U_{ij}^C(\beta_C)$ as $n \to \infty$.

$$
\begin{aligned}
(B.4) \quad \xrightarrow{a.s} \quad & E[w_{ij}^A(\hat{\theta}) \int_0^t w_{ij}^C(s) \exp\{\beta_C' Z_i(s)\} \int_0^s Z_i(u) - \bar{z}(u;\beta_C) d\Lambda_{0j}(u) G_i(s) dN_i^R(s)] \\
& \times A_j^C(\beta_C)^{-1} n^{-1/2} \sum_{i=1}^n U_{ij}^C(\beta_C)
\end{aligned}
$$

For the second piece in the above decomposition, we have

$$
\begin{aligned}
(B.5) \quad = \quad & \frac{1}{n} \sum_{i=1}^n \int_0^t w_{ij}^A(\theta_0) n^{1/2} \left\{ \exp\{ \int_0^s e^{\beta_C Z_i(u)} d\hat{\Lambda}_{0j}(u;\beta_C) \} \right. \\
& \left. - \exp\{ \int_0^s e^{\beta_C Z_i(u)} d\Lambda_{0j}(u;\beta_C) \} \right\} G_i(s) dN_i(s) \\
= \quad & \frac{1}{n} \sum_{i=1}^n \int_0^t w_{ij}^A(\theta_0) \exp\left\{ \int_0^s e^{\beta_C Z_i(u)} d\Lambda_{0j}(u;\beta_C) \right\} e^{\beta_C Z_i(s)} \\
& \times n^{1/2} \{ \hat{\Lambda}_{0j}(s;\beta_C) - \Lambda_{0j}(s;\beta_C) \} G_i(s) dN_i(s) \\
\xrightarrow{a.s.} \quad & \int_0^t E[w_{ij}^A(\theta_0) \exp\{\beta_C Z_i(s)\} w_{ij}^C(s) G_i(s) dN_i(s)] n^{1/2} \{ \hat{\Lambda}_{0j}(s;\beta_C) - \Lambda_{0j}(s;\beta_C) \}
\end{aligned}
$$

In the above term,

$$
\begin{aligned}
& n^{1/2} \left\{ \hat{\Lambda}_{0j}^C(s;\beta_C) - \Lambda_{0j}^C(s;\beta_C) \right\} \\
= \quad & n^{-\frac{1}{2}} \{ \int_0^s \frac{\sum_{i=1}^n dN_i^C(u;\beta_C)}{\sum_{i=1}^n Y_i(u) e^{\beta_C' Z_i(u)}} - \Lambda_{0j}^C(u;\beta_C) \} \\
= \quad & n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^s \frac{dM_{ij}^C(u;\beta_C)}{S_j^{(0)}(u;\beta_C)} \\
= \quad & n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^s \frac{dM_{ij}^C(u;\beta_C)}{s_j^{(0)}(u;\beta_C)} + n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^s [S_j^{(0)}(u;\beta_C)^{-1} - s_j^{(0)}(u;\beta_C)^{-1}] dM_{ij}^C(u;\beta_C)
\end{aligned}
$$

The second term goes to zero by the almost sure convergence, continuous mapping

theorem and USLLN. Therefore, we have

$$n^{\frac{1}{2}}\{\hat{\Lambda}_{0j}^{C}(s;\beta_C) - \Lambda_{0j}^{C}(s;\beta_C)\} \xrightarrow{a.s.} n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^s \frac{dM_{ij}^{C}(u;\beta_C)}{s_j^{(0)}(u;\beta_C)}.$$

Therefore, combining the above argument, we obtain

$$\begin{aligned}
(B.2) \quad \xrightarrow{a.s.} \quad & E\left[w_{ij}^{A}(\hat{\theta}) \int_0^t w_{ij}^{C}(s) \exp\{\beta_C' Z_i(s)\} \int_0^s \{Z_i(u) - \bar{z}(u;\beta_C)\}\, d\Lambda_{0j}(u) G_i(s) dN_i^{R}(s) \right] \\
& \times A_j^{C}(\beta_C)^{-1} n^{-1/2} \sum_{i=1}^{n} U_{ij}^{C}(\beta_C) \\
& + \int_0^t E[w_{ij}^{A}(\theta_0) \exp\{\beta_C Z_i(s)\} w_{ij}^{C}(s) G_i(s) dN_i(s)] n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^s \frac{dM_{ij}^{C}(u;\beta_C)}{s_j^{(0)}(u;\beta_C)} \\
= \quad & n^{-1/2} \sum_{i=1}^{n} \psi_{ij2}(t),
\end{aligned}$$

with $\psi_{ij2}(t)$ defined in Theorem IV.1. For the last term, it is straightforward that

$$\begin{aligned}
(B.3) \quad = \quad & n^{-1/2} \sum_{i=1}^{n} \int_0^t \{w_{ij}^{A}(\theta_0) w_{ij}^{C}(s) G_i(s) dN_i(s) - d\mu_j(s)\} \\
= \quad & n^{-1/2} \sum_{i=1}^{n} \psi_{ij3}(t)
\end{aligned}$$

After combining (B.1)-(B.3), we have, $n^{1/2}\{\hat{\mu}_j(t) - \mu_j(t)\} = n^{-1/2} \sum_{i=1}^{n} \psi_{ij}(t)$ $= n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{3} \psi_{ijk}(t)$, which is a scaled sum of independent and identically distributed zero mean random variates. By the Central Limit Theorem, $n^{1/2}\{\hat{\mu}_j(t) - \mu_j(t)\}$ converges to a zero mean Gaussian process with covariance function $\sigma_j(s,t) = E[\psi_{ij}(s)\psi_{ij}(t)]$.

BIBLIOGRAPHY

BIBLIOGRAPHY

Anstrom, K. J., and A. A. Tsiatis (2001), Utilizing Propensity Scores to Estimate Causal Treatment Effects with Censored Time-Lagged Data, *Biometrics*, *57*, 1207–1218.

Bilias, Y., M. Gu, and Z. Ying (1997), Towards a General Asymptotic Theory for the Cox Model with Staggered Entry, *Annals of Statisitics*, *25*, 662–682.

Breslow, N., and D. Clayton (1993), Approximate Inference on Generalized Linear Mixed Models, *JASA*, *88*, 9–25.

Cai, J., and D. E. Schaubel (2004), Marginal Means and Rates Models for Multi-type Recurrent Event Data, *Lifetime Data Analysis*, *10*, 121–138.

Cai, J., D. Zeng, and W. Pan (2010), Semiparametric Proportional Means Model for Marker Data Contigent on Recurrent Event, *Lifetime Data Analysis*, *16*, 250–270.

Canada-U.S.A. Peritoneal Dialysis Study Group (1996), Adequacy of Dialysis and Nutrition in Continuous Peritoneal Dialysis: Correlation with Clinical Outcomes, *Journal of the American Society of Nephrology*, *7*, 198–207.

Chen, P., and A. A. Tsiatis (2001), Causal Inference on the Difference of the Restricted Mean Lifetime between Two Groups, *Biometrics*, *57*, 1030–1038.

Churchill, D. N., K. E. Thorpe, K. Nolph, D. G. Oreopoulos, and P. D. (1998), Increased Peritoneal Membrane Transport is Associated with Decreased Patient and Technique Survival for Continuous Peritoneal Dialysis Dialysis Patients, *J Am Soc Nephrol*, *9*, 1285–1292.

Clayton, D. G. (1978), A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence, *Biometrika*, *65*, 141–151.

Cook, R. J., and J. F. Lawless (1997), Marginal Analysis of Recurrent Events and a Terminating Event, *Statistics in Medicine*, *16*, 911–924.

Cox, D. R. (1972), Regression Models and Life-Tables, *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*, 187–220.

Ghosh, D., and D. Y. Lin (2000), Nonparametric Analysis of Recurrent Events and Death, *Biometrics*, *56*, 554–562.

Ghosh, D., and D. Y. Lin (2002), Marginal Regression Models for Recurrent and Terminal Event, *Statistical. Sinica*, *12*, 663–688.

Hernan, M. A., B. Brumback, and J. M. Robins (2000), Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-positive Men, *Epidemiology*, *11*, 561–570.

Huang, C. Y., and M. C. Wang (2004), Joint Modeling and Estimation for Recurrent Event Processes and Failure Time Data, *Journal of American Statistical Association*, *99*, 1153–1165.

Huang, Y. (2002), Calibration Regression of Censored Lifetime Medical Cost, *Journal of American Statisitical Association*, *97*, 318–327.

Lawless, J., and J. Nadeau (1995), Some Simple Robust Methods for the Analysis of Recurrent Events, *Technometrics*, *37*, 18–68.

Li, Q., and S. Lagakos (1997), Use of the Wei-Lin-Weissfeld Method for the Analysis of a Recurring and a Terminating Event, *Statistics in Medicine*, *16*, 925–940.

Liang, K. Y., and S. L. Zeger (1986), Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika*, *73*, 13–22.

Lin, D. Y. (2000), Proportional Means Regression for Censored Medical Costs, *Biometrics*, *56*, 775–778.

Lin, D. Y., L. J. Wei, I. Yang, and Z. Ying (2000), Semiparametric Regression for the Mean and Rate Functions of Recurrent Events, *Journal of the Royal Statisitical Society, Series B*, *62*, 711–730.

Liu, L., R. A. Wolfe, and X. Huang (2004), Shared Frailty Models for Recurrent Events and a Terminating Event, *Biometrics*, *60*, 747–756.

Liu, L., M. R. Conaway, W. A. Knaus, and J. Bergin (2008), A Random Effects Four-Part Model with Application to Correlated Medical Costs, *Computational Statistics and Data Analysis*, *52*, 4458–4473.

McGilchrist, C. A., and C. W. Aisbett (1991), Regression with Frailty in Survival Analysis, *Biometrics*, *47*, 461–466.

Nielsen, G. G., R. D. Gill, P. K. Andersen, and T. I. A. Sorensen (1992), A Counting Process Approach to Maximum Likehood Estimation in Frailty Models, *Scandinavian Journal of Statistics*, *19*, 25–43.

Pan, Q., and D. E. Schaubel (2009), Flexible Estimation of Differences in Treatment-Specific Recurrent Event Means in the Presence of a Terminating Event, *Biometrics*, *65*, 753–761.

Pepe, M., and J. Cai (2004), Some Graphical Displays and Marginal Regression Analyses for Recurrent Failure Times and Time Dependent Covariates, *Journal of the American Statisitical Association*, *88*, 811–820.

Pollard, D. (1990), Empirical Processes: Theory and Applications, *Hayward, California: Institute of Mathematical Statistics*.

Robins, J. M. (1993), Information Recovery and Bias Adjustment in Proportional Hazards Regression Analysis of Randomized Trials Using Surrogate Markers, *In: Proceedings of the Biopharmaceutical Section, American Statistical Association, Alexandria, VA*.

Robins, J. M., and A. Rotnitzky (1992), Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers, *In: Jewell N, Dietz K, Farewell V (eds) AIDS Epidemiology-Methodological Issues, Proceedings of the Biopharmaceutical Section, American Statisitical Association, Alexandria VA*.

Robins, J. M., M. Hernan, and B. Brumback (2000), Marginal Structural Models and Causal Inference in Epidemiology, *Epidemiology*, *11*, 550–560.

Schaubel, D. E., and J. Cai (2005), Analysis of Clustered Recurrent Event Data with Application to Hospitalization Rates Among Renal Failure Patients, *Biostatistics*, *6*, 404–419.

Schaubel, D. E., and M. Zhang (2010), Estimating Treatment Effects on the Marginal Recurrent Event Mean in the Presence of a Terminating Event, *Lifetime Data Analysis*, *16*, 451–477.

Sen, P. K., and J. M. Singer (1993), Large Sample Methods in Statistics, *New York: Chapman and Hall*.

Tierney, L. J., and J. B. Kadane (1986), Accurate Approximations for Posterior Moments and Marginal Densities, *JASA*, *81*, 82–86.

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

Vaupel, J. W., K. G. Manton, and E. Stallard (1979), The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality, *Demography*, *16*, 439–454.

Wang, M. C., J. Qin, and C. T. Chiang (2001), Analyzing Recurrent Event Data with Informative Censoring, *American Statistical Association*, *96*, 1057–1065.

Ye, Y., J. D. Kalbfleisch, and D. Schaubel (2007), Semiparametric Analysis of Correlated Recurrent and Terminal Events, *Biometrics*, *63*, 78–87.