Flexible Methods for Clustered Event History Data

by

Dandan Liu

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

Doctoral Committee:

Professor John D. Kalbfleisch, Co-Chair
Associate Professor Douglas E. Schaubel, Co-Chair
Professor John R.C. Wheeler
Assistant Professor Min Zhang

ACKNOWLEDGEMENTS

This dissertation is a summary of my PhD study in the Department of Biostatistics at the University of Michigan. I can not finish it without the support and encouragement from my advisors Jack and Doug, who had the most important influence on my research and career. They leaded me step by step into the field of biostatistics and showed me with their respectful personalities how successful scientists behave. I received five-years research assistantship from the Kidney Epidemiology and Costs Center, which provided me enriched resources for motivation and application of my dissertation and precious collaboration experience. In addition, my other committee members, Jack and Min, have contributed valuable suggestion to the dissertation. I would also like to thank my parents and in-laws for their great love and support, and my husband, Dong, for everything. Finally, I would like to dedicate this dissertation to my beloved daughter, Flora, whose birth last year gave my the greatest power to finish my study.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER I

Introduction

This dissertation investigates three interesting problems in event history data. Chapter II considers the positive stable frailty model for clustered failure time data, but allows the frailty distribution to depend on cluster-level covariates. Chapter III proposes a computationally efficient marginal proportional rates model with piecewise-constant baseline rate function for recurrent event data and illustrates its implementation using SAS. Chapter IV introduces a new estimation method for fixed center effects under the proportional rates model for recurrent event data for the setting where comparisons across centers are of interest.

In biomedical studies, researchers often encounter data in which failure times from the same cluster are correlated due to sharing genetic and perhaps environmental factors; e.g. failure times among family members. It is often of interest to estimate both the hazard ratios and the degree of heterogeneity across clusters. In the presence of dependence within clusters, there are generally two approaches that can be adopted, with each addressing different aspects of the quantities of interest. Marginal hazards models can be used when comparisons between any two randomly selected subjects (i.e., possibly from different clusters) are of interest. For example, inference procedures for the regression parameters and baseline hazard function un-

der the marginal proportional hazards models have been established by Wei, Lin, and Weissfeld (1989); Lee, Wei, and Amato (1992) and Spiekerman and Lin (1998). However, the within-cluster dependence structure is left unspecified in the formulation of these models, which may result in loss of efficiency. When within-cluster interpretations of hazard ratios are of interest, the frailty (random effects) model may be preferred. In such cases, the degree of between-cluster heterogeneity is then described by the parameters indexing the distribution of the frailty; e.g. the scale parameter of a gamma distribution with unit mean. In many cases, a common frailty distribution is imposed, assuming equal intra-cluster dependence. However, this assumption may be violated. Moreover, it is typically inconvenient to obtain a marginal (population-averaged) interpretation of covariate effects using frailty model unless the frailty distribution follows a particular form, as in Glidden and Self (1999) and Pipper and Martinussen (2003) for Clayton-Oakes model.

In Chapter II, we propose a positive stable proportional hazards frailty model for clustered failure time data, with the frailty distribution allowed to depend on cluster-level covariates. We consider the positive stable frailty distribution due to its mathematical property enabling proportionality for both the marginal and conditional hazard functions. As a result, the marginal regression parameter equals the product of the dependence parameter and the conditional regression parameter. We propose a two-step estimation approach which capitalizes on this special feature. In addition, the model we propose connects the cluster-level covariate with the dependence parameter of the frailty through a logit link function, since the range of the dependence parameter is between 0 and 1. We establish large-sample properties and assess the performance of the proposed estimation procedure under finite samples through simulation. The proposed model is then applied to national kidney

transplant data.

Recurrent events represent another type of event history data, in which successive event occurrences within-subject are observed over time. The data structure is similar to that of clustered failure time data in that within-subject event times are correlated. When recurrent event data are subject to clustering, within-cluster correlation is present additionally, e.g., hospital admissions among patients treated at the same facility. Methods dealing with recurrent event data have been well developed and incorporated in standard statistical software (e.g. Andersen and Gill, 1982; Lin, Wei, Yang, and Ying, 2000; Lawless and Nadeau, 1995; Schaubel and Cai, 2005). Most existing methods, in handling the baseline rate/intensity function nonparametrically, are carried out based on the risk set defined at each distinct recurrent event time, which includes subjects who are under observation and are at risk of recurrent events. When the number of distinct recurrent event times is very large, those methods are usually slow in computation. For example, consider an analysis of hospitalization rates among U.S. dialysis patients over a span of three years. In this case, days of hospitalization (as opposed to hospital admissions) are the recurrent events of interest. A hospital admission with a length-of-stay of 8 days would then consist of 8 recurrent events, with event times at each day hospitalized. This study includes a total of 345,937 patients with an average of 22 hospital days during the 3-year period. Hence, the computation using traditional methods would be extremely slow for this study. If, on the other hand, we assume a piecewise-constant structure for the baseline rate/intensity function, recurrent event data can be grouped accordingly, leading to interval-grouped count data with appropriately defined exposure times for each interval. Such grouping for recurrent event data greatly reduces the computation time, especially when large databases are used or prevalence data is of

interest.

Various methods have been developed to analyze interval-grouped data, including the nonparametric methods of Thall and Lachin (1988); Sun and Kalbfleisch (1995); Wellner and Zhang (2000), semiparametric method of Sun and Wei (2000) and parametric model of Lawless and Zhan (1998). Many of these methods assume a common baseline rate function and may not be easily extended to the situation where the baseline is cluster-specific, especially when the number of clusters is relatively large. Moreover, most of the existing methods for interval-grouped recurrent event data cannot be carried out with standard software and thus have limited popularity.

The presence of a terminal event is a characteristic frequently encountered in recurrent event data. In practice, the occurrence recurrent events is typically stopped by the terminating event (e.g. death). Two popular approaches for analyzing recurrent/termial event data include the marginal model (e.g. Ghosh and Lin, 2002), which essentially averages the occurrence rate over the survival experience; and the partial marginal model (e.g. Cook and Lawless, 1997), which considers the recurrent event rate among survivors.

In Chapter III, we propose a partial marginal rates model for clustered recurrent event data with piecewise-constant baseline rate function. The parametric specification of the baseline rate function enables the modeling of intermittent counts instead of recurrent event times, which results in data reduction and remedies computational issues resulting from the size of the database. Within-subject correlations are accommodated by the proposed inference procedures through a robust covariance matrix. The proposed model is an extension of the Poisson model with piecewise-constant rate functions described in Cook and Lawless (2007) and can be implemented in standard statistical software for Cox regression (e.g. SAS, R). Asymptotic properties of

both the regression parameter estimates and the baseline rate functions are derived. Numerical studies are conducted to evaluate the asymptotic results of the estimators in finite samples. The proposed model is applied to national hospitalization data among dialysis patients.

In large registry studies such as the aforementioned national hospitalization study, comparisons among clinical centers are often of interest. Fixed effects models (FEM) are commonly used when the purpose of the analysis is to compare centers. In FEMs, center effects are usually estimated by incorporating center-specific indicator variables. However, the increasing dimension of the parameter space with the increasing number of centers might cause computational difficulties or even be infeasible, especially when the number of centers is large relative to sample size.

In Chapter IV, we study the estimation of center effects through fixed effects models for recurrent event data, applicable in both the presence or absence of a terminal event. In the proposed models, fixed center effects are assumed to multiplicatively influence the recurrent event rate functions. We then show that the center effect can be consistently estimated by the ratio of the observed cumulative number of events and the corresponding expected quantity based on a weighted average center effect and the same patient mix. The proposed estimation method is computationally more efficient since the dimension of the regression parameter space is not influenced by the number of centers. We provide asymptotic properties of the center effect estimators, as well as the cumulative baseline rate functions. The finite sample performance of the proposed estimators are investigated through simulation studies. We then apply the proposed method to a study of hospital admissions among dialysis patients using a national registry database.

CHAPTER II

A Positive Stable Frailty Model for Clustered Failure Time Data with Covariate Dependent Frailty

2.1 Introduction

Clustered failure time data are frequently observed in biomedical studies. For example, in the kidney transplantation setting, transplant failure times are of interest and can be taken as clustered failure times with transplant facilities as clusters. In family disease studies, time to disease onset is of interest and families are natural clusters. Subjects within cluster are correlated, with the intra-cluster dependence possibly due to sharing similar environmental and/or genetic conditions.

Several methods have been proposed for clustered failure time data. In general, these can be categorized into two broad strategies. In marginal models, the cluster structure is usually ignored when estimating the population averaged covariate effect, but is used to derive valid standard error estimates. Marginal models can be used when the comparison of lifetimes across clusters is of interest. Examples include Wei et al. (1989); Lee et al. (1992) and Spiekerman and Lin (1998). These authors used generalized estimating equations with an independence working assumption and the intra-cluster correlation structure left unspecified. As a result, some efficiency loss

may occur, potentially affecting the significance of estimated covariate effects.

When the comparison of lifetimes within the same cluster is of interest, frailty models may be more appropriate. In this case, the correlation structure is specified by incorporating a random effect (frailty) which is common to subjects within the same cluster. The covariate effect is then interpreted as being conditional on the frailties and is cluster specific. One can also obtain marginal covariate effects by making additional assumptions about the frailty distribution as was done by Glidden and Self (1999) and Pipper and Martinussen (2003) under the Clayton-Oakes model. In frailty models, it is usually assumed that the frailty variables follow the same distribution across clusters, which implies equal intra-cluster dependence as well as between-cluster heterogeneity. This assumption may be violated in practice.

In studies comparing U.S. kidney transplant centers to the national average, the ratio of observed to expected deaths, known as the standardized mortality ratio (SMR), is used, with the expected deaths obtained from a marginal Cox model. An SMR > 1 indicates a mortality rate above the national average. In the shared frailty model, this statistic is actually a nonparametric Poisson-type estimator (Glidden and Vittinghoff, 2004) for the corresponding frailty, given the observed data in the center. An investigation of the SMRs suggests that there may be greater heterogeneity for smaller facilities, since SMRs for smaller centers are more frequently seen at either the top or the bottom of the ordered list. Although this is partly due to sampling variance of the SMR estimator, it is also possible that an unequal degree of heterogeneity across centers results from varying cluster characteristics. This suggests a shared frailty model, but with the frailty distribution allowed to depend on cluster size. Other cluster level covariates may also have an effect on the frailty distribution. For example, urban transplant facilities may exhibit more uniform practices

than rural transplant hospitals, corresponding to less heterogeneity (smaller variance) for frailties of urban centers. In these examples of clustered failure time data, the population averaged effect is of primary interest. At the same time, however, the incorporation of cluster level covariate effects on the frailty distribution is of practical interest and should be considered.

Similar situations exist for other types of clustered data. Prentice (1986) proposed a regression model for clustered binary data, in which the correlation between pairs of binary observation within clusters were assumed to depend on cluster level covariates. Lin, Raz, and Harlow (1997) proposed a linear mixed model with heterogeneous within-cluster variances, where the within-cluster errors were assumed to follow a normal distribution with cluster-specific covariance matrix. Specifically, the variance of the measurement error was assumed to follow an inverse gamma distribution, where the mean depends on some linear combination of cluster level covariates through a log link. Heagerty (1999) proposed a marginally specified logistic-normal model for longitudinal binary data in which the marginal mean, rather than the conditional mean, was regressed on covariates. In addition, a conditional model on a Gaussian latent variable is specified, where the random effect additively influences the logit of the conditional mean. Wang and Louis (2004) further extended this method to clustered binary data, allowing the distribution parameters of the random effect to depend on some cluster level covariates. Their approach used a "bridge" distribution previously identified by Wang and Louis (2003) for the random effect to unify the form of the marginal and the conditional models. As a result, the conditional regression parameters can be expressed as functions of the marginal regression parameters and a parameter in the bridge distribution. Under this model, the regression parameter estimates have a direct marginal interpretation, while the

conditional regression parameter estimates can easily be obtained. Moreover, the influence of the cluster level covariates on the random effect can be estimated.

The positive stable distribution (Hougaard, 1986) serves as a bridge distribution for clustered failure time data under a Cox proportional hazards shared frailty model in the same sense as Wang and Louis (2003) since the resulting marginal regression parameter is a product of the conditional regression parameters and the frailty parameter. This relationship allows both marginal and conditional inference, while accounting for intra-cluster dependence. The shared positive stable frailty model has attracted renewed attention recently (e.g. Fine, Glidden, and Lee, 2003; Martinussen and Pipper, 2005).

In this chapter, we propose a covariate-dependent positive stable shared frailty model. The bridge type frailties are allowed to depend on cluster-level covariates and so to follow different distributions across clusters. Under this unified framework, the marginal regression parameters and the covariate effects on the frailty distribution can be consistently estimated. The major contributions of this project are the methods proposed for modeling the effects of the cluster-level covariates on the frailty distribution and the corresponding estimation of the marginal regression effects.

The remainder of this chapter is organized as follows. In Section 2.2, we introduce the proposed covariate dependent frailty model and describe the estimation procedures. We obtain the large sample properties of the model parameter estimators in Section 2.3 and Section 2.4 presents simulation studies. The proposed method is then applied to kidney transplant data from the Scientific Registry of Transplant Recipients (SRTR) in Section 2.5. In Section 2.6, we provide some concluding remarks and discussion. Proofs of the results are provided in an Appendix.

2.2 Model Specification and Estimation

2.2.1 The Positive Stable Shared Frailty Cox Proportional Hazards Model with Covariate Dependent Frailty

In this section, we specify a positive stable shared frailty Cox model, with the frailty distribution depending on cluster level covariates and the corresponding marginal hazard having a proportional hazards form. Our ultimate purpose is to estimate cluster level covariate effects on the frailty distribution, as well as the correlation within clusters and heterogeneity between clusters. We first define the Cox-type conditional and marginal hazard functions through the "bridge" property of the positive stable distribution. The relationship between the conditional hazard parameters, marginal hazard parameters and frailty distribution parameter can be obtained accordingly. Cluster level covariates are related to the frailty distribution parameter through a link function. Finally, we derive the individual intensity process given the observed history of all the individuals with the parameters of interest. We begin this section by establishing the requisite notation.

Suppose we have measurements from subjects in K clusters and that the cluster sizes n_k $(k = 1, 2, \ldots, K)$ are independent and identically distributed bounded random variables. Given n_k, let D_{ik} and C_{ik} be the failure and censoring times for the ith individual $(i = 1, \ldots, n_k)$ in the kth cluster; let $T_{ik} = D_{ik} \wedge C_{ik}$ be the follow-up time and $\Delta_{ik} = I(D_{ik} \leq C_{ik})$ the observed death indicator. Let W_k denote the positive stable distributed frailty with dependence parameter α_k for the kth cluster which we use to describes within-cluster dependence possibly due to unobserved covariate information. Let Z_{ik} be a p-vector of time independent covariates measured on individual (i, k). In addition, let X_k be a q-vector of time independent cluster level covariates which may influence α_k. Let $D_k = (D_{1k}, \ldots, D_{n_k k})$, with C_k and

Z_k defined similarly. We assume that $(D_k, C_k, Z_k, X_k, n_k, W_k)$ are independent and identically distributed for $k = 1, \ldots, K$. Define the at risk process $Y_{ik}(t) = I(T_{ik} \geq t)$ and the individual counting process $N_{ik}(t) = \Delta_{ik} I(T_{ik} \leq t)$. We define the filtrations

$$\mathcal{F}_t = \sigma\left\{N_{ik}(s), Y_{ik}(s), Z_{ik}, X_k, n_k : k = 1, \ldots, K, i = 1, \ldots, n_k, 0 \leq s \leq t\right\}$$

$$\mathcal{H}_t = \sigma\left\{N_{ik}(s), Y_{ik}(s), Z_{ik}, X_k, n_k, W_k : k = 1, \ldots, K, i = 1, \ldots, n_k, 0 \leq s \leq t\right\}.$$

Similar to Martinussen and Pipper (2005), we term $\mathcal{F}_t$ the observed filtration and $\mathcal{H}_t$ the conditional filtration.

We assume that W_k follows a positive stable distribution with shape parameter α_k $(0 < \alpha_k \leq 1)$. The positive stable distribution has been used by Hougaard (1986) for multivariate failure time data; its density function of and Laplace transform are respectively given by

$$f_{\alpha_k}(w) = -\frac{1}{\pi w} \sum_{i=1}^{\infty} \frac{\Gamma(i\alpha_k + 1)}{i!} (-w^{-\alpha_k})^i \sin(\alpha_k i \pi),$$

and

$$L(s) = \mathcal{E}\left\{\exp(-sW_k)\right\} = \exp(-s^{\alpha_k}) \quad (s \geq 0)$$

respectively.

Given (Z_k, X_k, W_k, n_k), the failure time D_{ik}, $i = 1, \ldots, n_k$ are assumed to be independent with hazard function

$$\lim_{h \to 0^+} P(t \leq D_{ik} \leq t + h | D_{ik} \geq t, Z_k, X_k, n_k, W_k)/h = W_k \lambda_{0k}(t) e^{\beta_k^T Z_{ik}}, \tag{2.1}$$

where $\lambda_{0k}(t)$ $(k = 1, \ldots, K)$ are unknown cluster specific baseline hazard functions and β_k $(k = 1, \ldots, K)$ are p-vectors of unknown cluster specific regression parameters, all of which rely on α_k through the derived marginal hazard function below.

Since W_k has the positive stable distribution, the marginal hazard function of D_{ik} is given by

$$\lim_{h\to 0^+} P(t \le D_{ik} \le t+h | D_{ik} \ge t, Z_k, X_k, n_k)/h = h_0(t) e^{\gamma^T Z_{ik}}, \tag{2.2}$$

where $h_0(t)$ is an unspecified baseline hazard and γ is a p-vector of unknown marginal regression parameters. In this, we have assumed a constant marginal log lazard ratio γ, which, given (2.1) and (2.2), impose the restriction $\gamma = \alpha_k \beta_k$, $k = 1, \ldots, K$. Note also that $\Lambda_{0k}(t) = H_0(t)^{\alpha_k^{-1}}$, where $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(s)ds$ and $H_0(t) = \int_0^t h_0(s)ds$.

We further relate X_k and α_k through a link function $\alpha_k = \widetilde{g}(\eta; X_k)$ and let $\alpha_k^{-1} = g(\eta; X_k)$, where η is a $(q+1)$-vector of unknown parameters. Here we assume that $g(\cdot)$ is monotone and twice differentiable with respect to η. Since $\alpha_k \in (0, 1]$, a natural choice for $\widetilde{g}$ is the logit link function and we set

$$g(\eta; X_k) = 1 + e^{-\eta^T \widetilde{X}_k}, \tag{2.3}$$

with $\widetilde{X}_k = (1, X_k^T)^T$ and $\eta = (\eta_1, \eta_2^T)^T$ where η_1 is a scalar intercept and η_2 is a q-vector of regression parameters.

In addition, we assume that the D_{ik} and C_{ik} are independent given Z_{ik} for $i = 1, \ldots, n_k$. Under this conditional independent censoring assumption, model (2.1) implies that the individual intensity process with respect to the conditional filtration $\mathcal{H}_t$ is

$$\lambda_{ik}(t|\mathcal{H}_{t^-}) = Y_{ik}(t) W_k \lambda_{0k}(t) e^{\beta_k^T Z_{ik}}. \tag{2.4}$$

By applying the innovation theorem (Andersen, Borgan, Gill, and Keiding, 1993) to (2.4) and inserting the link function (2.3), the individual intensity process with respect to the observed filtration $\mathcal{F}_t$ is

$$\lambda_{ik}(t|\mathcal{F}_{t^-}) \;=\; Y_{ik}(t) f_k(t) \lambda_{0k}(t) e^{g(\eta; X_k)\gamma^T Z_{ik}}, \tag{2.5}$$

where $f_k(t) = E(W_k|\mathcal{F}_{t^-})$ has the explicit form

$$f_k(t) = \frac{E_{W_k}\left[W_k^{N_{\cdot k}(t^-)+1} e^{-W_k \sum_{i=1}^{n_k} \int_0^{t^-} Y_{ik}(t) e^{g(\eta;X_k)\gamma^T Z_{ik}} dH_0^{g(\eta;X_k)}(s)}\right]}{E_{W_k}\left[W_k^{N_{\cdot k}(t^-)} e^{-W_k \sum_{i=1}^{n_k} \int_0^{t^-} Y_{ik}(t) e^{g(\eta;X_k)\gamma^T Z_{ik}} dH_0^{g(\eta;X_k)}(s)}\right]}, \tag{2.6}$$

with "$\cdot$" denoting summation over a subscript.

2.2.2 Estimation

Model (2.4) differs from the existing positive stable shared frailty Cox proportional model in that it allows the frailty distribution parameter α_k to depend on cluster level covariates, which induces the cluster specific conditional regression parameter $\beta_k = \alpha_k^{-1}\gamma$ and the cluster specific conditional baseline hazard $\lambda_{0k}(t)$. It can be easily seen that when $\eta_2 = 0$, α_k is a constant and the proposed model reduces to the common positive stable shared frailty model for which several estimation procedures have been developed. For example, Wang, Klein, and Moeschberger (1995) applied the E-M algorithm for parameter estimation. Fine et al. (2003) presented a simple estimation procedure which fitted a marginal model and stratified model separately and utilized the relationship $\alpha = \gamma/\beta$. Martinussen and Pipper (2005) proposed a likelihood based estimation procedure based on the individual intensity process with respect to an observed filtration similar to (2.5), but with $\alpha_k = \alpha$ and $\beta_k = \beta$. However, we are not able to extend these estimation procedures in the proposed model, since the regression parameter β_k in the conditional hazard is cluster specific.

As can be seen in the existing literature, simulations and applications of the positive stable shared frailty model are usually based on small clusters, such as twin or family studies, especially when the estimation of frailties is needed. In order to apply the positive stable frailty model to studies with large clusters, it is useful to

avoid the estimation of $f_k(t)$ in (2.6). We notice that model (2.5) can be written as

$$\lambda_{ik}(t|\mathcal{F}_{t^-}) \;=\; Y_{ik}(t)\widetilde{\lambda}_{0k}(t)e^{g(\eta;X_k)\gamma^T Z_{ik}},$$

where $\widetilde{\lambda}_{0k}(t) = \lambda_{0k}(t)f_k(t)$, which is actually a stratified Cox model, except that the covariate effect is cluster specific and depends on a function of cluster level covariates. The stratified partial likelihood approach (Cox, 1975; Kalbfleisch and Prentice, 2002) can be directly applied here. Due to the loss of information in $f_k(t)$ and the multiplicative relationship between g and γ, we cannot estimate the intercept term η_1 and the remaining parameters simultaneously. Therefore, our estimation procedure is actually based on two results from models (2.2) and (2.5) respectively.

Before proceeding, it is convenient to introduce the following two sets of notation for $k = 1.\ldots, K$ and $r = 0, 1, 2$,

$$\begin{aligned} S^{(r)}(\gamma,t) &= K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k} Y_{ik}(t)e^{\gamma^T Z_{ik}}Z_{ik}^{\otimes r}, \\ E(\gamma,t) &= S^{(1)}(\gamma,t)/S^{(0)}(\gamma,t), \quad V(\gamma,t) = S^{(2)}(\gamma,t)/S^{(0)}(\gamma,t) - \{E(\gamma,t)\}^{\otimes 2}, \end{aligned}$$

where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$, and

$$\begin{aligned} S_k^{(r)}(\eta;\gamma,t) &= \sum_{i=1}^{n_k} Y_{ik}(t)e^{g(\eta;X_k)\gamma^T Z_{ik}}\left\{g_1(\eta;X_k)\gamma^T Z_{ik}\right\}^{\otimes r}, \\ S_k^{(3)}(\eta;\gamma,t) &= \sum_{i=1}^{n_k} Y_{ik}(t)e^{g(\eta;X_k)\gamma^T Z_{ik}}g_2(\eta;X_k)\gamma^T Z_{ik}, \\ S_k^{(4)}(\eta;\gamma,t) &= \sum_{i=1}^{n_k} Y_{ik}(t)e^{g(\eta;X_k)\gamma^T Z_{ik}}g_1(\eta;X_k)Z_{ik}^T, \\ S_k^{(5)}(\eta;\gamma,t) &= \sum_{i=1}^{n_k} Y_{ik}(t)e^{g(\eta;X_k)\gamma^T Z_{ik}}g(\eta;X_k)Z_{ik}^T, \\ S_k^{(6)}(\eta;\gamma,t) &= \sum_{i=1}^{n_k} Y_{ik}(t)e^{g(\eta;X_k)\gamma^T Z_{ik}}g_1(\eta;X_k)\gamma^T Z_{ik}^{\otimes 2}g(\eta;X_k), \\ E_k^1(\eta;\gamma,t) &= S_k^{(1)}(\eta;\gamma,t)/S_k^{(0)}(\eta;\gamma,t), \quad E_k^3(\eta;\gamma,t) = S_k^{(3)}(\eta;\gamma,t)/S_k^{(0)}(\eta;\gamma,t), \\ E_k^4(\eta;\gamma,t) &= S_k^{(4)}(\eta;\gamma,t)/S_k^{(0)}(\eta;\gamma,t), \quad E_k^5(\eta;\gamma,t) = S_k^{(5)}(\eta;\gamma,t)/S_k^{(0)}(\eta;\gamma,t), \end{aligned}$$

$$
\begin{aligned}
V_k^1(\eta;\gamma,t) &= S_k^{(2)}(\eta;\gamma,t)/S_k^{(0)}(\eta;\gamma,t) - \left\{E_k^1(\eta;\gamma,t)\right\}^{\otimes 2}, \\
V_k^2(\eta;\gamma,t) &= S_k^{(6)}(\eta;\gamma,t)/S_k^{(0)}(\eta;\gamma,t) - E_k^1(\eta;\gamma,t)E_k^5(\eta;\gamma,t), \\
g_1(\eta;X) &= \partial g(\eta;X)/\partial\eta, \quad g_2(\eta;X) = \partial g_1(\eta;X)/\partial\eta^T.
\end{aligned}
$$

We first estimate γ from model (2.2) by maximizing the pseudo partial log-likelihood

$$\ell_1(\gamma) = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau \left\{\gamma^T Z_{ik} - \log S^{(0)}(\gamma,t)\right\} dN_{ik}(t)$$

under the working independence assumption (Wei et al., 1989). The corresponding estimating equation can be written as

$$U_1(\gamma) = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau \left\{Z_{ik} - E(\gamma,t)\right\} dN_{ik}(t).$$

Given an estimator $\hat{\gamma}$ of γ from model (2.2), we estimate η from model (2.5) by maximizing the pseudo-stratified partial log-likelihood

$$\ell_2(\eta;\hat{\gamma}) = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau \left\{g(\eta;X_k)\hat{\gamma}^T Z_{ik} - \log S_k^{(0)}(\eta;\hat{\gamma},t)\right\} dN_{ik}(t),$$

with corresponding score function,

$$U_2(\eta;\hat{\gamma}) = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau \left\{g_1(\eta;X_k)\hat{\gamma}^T Z_{ik} - E_k^1(\eta;\hat{\gamma},t)\right\} dN_{ik}(t).$$

Solving $U_2(\eta;\hat{\gamma}) = 0$, we can obtain the estimator $\hat{\eta}$ for η.

2.3 Asymptotic Properties

Denote γ_0 and η_0 as the true values of the parameters γ and η respectively. In this section, we emphasize the large sample results for $\hat{\eta}$. We begin by re-stating a previously derived result. We list the assumed conditions, state a previously derived result and then state the theorems for our estimators. Proofs are provided in the Appendix.

The following conditions are assumed throughout this chapter, where for all $k = 1, \ldots, K$ and some constant $\tau > 0$:

(a) $(D_k, C_k, Z_k, X_k, n_k, W_k)$ are independent and identically distributed;

(b) $P\{Y_{ik}(\tau) = 1\} > 0$ for $i = 1, \ldots, n_k$;

(c) $|Z_{ikl}| < B_Z < \infty$ and $|X_{kj}| < B_X < \infty$ for all $l = 1, \ldots, p$ and $j = 1, \ldots, q$ and some constants B_Z and B_X;

(d) $g(\cdot)$ is twice continuously differentiable with respect to η.

(e) γ_0 and η_0 are interior to the parameter space.

(f) The following matrices are positive definite,

$$A_1 = \mathcal{E}\left\{\int_0^\tau V(\gamma_0, t) S^{(0)}(\gamma_0, t) dH_0(t)\right\},$$
$$A_2 = \mathcal{E}\left\{\int_0^\tau V_k^1(\eta_0; \gamma_0, t) S_k^{(0)}(\eta_0; \gamma_0, t) f_k(t) d\Lambda_{0k}(t)\right\}.$$

Large sample results for $\hat{\gamma}$ have been provided by Lee et al. (1992), who showed that $K^{1/2}(\hat{\gamma} - \gamma_0)$ is asymptotically mean zero normal with variance $\Sigma_1 = A_1^{-1} B_1 A_1^{-1}$, where A_1 and B_1 can be consistently estimated by $\hat{A}_1 = K^{-1}\hat{I}$ and $\hat{B}_1 = K^{-1}\sum_{k=1}^K \hat{\psi}_k^{\otimes 2}$, with

$$\hat{I} = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^\tau V(\hat{\gamma}, t) dN_{ik}(t),$$
$$\hat{\psi}_k = \sum_{i=1}^{n_k} \int_0^\tau \{Z_{ik} - E(\hat{\gamma}, t)\} \left\{dN_{ik} - Y_{ik}(t) e^{\hat{\gamma}^T Z_{ik}} d\hat{H}_0(t)\right\},$$

where

$$\hat{H}_0(t) = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^t dN_{ik}(u) / S^{(0)}(\hat{\gamma}, u).$$

Theorem II.1. *Under conditions* $(a)-(f)$, $\hat{\eta}$ *is unique and converges almost surely to* η_0 *as* $K \to \infty$.

The proof of the consistency of $\hat{\eta}$ is similar to that of Prentice and Self (1983) and Lemma 3.1 in Andersen and Gill (1982) and is shown in Appendix (2.7.1).

Theorem II.2. *Under conditions* $(a)-(f)$, *the random vector* $K^{1/2}(\hat{\eta}-\eta_0)$ *converges weakly to a* $(q+1)$*-variate normal vector with mean 0 and covariance matrix*

$$\Sigma_2 = A_2^{-1}(A_2 + B_2\Sigma_1 B_2^T - 2CB_2^T)A_2^{-1},$$

where A_2 *is defined in condition* (f) *and*

$$B_2 = \mathcal{E}\left\{\int_0^\tau V_k^2(\eta_0;\gamma_0,t)S_k^{(0)}(\eta_0;\gamma_0,t)f_k(t)d\Lambda_{0k}(t)\right\},$$
$$C = \mathcal{E}\left\{u_k\psi_k^T\right\}A_1^{-1},$$

with

$$u_k = \sum_{i=1}^{n_k}\int_0^\tau \left\{g_1(\eta_0;X_k)\gamma_0^T Z_{ik} - E_k^1(\eta_0;\gamma_0,t)\right\}dN_{ik}(t),$$
$$\psi_k = \sum_{i=1}^{n_k}\int_0^\tau \left\{Z_{ik} - e(\gamma_0,t)\right\}\left\{dN_{ik} - Y_{ik}(t)e^{\gamma_0^T Z_{ik}}dH_0(t)\right\},$$

and

$$e(\gamma,t) = \frac{\mathcal{E}\left\{S^{(1)}(\gamma,t)\right\}}{\mathcal{E}\left\{S^{(0)}(\gamma,t)\right\}}.$$

Theorem II.2 is proved in Appendix 2.7.2. Using the proof of Theorems II.1 and II.2, together with the results from Lee et al. (1992), we can show that Σ_2 can be consistently estimated by

$$\hat{\Sigma}_2 = \hat{A}_2^{-1}(\hat{A}_2 + \hat{B}_2\hat{\Sigma}_1\hat{B}_2^T - 2\hat{C}\hat{B}_2^T)\hat{A}_2^{-1}$$

with

$$\begin{aligned}
\hat{A}_2 &= -K^{-1}\partial U_2(\eta;\gamma)/\partial\eta^T|_{\eta=\hat{\eta},\gamma=\hat{\gamma}} \\
&= K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau \left\{V_k^1(\hat{\eta};\hat{\gamma},t) - g_2(\hat{\eta};X_k)\hat{\gamma}^T Z_{ik} + E_k^3(\hat{\eta};\hat{\gamma},t)\right\} dN_{ik}(t), \\
\hat{B}_2 &= -K^{-1}\partial U_2(\eta;\gamma)/\partial\gamma^T|_{\eta=\hat{\eta},\gamma=\hat{\gamma}} \\
&= K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau \left\{V_k^2(\hat{\eta};\hat{\gamma},t) - g_1(\hat{\eta};X_k)Z_{ik}^T - E_k^4(\hat{\eta};\hat{\gamma},t)\right\} dN_{ik}(t), \\
\hat{C} &= K^{-1}\sum_{k=1}^{K}\hat{u}_k\hat{\psi}_k^T\hat{A}_1^{-1},
\end{aligned}$$

where

$$\hat{u}_k = \sum_{i=1}^{n_k}\int_0^\tau \left\{g_1(\hat{\eta};X_k)\hat{\gamma}^T Z_{ik} - E_k^1(\hat{\eta};\hat{\gamma},t)\right\} dN_{ik}(t).$$

2.4 Numerical Studies

Simulation studies were conducted to assess the finite sample behavior of $\hat{\eta}$. We also compare our method to that of Fine et al. (2003) under the special case where α_k is common among clusters.

In the first simulation study, clustered failure time data were simulated from models (2.3) and (2.4) with $K = 50, 100$; $H_0(t) = t$; $\gamma = (0.5, 1)^T$; $\eta_1 = -0.5, -0.25, 0, 0.25, 0.5$; and $\eta_2 = 0.5$. Cluster sizes were simulated from a discrete uniform distribution in the following four intervals $[5, 20]$,$[21, 50]$, $[51, 100]$ and $[101, 200]$ with approximately equal number of clusters in each interval. The cluster level covariate X_k was the cluster size measured in units of 100 subjects. The positive stable frailties, W_k, were simulated following the method in Chambers, Mallows, and Stuck (1976),

$$W_k = \frac{\sin(\alpha_k W_{1k})}{\sin(W_{1k})^{1/\alpha_k}}\left[\frac{\sin\{(1-\alpha_k)W_{1k}\}}{W_{2k}}\right]^{(1-\alpha_k)/\alpha_k},$$

where W_{1k} and W_{2k} are independent, with W_{1k} following a uniform distribution $U(0,\ \pi)$ and W_{2k} following an exponential distribution with mean 1. The individual level covariate $Z_{ik} = (Z_{ikalp02}, Z_{ik2})^T$ was independently generated, with $Z_{ikalp02}$ from a Bernoulli distribution with $p = 0.5$ and Z_{ik2} from $N(0, 1)$ distribution. The censoring times were simulated from the uniform distribution, $U(0.25,\ 1)$, yielding censoring probabilities of approximately 46%. For each scenario, 1000 replicates were carried out.

The results are summarized in Table 2.4. We report bias of the sampling mean of the estimators (BIAS), the mean of the standard error estimators (ASE), empirical standard deviation of the estimators (ESD), and the 95% empirical coverage probability (CP). In the last column, we present the approximate range of α_k for the simulated data. We also present the results for $\widehat{\gamma}$. We can see that the estimator $\widehat{\eta}$ is nearly unbiased. The (ASE) is generally fairly close to the ESD and, correspondingly, 95% empirical coverage probabilities are generally close to the nominal values. As the number of clusters increases from $K = 50$ to $K = 100$, the coverage probability is generally closer to the nominal value. In addition, as the value of α_k decreases, the coverage probability becomes lower. This may partly due to the fact that, for a fixed sample size, the amount of independent information decreases as α_k decreases; i.e., smaller value of α_k corresponds to stronger association within clusters.

To assess the asymptotic normality of the regression parameter estimates, we study the quantile-quantile (Q-Q) plots of $\widehat{\eta}$ after being standardized against standard normal variable. In Figure 2.4, we show the Q-Q plots of $\widehat{\eta}_1$ and $\widehat{\eta}_2$ when $K = 100$ and $\eta_1 = -0.5$, 0 and 0.5. All six plots exhibit diagonal lines which suggests that the asymptotic normal approximation is reasonable.

In the second simulation study, we compare the proposed method (LKS) with

Table 2.1: Summary of results for the first simulation study with $\eta_2 = 0.5$, $\gamma_1 = 0.5$ and $\gamma_2 = 1$ based on 1000 replicates.

	$\hat{\eta}$						$\hat{\gamma}$					
K	Param	True	BIAS	ASE	ESD	CP	Param	BIAS	ASE	ESD	CP	Range of α_k
50	η_1	0.5	0.01	0.27	0.27	0.96	γ_1	0.01	0.06	0.06	0.92	0.63-0.82
	η_2		0.03	0.22	0.23	0.94	γ_2	0.00	0.08	0.08	0.90	
	η_1	0.25	0.02	0.24	0.24	0.96	γ_1	0.01	0.06	0.06	0.91	0.57-0.78
	η_2		0.01	0.18	0.19	0.94	γ_2	0.00	0.08	0.09	0.90	
	η_1	0	0.03	0.22	0.23	0.94	γ_1	0.01	0.06	0.07	0.92	0.51-0.73
	η_2		0.00	0.15	0.16	0.93	γ_2	0.00	0.09	0.10	0.91	
	η_1	-0.25	0.04	0.21	0.22	0.93	γ_1	0.01	0.07	0.07	0.91	0.44-0.68
	η_2		-0.01	0.12	0.13	0.92	γ_2	0.00	0.10	0.11	0.91	
	η_1	-0.5	0.07	0.20	0.22	0.91	γ_1	0.01	0.07	0.07	0.92	0.38-0.62
	η_2		-0.04	0.10	0.12	0.87	γ_2	0.01	0.11	0.11	0.91	
100	η_1	0.5	0.01	0.20	0.19	0.95	γ_1	0.01	0.04	0.04	0.94	0.63-0.82
	η_2		0.02	0.15	0.15	0.95	γ_2	0.00	0.06	0.06	0.92	
	η_1	0.25	0.02	0.17	0.18	0.95	γ_1	0.01	0.05	0.05	0.93	0.57-0.78
	η_2		0.00	0.13	0.12	0.95	γ_2	0.00	0.06	0.06	0.93	
	η_1	0	0.02	0.16	0.16	0.93	γ_1	0.01	0.05	0.05	0.93	0.51-0.73
	η_2		0.00	0.10	0.10	0.94	γ_2	0.00	0.07	0.07	0.93	
	η_1	-0.25	0.03	0.15	0.16	0.93	γ_1	0.01	0.05	0.05	0.93	0.44-0.68
	η_2		-0.02	0.09	0.09	0.92	γ_2	0.00	0.07	0.07	0.94	
	η_1	-0.5	0.04	0.14	0.15	0.91	γ_1	0.01	0.05	0.05	0.94	0.38-0.62
	η_2		-0.03	0.07	0.08	0.87	γ_2	0.00	0.08	0.08	0.93	

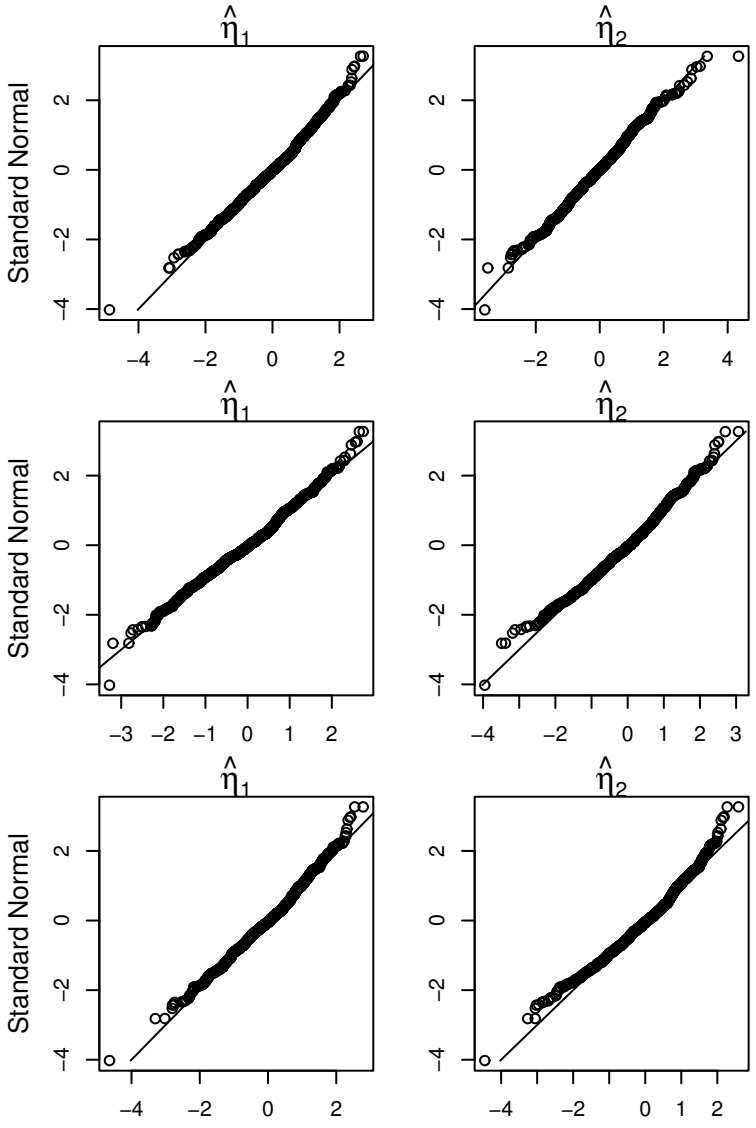

Figure 2.1: Q-Q plots for $\widehat{\eta}_1$ and $\widehat{\eta}_2$ when $K = 100$ and $\eta_1 = -0.5$, 0 and 0.5.

Table 2.2: Summary of results for the second simulation study comparing the proposed method (LKS) with FGL (see Fine et al., 2003) in the special case of constant $\alpha_k = \alpha$, $k = 1, \ldots, K$ with $\gamma_1 = 0.5$, $\gamma_2 = 1.$ and 1000 replicates.

			LKS				FGL			
K	$Param$	True	BIAS	ASE	ESD	CP	BIAS	ASE	ESD	CP
50	α	0.5	0.01	0.06	0.06	0.92	0.01	0.06	0.06	0.91
	γ_1		0.01	0.07	0.08	0.92				
	γ_2		0.01	0.11	0.12	0.91				
	α	0.75	0.01	0.06	0.06	0.91	0.01	0.06	0.06	0.87
	γ_1		0.02	0.06	0.06	0.91				
	γ_2		0.01	0.08	0.09	0.88				
100	α	0.5	0.00	0.04	0.04	0.94	0.00	0.04	0.05	0.94
	γ_1		0.01	0.05	0.05	0.94				
	γ_2		0.00	0.08	0.09	0.94				
	α	0.75	0.00	0.05	0.05	0.93	0.00	0.04	0.05	0.90
	γ_1		0.01	0.04	0.04	0.94				
	γ_2		0.00	0.06	0.06	0.92				

Fine et al. (2003) (FGL) when α_k is fixed for all clusters. We keep the same setting for H_0, γ and K. The individual level covariates and the censoring variable follow the same distribution as the first study. We fix $\alpha_k = 0.5$ or $\alpha_k = 0.75$ for all clusters. When using our method, we let $\eta_2 = 0$ and estimate η_1 only. For the FGL method, α is estimated by averaging the truncated ratio of the marginal and conditional regression parameter estimators. The results are displayed in Table 2.4, In order to facilitate the comparison, we show the results for $\widehat{\alpha}$ rather than $\widehat{\eta}$.

Both methods give an almost unbiased estimator for α, and the estimated standard error and coverage probability are reasonable. Similar to the results in Table 2.4, when the number of clusters increases from 50 to 100, the asymptotic standard errors of the estimators decrease and the coverage probability tends to be closer to the nominal value. The asymptotic standard error estimators from the two methods are very close. The LKS method gives somewhat better coverage probability than FGL.

Simulations have been done under covariate dependent frailty and common frailty

settings. Since there is no existing method to compare with under the covariate dependent frailty setting, we only make comparison under the common frailty setting. For this, three methods are available. Both the traditional EM method (Wang et al., 1995) and the Martinussen and Pipper (2005) method (MP) involve estimation of the frailties as missing data, which is computationally very slow when large number of deaths are observed for some clusters and dose not yield standard error easily. On the other hand, FGL does not involve the estimation of the frailties as is the case with the LKS method. Since our primary application of interest has clusters with large number of observed deaths, we have compared our method to FGL only.

2.5 Application

We applied the proposed methods to data on deceased donor kidney transplants performed between 2000 and 2004 in the United States. Data were obtained from the SRTR. Failure time (recorded in days) was defined as the time from transplantation to graft failure, retransplantation or death, whichever occurred first. There were 224 facilities and a total of 23,027 transplants included in the study. The facility size varied from 1 to 708 patients. We fitted the proposed covariate dependent frailty model to the data with the logit link function for the dependence parameter α_k. A total of 12 patient level covariates and 4 cluster level covariates are considered in the proportional hazards model. The same cluster level covariates are included in the link function for α_k. Patient level covariates included age at transplantation (by decade), race (African-American, Other), gender, time on dialysis (2 dummy variables), body mass index (BMI; 3 dummy variables) and primary cause of renal disease (4 dummy variables). Cluster level covariates included percentage of female patients, percentage of African-American patients, percentage of patients caused by

diabetes and center size (per 100 patients) in a center.

We expect that any covariate that is associated with the between-cluster variability may also be related to within-cluster variation. Moreover, it is easier to interpret a covariate's effect on the frailty variance after adjusting for its effect on the hazard function itself. Therefore, as a modeling strategy, covariates included in the logit link function should also be represented in the marginal hazards model. Naturally, such cluster-level covariates will not be used in the second stage of the estimation procedures, due to the stratification.

Results of our analysis are shown in Table 2.5. Percentage of female patients has a significant effect (p=0.0063) on the frailty parameter. It is found that facilities with fewer female patients tend to have a smaller value of α_k, which corresponds to greater heterogeneity in facility performance. The percentage of female patients also influences the hazards significantly. Upon examining the point estimates, one could interpret these results as being in the same direction, as higher percent female implies lower graft failure hazard and lower variation; both desirable outcomes.

2.6 Discussion

Covariate dependent frailty models for clustered failure time data have rarely been studied previously. Wassell and Moeschberger (1993) proposed a bivariate survival model with the gamma frailty parameter depending on a pair-wise covariate. Their approach only considered paired survival times in each cluster and cannot be applied to studies with larger cluster sizes. Wassell, Kulczycki, and Moyer (1995) also pointed out the increasing complexity of the application of a frailty model to clustered failure time data with larger group sizes. The model proposed in this chapter enables one to adjust for covariate effects on the frailty distribution and permits

Table 2.3: Analysis of SRTR kidney transplant data

Covariates	Estimates	SE	P-value
γ (Patient Level)			
Age (in decades)	0.1541	0.0104	$< .0001$
African-American	0.2738	0.0293	$< .0001$
Female	-0.0957	0.0254	0.0002
Time on Dialysis(in years)			
≤ 1	-0.1379	0.0372	0.0002
> 3	0.1153	0.0277	$< .0001$
Recipient BMI			
< 20	0.0732	0.0502	0.1450
$[25, 30)$	0.0391	0.0298	0.1904
$>= 30$	0.1369	0.0320	$< .0001$
Cause of ESRD			
Diabetes	0.2970	0.0359	$< .0001$
Hypertension	0.1646	0.0375	$< .0001$
Polycystic	-0.3106	0.0571	$< .0001$
Other	0.1156	0.0392	0.0032
γ (Cluster Level)			
Percent of Female (pct)	-0.0063	0.0022	0.0035
Percent of African-American (pct)	0.0039	0.0007	$< .0001$
Percent of Diabetes (pct)	0.0084	0.0017	$< .0001$
Center Size (in 100 pts)	0.0097	0.0068	0.1548
η			
Intercept	-2.3192	2.0945	0.2682
Percent of Female (pct)	0.1046	0.0382	0.0063
Percent of African-American (pct)	0.0389	0.0325	0.2316
Percent of Diabetes (pct)	0.0298	0.0531	0.5745
Center Size (in 100 pts)	-0.2288	0.2705	0.3977

both marginal and conditional inference for clustered failure time data regardless of the group size. Further consideration of the proposed method reveals two additional advantages. First, model (2.5), on which we make inference, allows for covariate-by-cluster interaction. The covariate effect is multiplicatively influenced by clusters through the cluster level covariate-dependent frailty parameter α_k. Second, with the rapid development of various methods for frailty models, researchers have begun to consider more carefully issues of ease of implementation and computation time (e.g. Fine et al., 2003; Liu and Huang, 2007). The proposed method performs well in both aspects. The method can be implemented using SAS IML. When we evaluated the computation time in the simulation study, it took approximately 4 hours for 1000 runs, with approximately 1/3 of the time spent on the PROC PHREG call.

Recalling that $\Lambda_{0k}(t) = H_0(t)^{\alpha_k^{-1}}$, we can estimate Λ_{0k} with $\widehat{H}_0(t)^{g(\widehat{\eta};X_k)}$, $k = 1, \ldots, K$, where the estimator $\widehat{H}_0(t)$ of $H_0(t)$ can be estimated from model (2.2) (see Spiekerman and Lin, 1998). Since the joint distribution of $\widehat{H}_0(t)$ and $\widehat{\eta}$ is complicated, we have not been able to obtain the asymptotic distribution of the the Λ_{0k}'s.

We noted that when a cluster level covariate is included in the conditional proportional hazard model, its effect is nearly nonidentifiable and does not interfere the estimation of other covariate effects. This is due to the use of the stratified partial likelihood approach in the estimation. Since the motivation of the proposed method is to model cluster level covariate effects on between-cluster heterogeneity and within-cluster association, the inclusion of a cluster level covariate in the conditional hazard is not needed. On the other hand, one is able to obtain the marginal effect of a cluster level covariate due to the proportional hazard in the marginal model.

For ease of computation and to avoid the estimation of the $f_k(t)$ (which is difficult for studies with large clusters), we first attempted using a stratified partial likelihood

approach based on model (2.5) only. We found that this approach does not lead to useful estimators for the parameter η_1. As an alternative, we estimate γ from model (2.2), then use the estimator $\widehat{\gamma}$ in model (2.5) to obtain a consistent estimator for η. The proposed estimation procedure is actually a two-step procedure. Such approach has been employed previously in the context of maximum likelihood by (e.g. Gong and Samaniego, 1981) and for the Clayton-Oakes model with a proportional hazards model for the margins by Glidden (2000). It should be noted that some efficiency is lost under the stratified partial likelihood approach in the second stage, as exemplified by the fact that the same estimation would be obtained if we let $f_k(t) = 1$.

Several areas of future research are possible. The proposed method relies on the specification of a link function, and model checking on this function is of potential interest. Future research on this method may also include the extension to other frailty distributions.

2.7 Appendix

2.7.1 Proof of Theorem II.1

The individual counting process martingale for the observed filtration is

$$M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(s) f_k(s) e^{g(\eta_0; X_k)\gamma_0^T Z_{ik}} d\Lambda_{0k}(s).$$

The proof of the consistency of $\widehat{\eta}$ considers the following two processes,

$$\begin{aligned} G(\eta, \widehat{\gamma}) &= K^{-1} \left\{ l_2(\eta, \widehat{\gamma}, t) - l_2(\eta_0, \gamma_0, t) \right\} \\ &= K^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^\tau \left[\left\{ g(\eta; X_k)\widehat{\gamma}^T - g(\eta_0; X_k)\gamma_0^T \right\} Z_{ik} - \log \frac{S_k^{(0)}(\eta, \widehat{\gamma}, t)}{S_k^{(0)}(\eta_0, \gamma_0, t)} \right] dN_{ik}(t), \end{aligned}$$

and

$$\begin{aligned} \Xi(\eta) &= K^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^\tau \left[\left\{ g(\eta; X_k)\gamma_0^T - g(\eta_0; X_k)\gamma_0^T \right\} Z_{ik} - \log \frac{S_k^{(0)}(\eta, \gamma_0, t)}{S_k^{(0)}(\eta_0, \gamma_0, t)} \right] \\ &\qquad Y_{ik}(t) f_k(t) e^{g(\eta_0; X_k)\gamma_0^T Z_{ik}} d\Lambda_{0k}(t). \end{aligned}$$

The difference between them can be decomposed into two parts,

$$
\begin{aligned}
& G(\eta, \widehat{\gamma}) - \Xi(\eta) \\
&= \{G(\eta, \widehat{\gamma}) - G(\eta, \gamma_0)\} + \{G(\eta, \gamma_0) - \Xi(\eta)\} \\
&= K^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^{\tau} \left\{ g(\eta; X_k) Z_{ik} (\widehat{\gamma} - \gamma_0) - \log \frac{S_k^{(0)}(\eta, \widehat{\gamma}, t)}{S_k^{(0)}(\eta, \gamma_0, t)} \right\} dN_{ik}(t) + \\
& \quad K^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^{\tau} \left[\{g(\eta; X_k) - g(\eta_0; X_k)\} \gamma_0^T Z_{ik} - \log \frac{S_k^{(0)}(\eta, \gamma_0, t)}{S_k^{(0)}(\eta_0, \gamma_0, t)} \right] dM_{ik}(t).
\end{aligned}
$$

For each η, the first term on the right-hand side of the equation converges almost surely to zero due to the consistency of $\widehat{\gamma}$ and under conditions (a)-(f) in Section 2.3, the second term is a summation of K independent and identical distributed zero mean random variables. By the Strong Law of Large Numbers (SLLN), as $K \to \infty$, $G(\eta, \widehat{\gamma})$ converges almost surely to the same limiting function of η as $\Xi(\eta)$.

By the boundness conditions (d)-(f) in Section 2.3, we can evaluate the first and the second derivatives of this limiting function by taking the partial derivatives inside the integral of $\Xi(\eta)$. The first derivative is thus

$$
\mathcal{E} \left[\sum_{i=1}^{n_k} \int_0^{\tau} \left\{ g_1(\eta; X_k) \gamma_0^T Z_{ik} - E_k^1(\eta; \gamma_0, t) \right\} Y_{ik}(t) f_k(t) e^{g(\eta_0; X_k) \gamma_0^T Z_{ik}} d\Lambda_{0k}(t) \right].
$$

It is 0 at $\eta = \eta_0$. The second derivative

$$
\mathcal{E} \left[- \sum_{i=1}^{n_k} \int_0^{\tau} V_k^1(\eta; \gamma_0, t) S_k^{(0)}(\eta; \gamma_0, t) f_k(t) d\Lambda_{0k}(t) \right]
$$

is minus a positive definite matrix at $\eta = \eta_0$ by condition (f). Therefore, $G(\eta, \widehat{\gamma})$ converges almost surely to a concave function of η with a unique maximum at $\eta = \eta_0$. Since $\widehat{\eta}$ maximizes $G(\eta, \widehat{\gamma})$, it follows that $\widehat{\eta} \stackrel{a.s.}{\to} \eta_0$ as $K \to \infty$.

2.7.2 Proof of Theorem II.2

The first order Taylor series expansion of $K^{-1/2}U_2(\widehat{\eta},\widehat{\gamma})$ about $\eta=\eta_0$ and $\gamma=\gamma_0$ gives

$$K^{-1/2}U_2(\widehat{\eta};\widehat{\gamma}) = K^{-1/2}U_2(\eta_0;\gamma_0) - \widehat{B}_2(\eta_0;\gamma^*)K^{1/2}(\widehat{\gamma}-\gamma_0) - \widehat{A}_2(\eta^*;\widehat{\gamma})K^{1/2}(\widehat{\eta}-\eta_0),$$

where η^* is on the line segment between $\widehat{\eta}$ and η_0 and γ^* is on the line segment between $\widehat{\gamma}$ and γ_0. Thus, we have

$$K^{1/2}(\widehat{\eta}-\eta_0) = \widehat{A}_2^{-1}(\eta^*;\widehat{\gamma})\left\{K^{-1/2}U_2(\eta_0;\gamma_0) - \widehat{B}_2(\eta_0;\gamma^*)K^{1/2}(\widehat{\gamma}-\gamma_0)\right\}.$$

With the consistency of $\widehat{\eta}$ and $\widehat{\gamma}$ and the SLLN, we can show that $\widehat{A}_2(\eta^*;\widehat{\gamma}) \xrightarrow{p} A_2$, and $\widehat{B}_2(\eta_0;\gamma^*) \xrightarrow{p} B_2$ and that A_2 and B_2 can be consistently estimated by $\widehat{A}_2$ and $\widehat{B}_2$ respectively.

It has been noted in Section 2.3 that $K^{1/2}(\widehat{\gamma}-\gamma_0)$ converges in distribution to $N(0,\Sigma_1)$. We will prove that $K^{-1/2}U_2(\eta_0;\gamma_0)$ converges in distribution to $N(0,A_2)$. It can be easily seen that the process $K^{-1/2}U_2(\eta_0;\gamma_0,t)$ can be written as a sum of orthogonal martingales,

$$K^{-1/2}U_2(\eta_0;\gamma_0,t) = K^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^t \left\{g_1(\eta_0;X_k)\gamma_0^T Z_{ik} - E_k^1(\eta_0;\gamma_0,s)\right\} dM_{ik}(s),$$

with predictable variation process

$$\begin{aligned}\left\langle K^{-1/2}U_2(\eta_0;\gamma_0)\right\rangle(t) &= K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^t \left\{g_1(\eta_0;X_k)\gamma_0^T Z_{ik} - E_k^1(\eta_0;\gamma_0,s)\right\}^{\otimes 2} \\ &\qquad \times Y_{ik}(s)f_k(s)e^{g(\eta_0;X_k)\gamma_0^T Z_{ik}(s)}d\Lambda_{0k}(s) \\ &= K^{-1}\sum_{k=1}^{K}\int_0^t V_k^1(\eta_0;\gamma_0,s)S_k^{(0)}(\eta_0;\gamma_0,s)f_k(s)d\Lambda_{0k}(s).\end{aligned}$$

From Rebolledo's Theorem, the Weak Law of Large Numbers (WLLN) and condition (f), we can easily show that $K^{-1/2}U_2(\eta_0;\gamma_0,\tau)$ converges in distribution to a zero

mean Gaussian vector with covariance matrix

$$\lim_{K\to\infty} \left\langle K^{-1/2} U_2(\eta_0;\gamma_0) \right\rangle (\tau) = A_2.$$

Finally, we need to obtain the asymptotic covariance matrix of $K^{-1/2}U_2(\eta_0;\gamma_0)$ and $K^{1/2}(\widehat{\gamma}-\gamma_0)$. We can see that both items can be written as a summation of K iid zero mean random vectors,

$$\begin{aligned} K^{-1/2}U_2(\eta_0;\gamma_0) &= K^{-1/2}\sum_{k=1}^{K} u_k, \\ K^{1/2}(\widehat{\gamma}-\gamma_0) &= K^{-1/2}\widehat{A}_1(\gamma^*)^{-1}\sum_{k=1}^{K}\psi_k + o_p(1), \end{aligned}$$

with

$$u_k(\eta_0;\gamma_0) = \sum_{i=1}^{n_k}\int_0^\tau \left\{ g_1(\eta_0;X_k)\gamma_0^T Z_{ik} - E_k^1(\eta_0;\gamma_0,t) \right\} dN_{ik}(t),$$

and

$$\psi_k(\gamma_0,H_0) = \sum_{i=1}^{n_k}\int_0^\tau \left\{ Z_{ik} - e(\gamma_0,t) \right\} \left\{ dN_{ik} - Y_{ik}(t)e^{\gamma_0^T Z_{ik}} dH_0(t) \right\}.$$

With the consistency of $\widehat{\gamma}$ and the WLLN, the asymptotic covariance matrix of $K^{-1/2}U_2(\eta_0;\gamma_0)$ and $K^{1/2}(\widehat{\gamma}-\gamma_0)$ is $C = \mathcal{E}\left\{u_k\psi_k^T\right\}A_1^{-1}$.

In summary, $K^{1/2}(\widehat{\eta}-\eta_0)$ converges in distribution to a $N(0,\Sigma_2)$, where

$$\Sigma_2 = A_2^{-1}(A_2 + B_2\Sigma_1 B_2^T - 2CB_2^T)A_2^{-1},$$

which can be consistently estimated by replacing each quantity with its corresponding estimator.

CHAPTER III

Computationally Efficient Marginal Model for Clustered Recurrent Event Data

3.1 Introduction

Hospitalizations are generally very costly events. For example, hospital stays represent over one third of total Medicare expenditures for dialysis patients (U.S. Renal Data System, 2006). Quantifying the impact of patient characteristics on the frequency and duration of hospitalization is an essential step towards the controlling of escalating medical costs, and can play an important role in providing cost-effective health care. In addition, assessment of dialysis facility outcomes in terms of hospitalization and comparison with outcomes at the national level can help to enhance a facility's understanding of its quality of care and how it relates to other facilities. Therefore, statistical modeling of hospitalization is needed to estimate and compare hospitalization rates. Since dialysis patients may have multiple hospital admissions, both hospital admissions (reflecting incidence) and hospital days (reflecting prevalence) can be considered as recurrent event data. Moreover, clustering is introduced both through the dependence among patients in the same facility and the correlation of outcomes over time for a given patient.

Many statistical methods have been proposed for recurrent event data (e.g. An-

dersen and Gill, 1982; Lawless and Nadeau, 1995; Lin, Wei, Yang, and Ying, 2000). The semiparametric proportional rates model of Lin et al. (2000) is widely used due to the ease of its implementation with standard statistical software such as SAS and R. The method has been extended to accommodate clustered recurrent event data. For example, Schaubel and Cai (2005) proposed two extensions applicable to clustered recurrent event data. The first assumes a cluster-specific baseline rate function, while the second assumes a common baseline rate function. It should be noted that each of the aforementioned methods requires the observation of each event occurrence time. Each is rank-based and, as such, uses the exact event times to order the failure and censoring time to construct the risk sets (and related summations) appropriately.

The analysis that motivated our current work considers the hospitalization experience among U.S. dialysis patients using both national end-stage renal disease (ESRD) registry data and that obtained from the Centers for Medicare and Medicaid Services (CMS). The pertinent analysis file is extremely large since there are over 5,000 dialysis facilities in the U.S. and more than 500,000 end-stage renal disease (ESRD) patients receiving dialysis treatment each year. Each dialysis patient may have multiple hospital admissions every year; on average, patients have 1.25 admissions with an average stay of 8 days for each admission. It has been known for some time that standard Cox regression software (e.g., R's *coxph(·)*, SAS's *PROC PHREG*) can be used to fit the proportional rates model of Lin et al. (2000). Specifically, each patient's follow-up is represented by a set of records, one per recurrent event (plus one for the final censoring event censoring). For example, the experience of a patient with events at 4, 7, 9 and censored at time 12 would be represented by 4 records: (0,4], (4,7], (7,9] and (9,12]; the event indicator would equal 1 for the first 3

records and 0 for the last record. This has come to be known as the 'counting process' style data structure; e.g., as described in Allison (2010). In our motivating example, days hospitalized is of interest, as opposed to hospital admissions. If one uses the left-truncated data structure just described, it is clear that even moderate-sized data sets can become unduly large if subjects are hospitalized frequently or tend to have long stays. For example, a hospitalization at time 7 with duration 8 days would result in 9 separate records: (6,7], (7,8] , ..., (13,14]; the event indicator set to 1 for each. Therefore, our use of the U.S. national ESRD and CMS databases will introduce computational difficulties. In settings such as these, the use of a piecewise-constant recurrent event rate model allows for the grouping of the recurrent event data, which leads to a flexible event rate model and a resulting data reduction which ameliorates the computational burden.

The proposed methods involve grouping recurrent events into intervals corresponding to the "pieces" implied by the assumed piece-wise constant baseline rate function. With respect to the interval-grouped event data, several authors have investigated nonparametric methods in estimating the mean and rate functions (e.g., Thall and Lachin, 1988; Sun and Kalbfleisch, 1995; Wellner and Zhang, 2000)). However, such methods do not consider covariate effects. Lawless and Zhan (1998) proposed a proportional rates model with a piecewise-constant baseline rate for interval-grouped recurrent event data. The authors developed robust estimation techniques based on generalized estimating equations, without assumptions on the event process. However, such methods assume a common baseline rate function and may not be easily extended to the situation where the baseline is cluster-specific, especially when the number of clusters is relatively large. Sun and Wei (2000) proposed semiparametric methods for the analysis of panel count data under informative observation and cen-

soring times. Such methods are also applicable to a proportional rates model with cluster-specific baseline rates. However, the authors emphasized regression parameter estimation only and did not consider estimation of baseline rate function. Most of the existing methods dealing with interval-grouped recurrent event data cannot be easily carried out using standard software. Cook and Lawless (2007) described a Poisson model with piecewise-constant rate functions for recurrent event data and illustrated the use of Poisson log-linear regression software for parameter estimation. This method assumes independent counting process increments given the covariates and, similar to Lawless and Zhan (1998), is not applicable to cluster-specific baseline rate function settings with relatively large number of clusters.

Another characteristic of the hospitalization data, common to many other recurrent event data settings, is the presence of a terminal event i.e., an event which stops the recurrent event process (e.g. death). Models for the rate function of recurrent event data in the presence of a terminal event can generally be categorized as (1) marginal model (e.g. Ghosh and Lin, 2002; Schaubel and Zhang, 2010) which can be interpreted as the occurrence rate averaging over mortality experience, (2) partial marginal model (e.g. Cook and Lawless, 1997; Ye, Kalbfleisch, and Schaubel, 2007) which models the rate of the recurrent events among survivors. In this chapter, we consider the partial marginal model for the rate function of the recurrent event with unspecified dependence structure between the recurrent events and the terminal event.

The remainder of this chapter is organized as follows. In Section 3.2, we first propose the proportional rates model with piecewise-constant baseline rate function for clustered recurrent event data, in the absence of a terminal event. The dependence structures for within-patient events are left completely unspecified. The extension to

the setting with a terminating event then follows, under a partial-marginal model. The essential parts of the estimation procedure are quite similar, although the interpretation of the covariate effects is different. The proposed estimation procedure requires only the interval-specific event and person-time totals, instead of the exact recurrent event times, which leads to considerable data reduction and hence reduced computing time. In Section 3.3, we compare the proposed estimation method to a joint estimating equation method based on pseudo likelihood. We derive the large-sample properties of the proposed estimators in Section 3.4 and assess their small-sample performance in Section 3.5 under various data configurations, including settings in which the model is misspecified. In Section 3.6, we apply the proposed the model to the study of days hospitalized among U.S. dialysis patients. The chapter then concludes with some discussion in Section 3.7.

3.2 Model Specification and Estimation

3.2.1 In the Absence of a Terminal Event

As the name implies, the proposed model assumes that the baseline rate is constant over pre-specified intervals and is applied to recurrent event data in the absence of terminal event. Denote the largest observation time by τ. Let $a_0 < a_1 < \ldots < a_L$ denote the cut points for the L intervals on $[0, \tau]$, where $a_0 = 0$, $a_L = \tau$ and $\Omega_\ell = (a_{\ell-1}, a_\ell]$ for $\ell = 1, \ldots, L$. Let k index cluster, with cluster sizes $n_1, \ldots, n_K$ and let i index the subject $(i = 1, \ldots, n)$ with $n = \sum_{k=1}^{K} n_k$. For subject i, let G_i denote cluster. Let C_i denote the right censoring time for subject i. Since data are often left-truncated, we explicitly allow for left-truncation in the formulation of the proposed methods. The left-truncation time is represented by B_i. We then define the at-risk process by $\widetilde{Y}_i(t) = I(B_i \leq t \leq C_i)$ with $I(\cdot)$ being the indicator function. Let $\widetilde{N}_i^*(t)$

denote the cumulative number of events up to time t and let $\widetilde{N}_i(t) = \int_0^t \widetilde{Y}_i(s) d\widetilde{N}_i^*(s)$ denote the observed number of events. We then specify the rate function for subject i from cluster k as

$$E\{d\widetilde{N}_i^*(t)|Z_i(t), G_i = k\} = \rho_{k\ell} e^{\gamma_0^T Z_{i\ell}} dt,$$

where, for $\ell = 1, \ldots, L$, $\rho_{k\ell}$ is the baseline rate function for the kth cluster, γ_0 is a p-vector parameter, $Z_{i\ell} = Z_i(t)$, $t \in \Omega_\ell$ is a p-vector possibly time-varying covariates for subject i. Note that any time-dependent covariates are assumed to be external (Kalbfleisch and Prentice, 2002) and constant within each interval Ω_ℓ. Define $G_{ik} = I(G_i = k)$, $\widetilde{Y}_{ik}(t) = G_{ik}\widetilde{Y}_i(t)$, $d\widetilde{N}_{ik}^*(t) = G_{ik} d\widetilde{N}_i^*(t)$ and $d\widetilde{N}_{ik}(t) = G_{ik} d\widetilde{N}_i(t)$. Under the assumption of independent left truncation and censoring, which can be specified as

$$E\{d\widetilde{N}_{ik}^*(t)|Z_i(t), G_i = k, \widetilde{Y}_i(t) = 1\} = E\{d\widetilde{N}_{ik}^*(t)|Z_i(t), G_i = k\},$$

we have

$$E\{d\widetilde{N}_{ik}(t)|Z_i(t), \widetilde{Y}_{ik}(t)\} = \widetilde{Y}_{ik}(t)\rho_{k\ell} e^{\gamma_0^T Z_{i\ell}} dt. \tag{3.1}$$

3.2.2 Piecewise-Constant Baseline Rates Model in the Presence of a Terminating Event

When the recurrent event is potentially stopped by a terminal event (e.g. death), we can similarly specify a partial marginal model with piecewise-constant baseline rates. Let D_i denote the death time for subject i. Define the follow-up time $X_i = C_i \wedge D_i$, with $a \wedge b = \min(a, b)$ and the at risk process $Y_i(t) = I(B_i \leq t \leq X_i)$. Then the counting process for the recurrent events $N_i^*(t) = N_i^*(t \wedge D_i)$, which acknowledges the fact that death stops the further occurrence of recurrent events, such that $N_i^*(t)$ is a constant after D_i. Similar to the model in the absence of terminal event, the

occurrence rate function for subject i from cluster k conditional on being alive is given as

$$E\{dN_i^*(t)|Z_i(t), D_i \geq t, G_i = k\} = \alpha_{k\ell} e^{\beta_0^T Z_{i\ell}} dt,$$

where, for $\ell = 1, \ldots, L$, $\alpha_{k\ell}$ is the baseline rate function for the kth cluster and β_0 is a p-vector parameter. Define $Y_{ik}(t) = G_{ik} Y_i(t)$, $dN_{ik}^*(t) = G_{ik} dN_i^*(t)$ and $dN_{ik}(t) = G_{ik} dN_i(t)$. Under the assumption of independent left truncation and censoring, which is written as

$$E\{dN_{ik}^*(t)|Z_i(t), Y_i(t) = 1, G_i = k\} = E\{dN_{ik}^*(t)|Z_i(t), D_i \geq t, G_i = k\},$$

we have

$$E\{dN_{ik}(t)|Z_i(t), Y_{ik}(t)\} = Y_{ik}(t)\alpha_{k\ell} e^{\beta_0^T Z_{i\ell}} dt. \tag{3.2}$$

3.2.3 Estimation

Next, we describe the estimation method for the model in the presence of a terminal event. Similar estimating procedure can be applied to the model in the absence of terminating event by setting $D_i = \tau$. We first define some notation. For subject i (from cluster k), let $t_{ik\ell} = \int_{a_{\ell-1}}^{a_\ell} Y_{ik}(t)dt$ denote the time at risk (exposure time) and $d_{ik\ell} = \int_{a_{\ell-1}}^{a_\ell} dN_{ik}(t)$ be the observed number of events experienced in Ω_ℓ . In addition, for $r = 0, 1, 2$, $k = 1, \ldots, K$ and $\ell = 1, \ldots, L$, we define

$$\begin{aligned} S_{k\ell}^{(r)}(\beta) &= n^{-1} \sum_{i=1}^{n} Z_{i\ell}^{\otimes r} t_{ik\ell} e^{\beta^T Z_{i\ell}}, \\ \overline{Z}_{k\ell}(\beta) &= S_{k\ell}^{(1)}(\beta)/S_{k\ell}^{(0)}(\beta), \\ V_{k\ell}(\beta) &= S_{k\ell}^{(2)}(\beta)/S_{k\ell}^{(0)}(\beta) - \overline{Z}_{k\ell}(\beta)^{\otimes 2} \end{aligned}$$

where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$. We next define the compensated counting process,

$$dM_{ik}(t) = dN_{ik}(t) - Y_{ik}(t)\alpha_{k\ell} e^{\beta_0^T Z_{i\ell}} dt, \quad t \in \Omega_\ell,$$

for $\ell = 1, \ldots, L$. By the specification of the model (3.2) and under the corresponding independent left truncation and censoring assumption, $E\{dM_{ik}(t)|Z_{i\ell}, Y_{ik}(t)\} = 0$ for $t \in \Omega_\ell$. Thus, it follows that

$$\xi_{ik\ell}(\beta_0) = \int_{a_{\ell-1}}^{a_\ell} dM_{ik}(t) = d_{ik\ell} - \alpha_{k\ell} t_{ik\ell} e^{\beta_0^T Z_{i\ell}}$$

has mean zero since

$$E\Big\{\int_{a_{\ell-1}}^{a_\ell} dM_{ik}(t)\Big\} = E\Big[\int_{a_{\ell-1}}^{a_\ell} E\left\{dM_{ik}(t)|Z_{i\ell}, Y_{ik}(t)\right\}\Big] = 0.$$

We consider the estimating function,

$$U(\beta) = \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{\ell=1}^{L}\{Z_{i\ell} - \overline{Z}_{k\ell}(\beta)\}\xi_{ik\ell}(\beta),$$

motivated by the fact that $U(\beta_0)$ can be shown to have mean 0 asymptotically, which can be proved by replacing $\overline{Z}_{k\ell}(\beta)$ with the corresponding limiting values in $U(\beta)$. We can simplify $U(\beta)$ to

$$U(\beta) = \sum_{i=1}^{n}\sum_{\ell=1}^{L}\sum_{k=1}^{K}\{Z_{i\ell} - \overline{Z}_{k\ell}(\beta)\}d_{ik\ell}, \tag{3.3}$$

such that an estimator for β_0, $\widehat{\beta}$, can be obtained by solving $U(\beta) = 0$. A Breslow-Aalen type estimator for $\alpha_{k\ell}$ is then given as

$$\widehat{\alpha}_{k\ell}(\widehat{\beta}) = \frac{d_{\bullet k\ell}}{nS_{k\ell}^{(0)}(\widehat{\beta})}, \tag{3.4}$$

where $\bullet$ denotes the summation over the corresponding subscript. Therefore, the corresponding Breslow-Aalen type estimator for the cumulative baseline rate function $\mu_{0k}(t) = \sum_{\ell=1}^{L}\alpha_{k\ell}(a_\ell \wedge t - a_{\ell-1})$ is given as

$$\widehat{\mu}_{0k}(t;\widehat{\beta}) = \sum_{\ell=1}^{L}\frac{d_{\bullet k\ell}}{nS_{k\ell}^{(0)}(\widehat{\beta})}(a_\ell \wedge t - a_{\ell-1}), \tag{3.5}$$

One may notice that (3.3) is similar to the partial score equation for recurrent event data except for an offset term, and a weight term. Therefore, the proposed

estimation method is easy to implement with SAS (PROC PHREG) or R (coxph) with the censoring variable $d_{ik\ell}$, the weight term $w_{ik\ell} = \max(d_{ik\ell}, 1)$, and the offset term $\log(t_{ik\ell}) - \log(w_{ik\ell})$. It should also be noted that, unlike the conventional partial score equation in which statistics are computed at each distinct recurrent event time, the proposed estimating equation is calculated only for each interval, which greatly speeds up the calculation, especially when the number of event occurrences is large.

A few additional notes are in order. First, if the data are not left-truncated, the proposed methods can be applied by setting $B_i = 0$ for all $i = 1, \ldots, n$. Second, in the absence of terminating event the unbiasedness of $U(\beta_0)$ can be proved based on conditional expectation argument (e.g. Appendix 7.1 in Schaubel and Cai, 2005). Finally, we emphasize cluster-specific baseline rates model in this chapter. When the baseline rate function is common to all clusters, an analogous estimation procedure can be carried out with $S^{(r)}_{k\ell}$ and $d_{\bullet k\ell}$ ($k = 1, \ldots, K$, $\ell = 1, \ldots, L$) replaced by the corresponding quantities summing over all the clusters in (3.3) and (3.4).

3.3 Comparison with Joint Estimating Equation Approach

An alternative estimation approach is based on pseudo-likelihood that ignores within-subject and within-cluster dependence. Let $\alpha = (\alpha_{11}, \ldots, \alpha_{1L}, \ldots, \alpha_{K1}, \ldots, \alpha_{KL})'$ and $\theta = (\alpha', \beta')'$. The pseudo-likelihood function is then $L(\theta) = \prod_{i=1}^{n} \prod_{k=1}^{K} \prod_{\ell=1}^{L} L_{ik\ell}(\theta)$, where $L_{ik\ell}(\theta)$ is given as

$$L_{ik\ell}(\theta) = (\alpha_{k\ell} e^{\beta^T Z_{i\ell}})^{d_{ik\ell}} e^{-\alpha_{k\ell} t_{ik\ell} e^{\beta^T Z_{i\ell}}}.$$

The resulting log-likelihood is then

$$\ell(\theta) = \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{\ell=1}^{L} \left\{ d_{ik\ell} (\log \alpha_{k\ell} + \beta^T Z_{i\ell}) - \alpha_{k\ell} t_{ik\ell} e^{\beta^T Z_{i\ell}} \right\}.$$

The score vector $U^J(\theta) = \{U_\alpha^T(\theta), U_\beta^T(\theta)\}^T$, with $U_\alpha = (U_{\alpha_{11}}, \ldots, U_{\alpha_{KL}})^T$, can be obtained by taking the partial derivative of $\ell(\theta)$ with respect to θ as

$$U_{\alpha_{k\ell}}(\theta) = \frac{d_{\bullet k\ell}}{\alpha_{k\ell}} - nS_{k\ell}^{(0)}(\beta), \quad k = 1, \ldots, K, \ell = 1, \ldots, L, \tag{3.6}$$

$$U_\beta(\theta) = \sum_{i=1}^{n}\sum_{\ell=1}^{L}\sum_{k=1}^{K} Z_{i\ell}\xi_{ik\ell}(\beta). \tag{3.7}$$

The solution of the joint estimating equation $U^J(\theta) = 0$, θ, is then an estimator for θ. It can be easily seen that this joint estimating equation method (JM) gives the same estimator as the proposed method (PM). A profile estimator $\widetilde{\alpha}(\beta)$ for α can be obtained from (3.6) given β, which equals the Breslow-Aalen estimator from PM. Replacing α with $\widetilde{\alpha}(\beta)$ in (3.7) then gives the same estimating function (3.3) for β in PM . Moreover, unlike PM that calculates the estimated covariance matrix for $\widehat{\alpha}$ and $\widehat{\beta}$ separately, JM estimates the joint covariance matrix for $\widehat{\alpha}$ and $\widehat{\beta}$, which involves inverting the observed information matrix I^J. As the minus second partial derivative of $\ell(\theta)$, I^J is of dimension $(KL + p)$ with the upper left square submatrix corresponding to α being a diagonal matrix with the $\{k(L-1)+\ell\}$th diagonal element equal to $d_{\bullet k\ell}/\alpha_{k\ell}^2$. When $d_{\bullet k\ell} = 0$, which is quite possible for clusters with small cluster size or less frequent recurrent events in interval Ω_ℓ, I^J is not positive definite. As a result, JM cannot give an estimator for the joint covariance matrix.

3.4 Asymptotic Properties

The asymptotic properties are derived for the model in the presence of a terminal event. As illustrated in Section 3.2.3, in the absence of terminal event, one can obtain similar results by letting $D_i = \tau$.

For $i = 1, \ldots, n$, we impose the following regularity conditions:

(a) $\{N_i(t), Y_i(t), Z_i(t), G_i\}_{i=1}^n$ are independent and identically distributed;

(b) $P\{G_{ik} = 1\} \in (0, 1]$;

(c) $E\{Y_i(t)\} > 0$, for all $t \in (0, \tau]$;

(d) $N_i(t)$, are bounded by a constant;

(e) $Z_{i\ell}$, $\ell = 1, \ldots, L$ are bounded by a constant;

(f) Let $\mathcal{B}$ be a neighborhood of β_0. For $d = 0, 1, 2$, $s_{k\ell}^{(d)}(\beta)$ are continuous functions of $\beta \in \mathcal{B}$, where $s_{k\ell}^{(d)}(\beta)$ is the limiting values of $S_{k\ell}^{(d)}(\beta)$; $s_{k\ell}^{(1)}(\beta)$ and $s_{k\ell}^{(2)}(\beta)$ are bounded and $s_{k\ell}^{(0)}(\beta)$ is bounded away from 0 on $\mathcal{B}$ with

$$s_{k\ell}^{(1)}(\beta) = \frac{\partial}{\partial \beta} s_{k\ell}^{(0)}(\beta), \; s_{k\ell}^{(2)}(\beta) = \frac{\partial^2}{\partial \beta \partial \beta^T} s_{k\ell}^{(0)}(\beta).$$

(g) Positive-definiteness of the matrix

$$A = \lim_{n \to \infty} n^{-1} \sum_{k=1}^{K} \sum_{\ell=1}^{L} \alpha_{k\ell} v_{k\ell}(\beta_0) s_{k\ell}^{(0)}(\beta_0),$$

where $v_{k\ell}(\beta) = s_{k\ell}^{(2)}(\beta)/s_{k\ell}^{(0)}(\beta) - \bar{z}_{k\ell}(\beta)^{\otimes 2}$ and $\bar{z}_{k\ell}(\beta) = s_{k\ell}^{(1)}(\beta)/s_{k\ell}^{(0)}(\beta)$.

Assumption (a) specifies that the independent units in the proposed method are subjects. Assumption (b) states that the probability of a randomly selected subject being assigned to a cluster is nonzero for any cluster. Both conditions are necessary so that parameter estimators for the cluster-specific baseline rate functions are estimable for all clusters.

We next summarize the theoretical results for $\widehat{\beta}$ by the following theorem.

Theorem III.1. *Under regularity conditions* $(a) - (g)$, $\widehat{\beta}$ *converges almost surely to* β_0 *as* $n \to \infty$, *while* $n^{1/2}(\widehat{\beta} - \beta_0)$ *converges to a p-variate normal vector with mean 0 and covariance matrix* $\Sigma = A^{-1}BA^{-1}$, *where* $B = E\{U_1(\beta_0)^{\otimes 2}\}$, *with*

$$U_i(\beta) = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \{Z_{i\ell} - \bar{z}_{k\ell}(\beta)\} \xi_{ik\ell}(\beta).$$

A consistent estimator for Σ can be obtained as $\widehat{\Sigma} = \widehat{A}(\widehat{\beta})\widehat{B}(\widehat{\beta})\widehat{A}(\widehat{\beta})$, where

$$\begin{aligned}
\widehat{A}(\beta) &= n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{\ell=1}^{L} V_{k\ell}(\beta)d_{ik\ell}, \\
\widehat{B}(\beta) &= n^{-1}\sum_{i=1}^{n}\widehat{U}_i(\beta)^{\otimes 2}, \\
\widehat{U}_i(\beta) &= \sum_{k=1}^{K}\sum_{\ell=1}^{L}\{Z_{i\ell} - \overline{Z}_{k\ell}(\beta)\}\widehat{\xi}_{ik\ell}(\beta), \\
\widehat{\xi}_{ik\ell}(\beta) &= d_{ik\ell} - \widehat{\alpha}_{k\ell}(\beta)t_{ik\ell}e^{\beta^T Z_{ik\ell}}
\end{aligned}$$

Theorem (III.1) can be proved by combining the Uniform Strong Law of Large Numbers (USLLN) and the Central Limit Theorem (CLT), as is shown in the Appendix. We next present the essential asymptotic properties for $\widehat{\mu}_{0k}(t;\widehat{\beta})$.

Theorem III.2. *Under regularity conditions $(a)-(f)$, for $k = 1,\ldots,K$, $\widehat{\mu}_{0k}(t;\widehat{\beta})$ converges almost surely to $\mu_{0k}(t)$ uniformly in $t \in [0,\tau]$; the process $n^{1/2}\{\widehat{\mu}_{0k}(t;\widehat{\beta}) - \mu_{0k}(t)\}$ converges to a zero-mean Gaussian process with covariance function $\Psi_k(s,t) = E\{\psi_{1k}(s)\psi_{1k}(t)\}$, where*

$$\begin{aligned}
\psi_{ik}(t) &= \sum_{\ell=1}^{L}\psi_{ik\ell}(\beta_0)(a_\ell \wedge t - a_{\ell-1}), \\
\psi_{ik\ell}(\beta) &= \frac{\xi_{ik\ell}(\beta)}{s_{k\ell}^{(0)}(\beta)} - \alpha_{k\ell}\bar{z}_{k\ell}(\beta)A^{-1}U_i(\beta_0).
\end{aligned}$$

We show in Appendix that $n^{1/2}\{\widehat{\mu}_{0k}(t;\widehat{\beta}) - \mu_{0k}(t)\}$ is asymptotically equivalent to $n^{-1/2}\sum_{i=1}^{n}\psi_{ik}(t)$. A consistent estimator for $\Psi_k(s,t)$ is then $n^{-1}\sum_{i=1}^{n}\widehat{\psi}_{ik}(s)\widehat{\psi}_{ik}(t)$ with

$$\begin{aligned}
\widehat{\psi}_{ik}(t) &= \sum_{\ell=1}^{L}\widehat{\psi}_{ik\ell}(\widehat{\beta})(a_\ell \wedge t - a_{\ell-1}), \\
\widehat{\psi}_{ik\ell}(\beta) &= \frac{\widehat{\xi}_{ik\ell}(\beta)}{S_{k\ell}^{(0)}(\beta)} - \frac{d_{\bullet k\ell}\overline{Z}_{k\ell}(\beta)}{S_{k\ell}^{(0)}(\beta)}\widehat{A}^{-1}(\beta)\widehat{U}_i(\beta).
\end{aligned}$$

3.5 Simulation Study

Simulation studies were conducted to assess the performance of the estimation method in finite samples both in the absence and presence of a terminal event. We present the results in the presence of a terminal event only, since they exhibit similar patterns as those in the absence of terminal event.

In the first simulation study, for subject (i,k), we generate recurrent event times from

$$E\{dN_{ik}^*(t)|Z_i, W_i, D_i \geq t\} = W_i\lambda_{0k}(t)e^{\beta_0^T Z_i}dt,$$

where the subject-level random effect W_i follows Gamma distribution with unit mean and variance $\sigma^2 = 1$, and the cluster-specific baseline rate function $d\mu_{0k}(t) = 1$, $Z_i = \{Z_{i1}, Z_{i2}\}$ are 2-vector covariates with $Z_{i1} \sim Bernoulli(0.5)$ and $Z_{i2} \sim N(0, 0.25)$, $\beta_1 = 0.5$ and $\beta_2 = 0.25, 0.5, 0.75, 1$. In addition, we let $D_i \sim Exp(0.1 + 0.1Z_{i1})$ and $C_i \sim U(5, 10)$. The average number of recurrent events ranged from 6 to 8. We set $K = 50, 100$ and $n_k = 20, 50, 100$. For each simulated data set, we estimated β_0 under model (3.2) with three settings for the piecewise-constant baseline rate function: the first setting is with $L = 3$ pieces defined by $0, 2, 4, 10$; the second setting is with $L = 6$ pieces defined by $0, 1, \ldots, 5, 10$; the third setting is with $L = 12$ pieces resulting from adding 6 midpoints of the intervals in the second setting. The results are shown in Table 3.1 and Table 3.2 for $\widehat{\beta}_1$ and $\widehat{\beta}_2$ respectively based on 1000 simulations.

For the first simulation study where the true model is actually piecewise-constant, we do not present the results based on $L = 3$ since it gives similar results to $L = 6$ and $L = 12$. For all of the data configurations in Table 3.1, the estimator for β_1 corresponding to the binary covariate, $\widehat{\beta}_1$, is approximately unbiased with the bias reduced with increasing cluster size. The mean of the asymptotic standard error

(ASE) of $\widehat{\beta}_1$ is generally close to the empirical standard deviation (ESD) of $\widehat{\beta}_1$, and the coverage probabilities (CP) are fairly close to the nominal value. Adding more cut points does not seem to improve the performance of the estimator. Results in Table 3.2 for $\widehat{\beta}_2$, the estimator corresponding to the normal covariate, are similar to those of $\widehat{\beta}_1$ except that the CPs for the settings with smaller cluster size tend to underestimate the nominal values, but do get closer to 95% as the true value decrease.

In the second simulation study, we let $\mu_{0k}(t) = 0.5t^2$ and keep everything else the same. The average number of recurrent events per subject ranged from 18 to 20. For each setting, 1000 data sets are simulated. We then assess the proposed methods under mis-specification assuming piecewise-constant baseline rates with the same setting as the first simulation study. The results for $\widehat{\beta}_1$ are shown in Table 3.4, again based on 1000 replicates.

Under mis-specification of the baseline rate function, results for $\widehat{\beta}_1$ are similar to those in the first simulation study; except for the comparison among the three choices of cut points. In particular, it appears that $\widehat{\beta}_1$ is approximately unbiased, with the bias reduced by adding more pieces to the baseline rate function. The ASEs and ESDs are fairly close to each other under all settings examined. Adding more pieces do not seem to improve the efficiency, although the CP gets closer to the nominal level as more prices are added to the assumed baseline rate function. The improvement is more obvious comparing $L = 3$ and $L = 12$. Results for $\widehat{\beta}_2$ are very similar to those for $\widehat{\beta}_1$ and, hence, are not presented.

In the third simulation study, we assess the asymptotic properties for $\widehat{\mu}_{0k}(t)$ with $\beta_0 = (0.5, 1)$ and $K = 50$ under two scenarios; one in which $\mu_{0k}(t) = t$ and a second in which $\mu_{0k}(t) = 0.5\, t^2$. Remaining characteristics are as described previously. We let

Table 3.1: Results of $\widehat{\beta}_1$ in the first simulation study with $\beta_1 = 0.5$, $\mu_{0k}(t) = t$ and 1000 replicates.

			$L = 6$				$L = 12$			
K	n_k	β_2	BIAS	ASE	ESD	CP	BIAS	ASE	ESD	CP
50	20	1	-0.003	0.064	0.060	0.958	-0.003	0.063	0.060	0.957
		0.75	-0.004	0.061	0.059	0.962	-0.004	0.061	0.059	0.963
		0.5	-0.004	0.060	0.058	0.958	-0.004	0.060	0.058	0.952
		0.25	-0.004	0.059	0.056	0.959	-0.004	0.059	0.056	0.958
50	50	1	0.008	0.048	0.046	0.958	0.008	0.048	0.046	0.959
		0.75	0.007	0.047	0.044	0.962	0.007	0.046	0.044	0.960
		0.5	0.007	0.045	0.043	0.959	0.007	0.045	0.043	0.961
		0.25	0.007	0.045	0.042	0.962	0.007	0.045	0.042	0.960
50	100	1	0.010	0.037	0.037	0.949	0.010	0.037	0.037	0.948
		0.75	0.009	0.036	0.035	0.952	0.010	0.036	0.035	0.950
		0.5	0.009	0.035	0.033	0.953	0.009	0.035	0.033	0.954
		0.25	0.010	0.034	0.033	0.944	0.010	0.034	0.033	0.946
100	20	1	0.012	0.050	0.049	0.944	0.012	0.050	0.049	0.947
		0.75	0.010	0.049	0.047	0.949	0.010	0.049	0.047	0.950
		0.5	0.011	0.048	0.046	0.949	0.011	0.048	0.045	0.949
		0.25	0.011	0.047	0.045	0.957	0.011	0.047	0.045	0.958
100	50	1	0.010	0.036	0.036	0.938	0.010	0.036	0.036	0.935
		0.75	0.010	0.035	0.034	0.952	0.010	0.035	0.035	0.951
		0.5	0.010	0.034	0.033	0.949	0.010	0.034	0.033	0.945
		0.25	0.011	0.033	0.032	0.940	0.011	0.033	0.032	0.939
100	100	1	-0.000	0.027	0.027	0.962	-0.000	0.027	0.027	0.960
		0.75	-0.001	0.026	0.025	0.959	-0.001	0.026	0.025	0.958
		0.5	-0.001	0.025	0.024	0.966	-0.001	0.025	0.024	0.967
		0.25	-0.000	0.025	0.023	0.961	-0.000	0.025	0.023	0.961

Table 3.2: Results of $\widehat{\beta}_2$ in the first simulation study with $\beta_1 = 0.5$, $\mu_{0k}(t) = t$ and 1000 replicates.

			$\widehat{\beta}_2$							
			$L = 6$				$L = 12$			
K	n_k	β_2	BIAS	ASE	ESD	CP	BIAS	ASE	ESD	CP
50	20	1	-0.001	0.067	0.072	0.927	-0.001	0.067	0.072	0.925
		0.75	-0.001	0.063	0.067	0.930	-0.001	0.063	0.067	0.927
		0.5	-0.001	0.061	0.063	0.929	-0.001	0.061	0.063	0.929
		0.25	-0.001	0.059	0.062	0.935	-0.001	0.059	0.062	0.935
50	50	1	0.000	0.052	0.055	0.937	0.000	0.051	0.055	0.938
		0.75	0.001	0.049	0.051	0.941	0.001	0.048	0.051	0.943
		0.5	0.000	0.046	0.048	0.945	0.000	0.046	0.048	0.947
		0.25	0.001	0.045	0.046	0.940	0.001	0.045	0.046	0.941
50	100	1	0.000	0.040	0.043	0.936	0.000	0.040	0.043	0.937
		0.75	0.001	0.037	0.039	0.945	0.001	0.037	0.039	0.944
		0.5	0.000	0.036	0.036	0.950	0.000	0.035	0.036	0.950
		0.25	0.001	0.034	0.035	0.948	0.001	0.034	0.035	0.947
100	20	1	-0.000	0.053	0.059	0.921	-0.000	0.052	0.059	0.921
		0.75	0.000	0.050	0.055	0.928	0.000	0.050	0.055	0.924
		0.5	0.000	0.048	0.051	0.943	0.000	0.048	0.051	0.943
		0.25	0.001	0.047	0.050	0.934	0.001	0.047	0.050	0.934
100	50	1	0.000	0.038	0.042	0.930	0.000	0.038	0.042	0.928
		0.75	0.000	0.036	0.039	0.938	0.000	0.036	0.039	0.936
		0.5	-0.000	0.035	0.036	0.942	-0.000	0.034	0.036	0.942
		0.25	0.001	0.034	0.035	0.939	0.001	0.033	0.035	0.941
100	100	1	0.001	0.029	0.031	0.934	0.001	0.029	0.031	0.934
		0.75	0.000	0.027	0.029	0.938	0.000	0.027	0.029	0.938
		0.5	0.000	0.026	0.027	0.933	0.000	0.026	0.027	0.933
		0.25	0.000	0.025	0.026	0.930	0.000	0.025	0.026	0.930

Table 3.3: Results of the second simulation study for $\widehat{\beta}_1$ with $\beta_1 = 0.5$, $\mu_{0k}(t) = 0.5t^2$ and 1000 replicates.

			$L = 3$				$L = 6$				$L = 12$			
K	n_k	β_2	BIAS	ASE	ESD	CP	BIAS	ASE	ESD	CP	BIAS	ASE	ESD	CP
50	20	1	-0.042	0.071	0.071	0.907	-0.026	0.071	0.070	0.935	-0.018	0.071	0.070	0.946
		0.75	-0.042	0.069	0.067	0.909	-0.026	0.068	0.065	0.934	-0.018	0.068	0.067	0.952
		0.5	-0.041	0.067	0.065	0.912	-0.026	0.067	0.065	0.944	-0.017	0.066	0.065	0.947
		0.25	-0.042	0.066	0.064	0.917	-0.026	0.066	0.064	0.937	-0.018	0.065	0.064	0.945
50	50	1	-0.025	0.055	0.053	0.922	-0.010	0.054	0.052	0.957	-0.002	0.054	0.052	0.957
		0.75	-0.025	0.053	0.050	0.928	-0.010	0.052	0.050	0.959	-0.003	0.052	0.050	0.961
		0.5	-0.025	0.051	0.049	0.925	-0.010	0.051	0.048	0.956	-0.002	0.050	0.048	0.960
		0.25	-0.025	0.050	0.047	0.938	-0.010	0.050	0.047	0.962	-0.002	0.050	0.047	0.962
50	100	1	-0.022	0.042	0.042	0.907	-0.001	0.042	0.042	0.947	0.001	0.042	0.042	0.951
		0.75	-0.022	0.040	0.040	0.915	-0.006	0.040	0.040	0.951	0.002	0.040	0.039	0.948
		0.5	-0.021	0.039	0.038	0.910	-0.007	0.039	0.038	0.948	0.001	0.039	0.038	0.955
		0.25	-0.021	0.038	0.037	0.912	-0.007	0.038	0.037	0.955	0.001	0.038	0.037	0.960
100	20	1	-0.022	0.056	0.056	0.921	-0.006	0.056	0.056	0.940	0.002	0.056	0.056	0.950
		0.75	-0.022	0.054	0.054	0.928	-0.006	0.054	0.054	0.948	0.002	0.054	0.054	0.949
		0.5	-0.022	0.053	0.052	0.926	-0.006	0.053	0.052	0.955	0.002	0.053	0.052	0.949
		0.25	-0.022	0.052	0.051	0.929	-0.007	0.052	0.051	0.950	0.002	0.052	0.051	0.942
100	50	1	-0.022	0.041	0.041	0.904	-0.001	0.040	0.041	0.937	0.001	0.040	0.041	0.942
		0.75	-0.022	0.039	0.039	0.909	-0.001	0.039	0.039	0.942	0.001	0.039	0.039	0.947
		0.5	-0.022	0.038	0.037	0.907	-0.001	0.038	0.037	0.940	0.001	0.038	0.037	0.949
		0.25	-0.022	0.037	0.036	0.910	-0.001	0.037	0.036	0.950	0.001	0.037	0.036	0.957
100	100	1	-0.029	0.031	0.030	0.836	0.000	0.030	0.030	0.942	0.000	0.030	0.030	0.945
		0.75	-0.028	0.029	0.029	0.831	0.000	0.029	0.029	0.945	0.000	0.029	0.028	0.948
		0.5	-0.028	0.028	0.027	0.824	0.000	0.028	0.027	0.955	0.000	0.028	0.027	0.950
		0.25	-0.028	0.028	0.026	0.820	0.000	0.028	0.026	0.950	0.000	0.028	0.026	0.958

$n_1 = 50$, $n_2 = 100$, $n_3 = 200$, $n_4 = 500$ and $n_k = 50$ for $k = 5, \ldots, 50$. In both cases, the rates are assumed to be piece-wise constant. Specifically, we estimated $\mu_{0k}(t)$ under model (3.2) with two settings for the cut points of the piecewise-constant baseline rate function: (i) with $L = 6$ and $a_\ell = 0, 1, \ldots, 5, 10$ (ii) with $L = 12$ and $a_\ell = 0, 0.5, 1, 1.5, \ldots, 5, 7.5, 10$. In the second setting, we double the number of cut points by including the mid-points of all the intervals from the first setting. We then evaluate the performance of the estimator for $\mu(t)$ at 5 selected time points, $t = 1, \ldots, t = 5$ respectively. For each setting, 1,000 data sets are simulated.

Results for the third simulation study are shown in Table 3.4 and Table 3.5. The estimator for $\mu(t)$ is approximately unbiased under both piecewise-constant baseline rates and linear baseline rates settings for $d\mu(t)$, with the bias reduced with increasing cluster size. On the other hand, the bias increases with t since the number of subjects at risk decreases with increasing t. The ASE is generally similar to the ESD, and the CP is close to 95%. Results for $L = 6$ and $L = 12$ are almost the same under piecewise-constant baseline rates setting. When data are simulated from the linear baseline rates model, the piecewise-constant baseline rates model with $L = 12$ does not seem to produce a better estimator than the model with $L = 6$ for $\mu(t)$ at the 5 selected time points, in terms of unbiasedness and efficiency.

3.6 Application

We applied the proposed marginal models with piecewise-constant baseline rates to the study of hospitalization days among Medicare dialysis patients. Between 2005 and 2007, there were 345,937 Medicare dialysis patients from 5,302 dialysis facilities being hospitalized in the U.S. with facility sizes varying from 3 to 2923 dialysis patients. Hospitalization days per patient ranged from 1 to 788 with an average of

Table 3.4: Results of the third simulation study with $\beta_1 = 0.5$, $\beta_2 = 1$, $\mu_{0k}(t) = t$ and 1000 replicates.

				$L = 6$				$L = 12$			
k	n_k	t	$\mu_{0k}(t)$	BIAS	ASE	ESD	CP	BIAS	ASE	ESD	CP
1	50	1	1	0.002	0.192	0.201	0.926	0.003	0.193	0.201	0.928
		2	2	0.034	0.364	0.376	0.932	0.035	0.364	0.376	0.932
		3	3	0.066	0.541	0.570	0.926	0.067	0.541	0.570	0.928
		4	4	0.129	0.728	0.755	0.933	0.130	0.728	0.755	0.934
		5	5	0.215	0.932	0.962	0.924	0.216	0.932	0.962	0.925
2	100	1	1	0.004	0.141	0.145	0.935	0.004	0.141	0.149	0.935
		2	2	0.024	0.260	0.277	0.932	0.024	0.260	0.277	0.934
		3	3	0.084	0.389	0.415	0.929	0.084	0.389	0.415	0.929
		4	4	0.154	0.528	0.552	0.938	0.154	0.529	0.552	0.938
		5	5	0.218	0.675	0.706	0.938	0.218	0.675	0.707	0.940
3	200	1	1	0.017	0.102	0.110	0.921	0.017	0.102	0.110	0.921
		2	2	0.018	0.190	0.207	0.922	0.017	0.190	0.207	0.924
		3	3	0.006	0.280	0.302	0.927	0.005	0.280	0.302	0.925
		4	4	-0.013	0.372	0.398	0.929	-0.014	0.373	0.398	0.930
		5	5	-0.039	0.468	0.503	0.927	-0.040	0.469	0.503	0.928
4	500	1	1	0.008	0.063	0.069	0.923	0.008	0.063	0.069	0.923
		2	2	0.015	0.116	0.131	0.918	0.015	0.116	0.131	0.918
		3	3	0.033	0.172	0.191	0.924	0.032	0.172	0.192	0.921
		4	4	0.058	0.230	0.257	0.919	0.057	0.231	0.257	0.918
		5	5	0.088	0.292	0.328	0.913	0.087	0.293	0.328	0.913

Table 3.5: Results of the third simulation study with $\beta_1 = 0.5$, $\beta_2 = 1$, $\mu_{0k}(t) = 0.5t^2$ and 1000 replicates.

				$L = 6$				$L = 12$			
k	n_k	t	$\mu_{0k}(t)$	BIAS	ASE	ESD	CP	BIAS	ASE	ESD	CP
1	50	1	0.5	-0.005	0.113	0.119	0.932	0.002	0.115	0.121	0.932
		2	2	0.036	0.375	0.396	0.930	0.048	0.378	0.399	0.933
		3	4.5	0.136	0.832	0.874	0.932	0.149	0.835	0.877	0.934
		4	8	0.351	1.496	1.567	0.918	0.364	1.499	1.570	0.917
		5	12.5	0.789	2.438	2.571	0.919	0.798	2.441	2.573	0.922
2	100	1	0.5	-0.005	0.082	0.089	0.929	0.003	0.083	0.090	0.928
		2	2	0.031	0.266	0.290	0.927	0.043	0.268	0.291	0.925
		3	4.5	0.158	0.589	0.626	0.928	0.170	0.591	0.628	0.922
		4	8	0.403	1.081	1.131	0.926	0.412	1.083	1.133	0.924
		5	12.5	0.733	1.760	1.837	0.925	0.734	1.761	1.837	0.925
3	200	1	0.5	-0.002	0.059	0.063	0.921	0.006	0.060	0.064	0.925
		2	2	-0.007	0.195	0.216	0.921	0.004	0.196	0.217	0.925
		3	4.5	-0.046	0.424	0.459	0.920	-0.034	0.425	0.461	0.920
		4	8	-0.120	0.756	0.817	0.914	-0.113	0.757	0.817	0.918
		5	12.5	-0.206	1.201	1.281	0.915	-0.207	1.202	1.280	0.916
4	500	1	0.5	-0.004	0.037	0.039	0.934	0.004	0.038	0.040	0.944
		2	2	0.009	0.120	0.135	0.910	0.020	0.120	0.135	0.916
		3	4.5	0.053	0.261	0.300	0.904	0.063	0.262	0.300	0.901
		4	8	0.156	0.470	0.545	0.904	0.162	0.471	0.544	0.901
		5	12.5	0.329	0.758	0.868	0.895	0.326	0.758	0.866	0.899

22 days during the 3-year period. In this study, hospital days are viewed as recurrent event data, with time of follow-up defined as time from 90 days after the initiation of end-stage renal disease (ESRD) therapy. We use the 90-day period to assure that most patients are eligible for Medicare insurance either as their primary or secondary insurer. Patients who died during the first 90 days of ESRD are excluded from the analysis. Patients are subject to left truncation at the start of the observation period, Jan 1, 2005. Subjects are followed until the earliest of death and right censoring, with the latter defined as the earliest of Dec 31, 2007, 3 days prior to transplant and loss to follow-up. Since a patient's hospitalization rate may be influenced by the facility at which (s)he receives dialysis, we fitted a facility-stratified model to adjust for facility effects.

Patient characteristics of interest include age, race, gender, diabetes, ethnicity, nursing home (NH) status and body mass index (BMI). All covariates are coded as categorical variables through binary indicators. According to the proposed methods, we summarize patient hospital days as intermittent counts and exposure times in 6 time since ESRD intervals with cut points 90 days (time 0), 6 months, 1 year, 2 years, 3 years and 5 years. Patient age is recorded at the beginning of each interval. Nursing home status is recorded as whether a patient was in nursing home in the previous calendar year. All the rest variables are measured at the beginning of the study, thus are time independent. Results from the proposed method are summarized in Table 3.6 below.

All the included covariates significantly influence the recurrent rate of hospitalization days. When comparing within the same cluster, patients at the age 25-44 have lowest hospitalization rates among survivors with all the other patient mix held the same. Asian dialysis patients are more frequently hospitalized among sur-

Table 3.6: Analysis of hospitalization days for Medicare dialysis patients in the U.S.

Covariates	Estimates	SE	p-value
Age (in Years)			
0-24	0.0091	0.0265	0.7331
25-44	-0.0885	0.0085	$< .0001$
45-59	-0.0209	0.0057	0.0002
60-74	0	.	.
75+	-0.0140	0.0054	$< .0001$
Race			
African-American	-0.0706	0.0061	$< .0001$
Asian	-0.3430	0.0160	$< .0001$
Native	-0.0370	0.0230	0.0014
Other	0.0125	0.0203	0.5377
Caucasian	0	.	.
Gender			
Female	0.1147	0.0044	$< .0001$
Male	0	.	.
Diagnoses			
Diabetes	0.2524	0.0048	$< .0001$
Non-Diabetes	0	.	.
Ethnicity			
Hispanic	-0.1530	0.0091	$< .0001$
Non-Hispanic	0	.	.
Nursing Home Status			
Yes	0.7685	0.0057	$< .0001$
No	0	.	.
BMI			
Underweight	0.0371	0.0088	$< .0001$
Normal	0	.	.
Overweight	-0.0351	0.0057	$< .0001$
Obese	-0.0406	0.0057	$< .0001$

vivors. Female patients, diabetic patients, non-Hispanic patients and under-weighted patients have higher hospitalization rates among survivors than the corresponding comparable groups. Conditional on being alive, dialysis patients are more frequently hospitalized if h/she was in the nursing home in the previous calendar year.

3.7 Discussion

In this report, we propose a proportional rates model with cluster-specific piecewise-constant baseline rate function for recurrent event data, which applies to the settings with and without a terminal event. With the parametric setting for the baseline rate function, we are able to estimate the regression parameter and cumulative baseline rates based on intermittent counts and exposure times within each pre-specified interval, which is defined according to the "pieces" in the baseline rate function. The proposed method reduces data storage volume and speed up the computation. The Cox format of the estimating equation enables the feasibility of stratification, which is difficult to implement under the joint estimating equation approach when the number of clusters is relatively large, as the illustrating example in Section 3.6.

The proposed method is applicable to both recurrent event and failure time data from large registry study or large observational study such as claims data in insurance or hospitalization data. When the number of distinct event times is large, we can fold the data by recording the counts and exposure time in pre-specified intervals and analyze the folded data using the proposed method.

3.8 Appendix

3.8.1 Proof of Theorem III.1

Define

$$P_n(\beta) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{\ell=1}^{L} \{\beta^T Z_{i\ell} - \log S_{k\ell}^{(0)}(\beta)\} d_{ik\ell},$$

and let $W_n(\beta) = P_n(\beta) - P_n(\beta_0)$, which can be written as

$$W_n(\beta) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{\ell=1}^{L} \left\{ (\beta^T - \beta_0^T) Z_{i\ell} - \log \frac{S_{k\ell}^{(0)}(\beta)}{S_{k\ell}^{(0)}(\beta_0)} \right\} d_{ik\ell}.$$

With condition (a) to (e) in Section 3.4, the Strong Law of Large Number (SLLN) and the fact that $d_{ik\ell}$ and $S_{k\ell}^{(0)}(\beta)$ have bounded variation, we can show that $W_n(\beta)$ converges almost surely to

$$\mathcal{W}(\beta) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{\ell=1}^{L} \left\{ (\beta^T - \beta_0^T) Z_{i\ell} - \log \frac{s_{k\ell}^{(0)}(\beta)}{s_{k\ell}^{(0)}(\beta_0)} \right\} \alpha_{k\ell} t_{ik\ell} e^{\beta_0^T Z_{i\ell}},$$

for every β. Obviously,

$$\begin{aligned} \frac{\partial^2 W_n(\beta)}{\partial\beta\partial\beta^T} &= -n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{\ell=1}^{L} V_{k\ell}(\beta) d_{ik\ell} \\ &= -n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{\ell=1}^{L} \{Z_{i\ell} - \overline{Z}_{k\ell}(\beta)\}^{\otimes 2} d_{\bullet k\ell} / S_{k\ell}^{(0)}(\beta) t_{ik\ell} e^{\beta^T Z_{i\ell}} \end{aligned}$$

is negative semidefinite. Therefore, $W_n(\beta)$ is concave. By Theorem 10.8 of Rockafellar (1970), the convergence of $W_n(\beta)$ to $\mathcal{W}(\beta)$ is uniform on any compact set of β. Specifically, defining a compact set of β, $\mathcal{B}_r = \beta : \|\beta - \beta_0\| \le r$, we have

$$\sup_{\beta\in\mathcal{B}_r} \|W_n(\beta) - \mathcal{W}(\beta)\| \to 0. \tag{3.8}$$

In addition, $\partial\mathcal{W}(\beta_0)/\partial\beta = 0$ and $\partial^2\mathcal{W}(\beta_0)/\partial\beta\partial\beta^T = -A$, which is assumed to be negative semidefinite through condition (g). Hence, $\mathcal{W}(\beta)$ has a unique maximizer at β_0. In particular, $\sup_{\beta\in\partial\mathcal{B}_r}\{\mathcal{W}(\beta)\} < \mathcal{W}(\beta_0)$, where $\partial\mathcal{B}_r = \beta : \|\beta - \beta_0\| = r$ is

the boundary of $\mathcal{B}_r$. This fact, together with expression (3.8), implies that $W_n(\beta) < W_n(\beta_0)$ for all $\beta \in \partial\mathcal{B}_r$ and large n. Therefore, there must exists a maximizer of $W_n(\beta)$, i.e. the solution to $\partial W_n(\beta)\ \partial\beta = 0$, say $\widehat{\beta}$, in the interior of $\mathcal{B}_r$, and the argument in Jacobsen (1989) can be used to show the uniqueness of this maximizer. Since r can be arbitrarily small, letting $r \to 0$ yields that $\widehat{\beta} \overset{a.s.}{\to} \beta_0$ as $n \to \infty$.

The weak convergence of $\widehat{\beta}$ can be shown through the first order Taylor series expansion about $\beta = \beta_0$ on $U(\widehat{\beta})$ as

$$U(\widehat{\beta}) = U(\beta_0) + \frac{\partial U(\beta^*)}{\partial \beta}(\widehat{\beta} - \beta_0),$$

where β^* is on the line segment joining $\widehat{\beta}$ and β_0. It follows that

$$n^{1/2}(\widehat{\beta} - \beta_0) = \widehat{A}^{-1}(\beta^*) n^{-1/2} U(\beta_0).$$

The almost sure convergence of $\widehat{\beta}$ to β_0 and the fact that $\xi_{ik\ell}(\beta_0)$ is zero mean implies that $\widehat{A}(\beta^*) \overset{a.s.}{\to} A$ as $n \to \infty$.

Next, we derive the distribution of $n^{-1/2}U(\beta_0)$ beginning by the following decomposition

$$U(\beta_0) = \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{\ell=1}^{\ell}\{Z_{i\ell} - \bar{z}_{k\ell}(\beta_0)\}\xi_{ik\ell}(\beta_0) - \sum_{k=1}^{K}\sum_{\ell=1}^{\ell}\{\overline{Z}_{k\ell}(\beta_0) - \bar{z}_{k\ell}(\beta_0)\}\xi_{\bullet k\ell}(\beta_0). \tag{3.9}$$

The first term on the right-hand side of equation (3.9) is a sum of n i.i.d. distributed random vectors with zero mean and finite variance. The second term on the right-hand side of equation (3.9) is $o_p(n^{1/2})$ since $\overline{Z}_{k\ell}(\beta_0) - \bar{z}_{k\ell}(\beta_0) \overset{p}{\to} 0$ and $\|n^{-1/2}\xi_{\bullet k\ell}(\beta_0)\| = O(1)$ by the boundedness conditions (d) and (e).

Thus $n^{-1/2}U(\beta_0)$ converges weakly to a p-variate normal vector with mean 0 and covariance matrix $B(\beta_0)$ by the multivariate central limit theorem. From Slutsky's

theorem and the consistency of $\widehat{A}(\beta^*)$ to A, $n^{1/2}(\widehat{\beta} - \beta_0)$ converges to a p-variate normal vector with mean 0 and covariance matrix Σ.

3.8.2 Proof of Theorem III.2

We now derive the asymptotic properties for $\widehat{\alpha}_{k\ell}(\widehat{\beta})$. The asymptotic results of $\widehat{\mu}_{0k}(t;\widehat{\beta})$ then directly applies by combining the results of $\widehat{\alpha}_{k\ell}(\widehat{\beta})$ over t.

We first consider the following decomposition

$$\widehat{\alpha}_{k\ell}(\widehat{\beta}) - \alpha_{k\ell} = \phi_1 + \phi_2,$$

where $\phi_1 = \widehat{\alpha}_{k\ell}(\beta_0) - \alpha_{k\ell}$, and $\phi_2 = \widehat{\alpha}_{k\ell}(\widehat{\beta}) - \widehat{\alpha}_{k\ell}(\beta_0)$. We can write $\phi_1 = n^{-1}\xi_{\bullet k\ell}(\beta_0)/S^{(0)}_{k\ell}(\beta_0)$. The SLLN and condition (f) implies that $\phi_1 \stackrel{a.s.}{\to} 0$. By Taylor series expansion,

$$\phi_2 = -n^{-1}\frac{d_{\bullet k\ell}\overline{Z}_{k\ell}(\beta_0)}{S^{(0)}_{k\ell}(\beta_0)}(\widehat{\beta} - \beta_0) + o_p(n^{-1/2}).$$

By the boundedness conditions (d) and (f), and the almost sure convergence of $\widehat{\beta}$ to β_0, $\phi_2 \stackrel{a.s.}{\to} 0$. The almost sure convergence of $\widehat{\alpha}_{k\ell}(\widehat{\beta})$ to $\alpha_{k\ell}$ then follows. This result, together with (3.5) implies that $\widehat{\mu}_{0k}(t;\widehat{\beta})$ converges almost surely to $\mu_{0k}(t)$ uniformly in t.

Next, we prove the weak convergence of $n^{1/2}\{\widehat{\alpha}_{k\ell}(\widehat{\beta}) - \alpha_{k\ell}\}$. With the previously derived arguments,

$$\begin{aligned} n^{1/2}\phi_1 &= n^{-1/2}\sum_{i=1}^{n}\frac{\xi_{ik\ell}(\beta_0)}{S^{(0)}_{k\ell}(\beta_0)}, \\ n^{1/2}\phi_2 &= -n^{-1}\sum_{i=1}^{n}\frac{d_{ik\ell}\overline{Z}_{k\ell}(\beta_0)}{S^{(0)}_{k\ell}(\beta_0)}n^{1/2}(\widehat{\beta} - \beta_0) + o_p(1). \end{aligned}$$

Using condition (f),

$$n^{1/2}\phi_1 = n^{-1/2}\sum_{i=1}^{n}\frac{\xi_{ik\ell}(\beta_0)}{s^{(0)}_{k\ell}(\beta_0)}, \tag{3.10}$$

Using condition (a) to (f) and the SLLN,

$$-n^{-1}\sum_{i=1}^{n}\frac{d_{ik\ell}\overline{Z}_{k\ell}(\beta_0)}{S^{(0)}_{k\ell}(\beta_0)} \stackrel{a.s.}{\to} \alpha_{k\ell}\bar{z}_{k\ell}(\beta_0).$$

Combing with the fact that

$$n^{1/2}(\widehat{\beta}-\beta_0) = A^{-1}(\beta_0)n^{-1/2}\sum_{i=1}^{n}U_i(\beta_0)+o_p(1),$$

it follows that

$$n^{1/2}\phi_2 \;=\; n^{-1/2}\sum_{i=1}^{n}\alpha_{k\ell}\bar{z}_{k\ell}(\beta_0)A^{-1}(\beta_0)U_i(\beta_0)+o_p(1). \tag{3.11}$$

With (3.10) and (3.11),

$$n^{1/2}\{\widehat{\alpha}_{k\ell}(\widehat{\beta})-\alpha_{k\ell}\} = n^{-1/2}\sum_{i=1}^{n}\psi_{ik\ell}(\beta_0)+o_p(1).$$

This result, together with (3.5) implies that

$$n^{1/2}\{\widehat{\mu}_{0k}(t;\widehat{\beta})-\mu_{0k}(t)\} \;=\; n^{-1/2}\sum_{i=1}^{n}\sum_{\ell=1}^{L}\psi_{ik\ell}(\beta_0)(a_\ell\wedge t-a_{\ell-1})+o_p(1),$$

which converges to a zero-mean Gaussian process with covariance function $\Psi_k(s,t)$.

CHAPTER IV

Fixed Center Effect Model for Recurrent Event Data

4.1 Introduction

In many large registry studies, comparisons among clinical centers are often of interest. Comparisons of center-specific outcomes to those of the population average help in understanding the facility's influence on patient prognosis and, hence, may ultimately improve quality of care. For example, dialysis facility-specific measures of hospitalization rates reflect a facility's quality of service (in terms of morbidity) relative to the population average level. Such measures thus assist patients in selecting a health care provider. Statistical models for multi-center studies have been developed for various outcomes, including clustered continuous, binary and time-to-failure data (e.g. Gould, 1998; Agresti and Hartzel, 2000; Localio, Berlin, Ten Have, and Kimmel, 2001; Glidden and Vittinghoff, 2004). The majority of such models devote primary attention to treatment differences, with center effects viewed as merely requiring adjustment. However, when the evaluation of center performance is of primary interest, inference on center effects becomes the ultimate purposes (as illustrated in the example above).

Existing methods of estimating center effects can be classified as fixed effects

models (FEM) or random effects models (REM). FEMs treat the center effects as fixed factors and typically yield a direct contrast between each center's performance relative to a pre-specified reference level. Conclusions drawn from an analysis based on FEMs will generally concern the centers included in the study only. In FEMs, parameters for center effects are usually estimated by including indicator variables for each center. However, when the number of centers is large, and with relatively small center sizes present, this estimation method leads to difficulty, as pointed out by Glidden and Vittinghoff (2004). On the other hand, REMs typically treat the center effects as random variates, assumed to follow a specific distribution across centers; often with only one additional parameter indexing the degree of dependence. The estimation procedure depend on the distribution of the random effects and the likelihood typically dose not have a closed form unless the random effect distribution is carefully chosen. Conclusions regarding center effects will be influenced by the whole population from which this particular center is randomly selected.

In this chapter, we propose fixed effects methods for estimating center effects in the context of recurrent event data. The problem motivating this research arose in a nationwide dialysis facility assessment study on hospitalization. Dialysis patients may be admitted from dialysis facilities to hospitals at any time since initiation of end-stage renal disease (ESRD) therapy. Frequency of hospitalization is an important factor to be considered when attempting to address issues regarding escalating medical costs. As a measure of hospitalization, hospital admissions can be viewed as recurrent events since dialysis patients may have multiple hospitalizations. Each year, over 5,000 dialysis facilities are assessed in the U.S., with facility sizes ranging from fewer than 10 to more than 2,000 dialysis patients. Each facility's hospitalization experience is then evaluated by comparing its outcomes to a risk adjusted

national average. The proportional rates model of Lin, Wei, Yang, and Ying (2000) is widely used for the analysis of recurrent event data, due to flexibility and its ease of implementation. The method proposed by Lin et al. (2000) assumes that subjects are independent units. To account for within cluster correlation, Schaubel and Cai (2005) proposed two semiparametric proportional rates models for clustered recurrent event data with 1) common baseline rates and 2) cluster specific baseline rates. When the estimation of center effects for recurrent event data is of interest, one might consider multiplicative center effects in the common baseline rates model of Schaubel and Cai (2005). We will discuss this model in detail in the next section. Due to the large number of facilities in the motivating example, traditional estimation methods that treat centers as categorical variables have many parameters and are thus often not feasible to implement; especially with a large number of distinct recurrent event times. In order to circumvent this problem, we propose a new estimation method for FEMs for recurrent event data, which is based on the ratio of center-specific observed to expected numbers of recurrent events. The proposed methods can also be generalized to a model in which a terminal event (e.g. death) is also present.

This chapter is organized as follows. We specify fixed center effects models (Section 4.2); describe the estimation method (Section 4.3); derive pertinent theoretical properties of the proposed methods (Section 4.4); assess the performance of the proposed estimators and compare them with those obtained from the traditional estimation method (Section 4.5); apply the proposed method to the motivating example (Section 4.6); and conclude the chapter with some discussion (Section 4.7).

4.2 Fixed Center Effect Model

Let $N^*(t)$ denote the number of events that occur over interval $(0, t]$. Let $Z(t)$ be a p-vector of external covariates (Kalbfleisch and Prentice, 2002). Let $k = 1, \ldots, K$ index the centers which are assumed to be independent of each other with center size n_k and let $n = \sum_{k=1}^{K} n_k$ be the population size. Let G be the center index. Assuming a multiplicative center effect on the rates of the counting process $N^*(t)$, the fixed center effects proportional rates model can be specified as

$$E\{dN^*(t)|G = k, Z(t)\} = \exp\{\beta^{*T}_0 Z(t)\}\theta^*_k d\mu^*_0(t), \tag{4.1}$$

where β^*_0 is an unknown p-vector of regression parameters, $\mu^*_0(t)$ is an unspecified baseline cumulative rate function, and θ^*_k represents the fixed center effects for the k-th center, $k = 1, \ldots, K$. The θ^*_k's are positive valued parameters which, for identifiability, are subject to the constraint $\sum_{k=1}^{K} w_k\theta^*_k = 1$, for some pre-specified weight w_k, $k = 1, \ldots, K$. Thus, θ^*_k can be interpreted as the center effect relative to this weighted average.

When a terminal event (death) is present, the recurrent event process stops at the terminal event (death) time D. Analogous to model (4.1), a partial marginal model (e.g. Cook and Lawless, 1997; Ye, Kalbfleisch, and Schaubel, 2007) with fixed center effects is considered. This specifies the occurrence rates for subject in the kth center conditional on being alive as

$$E\{dN^*(t)|G = k, Z(t), D \geq t\} = \exp\{\beta^T_0 Z(t)\}\theta_k d\mu_0(t). \tag{4.2}$$

A constraint on the θ_k's, such as $\sum_{k=1}^{K} w_k\theta_k = 1$ with known weight terms w_k, is again necessary for the identifiability of $\mu_0(t)$.

One may notice that by letting $D \to \infty$, model (4.2) reduces to model (4.1).

Thus, we emphasize model (4.2) henceforth with estimation and inference under model (4.1) being similar to model (4.2).

Another FEM incorporates center effects within the exponential function as

$$E\{dN^*(t)|G=k,Z(t),D\geq t\} \;=\; \exp\{\eta_k+\beta_0^TZ(t)\}d\mu_0(t), \tag{4.3}$$

where η_k denote center effects with $\exp(\eta_k)=\theta_k$. Model (4.3) is equivalent to model (4.2), but this form is used more frequently due to the fact that η_k in model (4.3) can be estimated directly together with the regression parameters β_0. The drawback, however, is the large dimension of the parameter space when the number of centers are large. In addition, the estimates of η_k's can be $-\infty$ which plays havoc with standardization. In the next section, we propose a new estimator for center effects θ_k under model (4.2) without increasing the dimension of the design matrix as the number of centers increases. We also derive an alternative estimator based on the commonly used estimator for η_k so that it is comparable to $\widehat{\theta}_k$, i.e., it is relative to the same reference.

In the following, we specify the independent left truncation and right censoring assumption. In many applications, subjects may enter the study at different stages of follow-up and be followed for a limited time. Thus, $N^*(t)$ may not be fully observed before D. Let B and C denote the left truncation and right censoring time respectively. Define the at risk process $Y(t)=I(B<t\leq C\wedge D)$ with $a\wedge b=\min(a,b)$. The observed cumulative number of events as of time t is $N(t)=\int_0^tY(s)dN^*(s)$. Under the assumption that $N^*(t)$ is subject to independent left truncation and right censoring, which can be specified as

$$E\{dN^*(t)|G=k,Z(t),Y(t)=1\}=E\{dN^*(t)|G=k,Z(t),D\geq t\},$$

we have that

$$E\{dN(t)|G=k,Z(t),Y(t)\} = Y(t)\exp\{\beta_0^T Z(t)\}\theta_k d\mu_0(t). \quad (4.4)$$

4.3 Estimation Methods

In this section, we propose new methods for estimating fixed center effects, θ_k, based on the ratio of the observed number of recurrent events to that expected under model (4.2). One may notice that model (4.2) is actually a special case of cluster-specific proportional rates model with $d\mu_{0k}(t) = \theta_k d\mu_0(t)$. Schaubel and Cai (2005) studied the cluster-specific proportional rates model for recurrent event data. With the estimators for β and $\mu_{0k}(t)$ as in Schaubel and Cai (2005), one is then able to obtain an estimator for $\mu_0(t)$ based on the constraint of θ_k specified in Section 4.2. An estimator of θ_k follows by replacing the true values with the corresponding estimators for β and $\mu_{0k}(t)$ in the expected number of recurrent events for center k. To compare with the proposed estimators, we consider transforming the commonly used estimators for η_k from model (4.3). After transformation, the familiar estimators have the same reference level as the proposed estimators.

We begin by defining the following relevant notation. For subject i ($i = 1,\ldots,n$), let $N_i(t)$, $Y_i(t)$, $Z_i(t)$, G_i, B_i and C_i be as defined above. Let $G_{ik} = I(G_i = k)$, where $I(\cdot)$ is an indicator function. In addition, let $Y_{ik}(t) = G_{ik}Y_i(t)$ and $dN_{ik}(t) = G_{ik}dN_i(t)$. We assume that $\{N_i(t), Y_i(t), Z_i(t), G_i\}_{i=1}^n$ are independent and identically distributed. For $d = 0, 1, 2$ and $k = 1,\ldots,K$, let

$$\begin{aligned} S_k^{(d)}(\beta,t) &= n^{-1}\sum_{i=1}^n Y_{ik}(t)Z_i(t)^{\otimes d}\exp\{\beta^T Z_i(t)\}, \\ \overline{Z}_k(\beta,t) &= S_k^{(1)}(\beta,t)/S_k^{(0)}(\beta,t), \\ V_k(\beta,t) &= S_k^{(2)}(\beta,t)/S_k^{(0)}(\beta,t) - \overline{Z}_k(\beta,t)^{\otimes 2} \end{aligned}$$

where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$ for a vector a.

Under the independent left truncation and right censoring assumptions, one can see that the score function is

$$U(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^{\tau} \{Z_i(t) - \overline{Z}_k(\beta, t)\} dM_{ik}(t), \tag{4.5}$$

where

$$dM_{ik}(t) = dN_{ik}(t) - Y_{ik}(t) \exp\{\beta_0^T Z_i(t)\} \theta_k d\mu_0(t).$$

We can simplify (4.5) to

$$U(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^{\tau} \{Z_i(t) - \overline{Z}_k(\beta, t)\} dN_{ik}(t), \tag{4.6}$$

and thus an estimator $\widehat{\beta}$ of β can be obtained by solving $U(\beta) = 0$. Schaubel and Cai (2005) derived the large sample properties for $\widehat{\beta}$ under the setting that $K \to \infty$. Theoretical results for $\widehat{\beta}$ under the setting that $n \to \infty$ can be obtained similarly.

The Breslow-Aalen-type estimators of $\mu_{0k}(t) = \theta_k \mu_0(t)$ for $k = 1, \ldots, K$ are given by

$$\widehat{\mu}_{0k}(t)(\widehat{\beta}, t) = n^{-1} \sum_{i=1}^{n} \int_0^t \frac{dN_{ik}(s)}{S_k^{(0)}(\widehat{\beta}, s)}.$$

With $\mu_{0k}(t) = \theta_k \mu_0(t)$, our chief interest is focused on $\mu_0(t)$ and θ_k, $k = 1, \ldots, K$. Under the constraint for the θ_k's in model (4.2), $\sum_{k=1}^{K} w_k \theta_k = 1$, it follows that

$$\sum_{k=1}^{K} \{w_k \theta_k \mu_0(t)\} = \mu_0(t).$$

Therefore, a natural estimator of $\mu_0(t)$ is

$$\widehat{\mu}_0(\widehat{\beta}, t) = \sum_{k=1}^{K} w_k \widehat{\mu}_{0k}(\widehat{\beta}, t). \tag{4.7}$$

Next, we consider estimation for θ_k in model (4.2). From (4.4), we have that

$$E\{dN_{ik}(t) | G_i = k, Z_i(t), Y_{ik}(t)\} \quad = \quad Y_{ik}(t) \exp\{\beta_0^T Z_i(t)\} \theta_k d\mu_0(t).$$

It then follows by summing over n on both sides of the above equation that

$$\sum_{i=1}^{n} E\{dN_{ik}(t)|G_i = k, Z_i(t), Y_{ik}(t)\} \;\; = \;\; \theta_k d\mu_0(t) n S_k^{(0)}(\beta_0, t).$$

Thus, a natural estimator for θ_k is

$$\widehat{\theta}_k = \frac{\overline{N}_k(\tau)}{\int_0^\tau S_k^{(0)}(\widehat{\beta}, t) d\widehat{\mu}_0(\widehat{\beta}, t)}, \tag{4.8}$$

where $\overline{N}_k(\tau) = n^{-1} \sum_{i=1}^n \int_0^\tau dN_{ik}(t) = n^{-1} \sum_{i=1}^n N_{ik}(\tau)$. This estimator $\widehat{\theta}_k$ has a standardized ratio format. Specifically, the numerator is the observed number of events up to time τ, while the denominator is the estimated expected number of events based on a (weighted) average center effect for subjects with characteristics and at-risk patterns observed in the kth center.

Finally, we consider an alternative estimator based directly on the model (4.3). In order to avoid identifiability issues, investigators usually choose a reference center and estimate the difference between the selected center effects and the reference level. With center r as the reference center, the estimators $\widetilde{\eta}_k$ from model (4.3) can be thought of as estimating $\eta_k - \eta_r$ with $\widetilde{\eta}_r = 0$. Since the reference level under model (4.2) is $\sum_{k=1}^K w_k \theta_k$ and $\exp(\eta_k) = \theta_k$, an alternative estimator for θ_k is

$$\widetilde{\theta}_k = \frac{\exp(\widetilde{\eta}_k + \eta_r)}{\sum_{k=1}^K w_k \exp(\widetilde{\eta}_k + \eta_r)}.$$

This estimator can be written as

$$\widetilde{\theta}_k = f_k(\widetilde{\boldsymbol{\eta}}^r), \tag{4.9}$$

where $\widetilde{\boldsymbol{\eta}}^r = (\widetilde{\eta}_1, \ldots, \widetilde{\eta}_K)^T$ and

$$f_k(x_1, \ldots, x_K) = \frac{\exp(x_k)}{\sum_{k=1}^K w_k \exp(x_k)}.$$

4.4 Asymptotic Results

To derive large sample properties of the parameters of interest, we first impose the following regularity conditions for $i = 1, \ldots, n$ and $k = 1, \ldots, K$.

(a) $\{N_i(t), Y_i(t), Z_i(t), G_i\}_{i=1}^n$ are independent and identically distributed;

(b) $E\{Y_i(t)\} > 0$, for all $t \in (0, \tau]$;

(c) $N_{ik}(\tau)$ are bounded;

(d) $Z_i(\cdot)$ have bounded total variations, i.e. $|Z_{ji}(0)| + \int_0^\tau |dZ_{ji}(t)| < c_Z < \infty$ for all $j = 1, \ldots, p$, where $Z_{ji}(\cdot)$ is the jth component of $Z_i(\cdot)$ and c_Z is a constant;

(e) $\mu_0(\tau)$ is bounded;

(f) $P\{G_i = k\} > 0$;

(g) There exists a neighborhood $\mathcal{B}$ of β_0 such that the following hold. For $d = 0, 1, 2$ and $k = 1, \ldots, K$, $s_k^{(d)}(\beta, t)$ is a continuous function of $\beta \in \mathcal{B}$ uniformly in $t \in [0, \tau]$, where $s_k^{(d)}(\beta, t)$ is the limiting values of $S_k^{(d)}(\beta, t)$; functions $s_k^{(1)}(\beta, t)$ and $s_k^{(2)}(\beta, t)$ are bounded and $s_k^{(0)}(\beta, t)$ is bounded away from 0 on $\mathcal{B} \times [0, \tau]$ with

$$s_k^{(1)}(\beta, t) = \frac{\partial}{\partial \beta} s_k^{(0)}(\beta, t), \; s_k^{(2)}(\beta, t) = \frac{\partial^2}{\partial \beta \partial \beta^T} s_k^{(0)}(\beta, t).$$

(h) Positive-definiteness of the matrix

$$A \;=\; \sum_{k=1}^{K} \int_0^\tau v_k(\beta_0, t) s_k^{(0)}(\beta_0, t) \theta_k d\mu_0(t),$$

where $v_k(\beta, t) = s_k^{(2)}(\beta, t)/s_k^{(0)}(\beta, t) - \bar{z}_k(\beta, t)^{\otimes 2}$ and $\bar{z}_k(\beta, t) = s_k^{(1)}(\beta, t)/s_k^{(0)}(\beta, t)$.

We now describe the asymptotic property of $\widehat{\mu}_0(\widehat{\beta}, t)$.

Theorem IV.1. *Under regularity conditions* $(a)-(h)$, $\widehat{\mu}_0(\widehat{\beta},t)$ *converges almost surely to* $\mu_0(t)$, *uniformly in* $t\in[0,\tau]$; $n^{1/2}\{\widehat{\mu}_0(\widehat{\beta},t)-\mu_0(t)\}$ *converges in distribution to a Gaussian process with covariance function* $\xi(s,t)=E\{\Phi_i(\beta_0,s)\Phi_i(\beta_0,t)\}$, *where*

$$\begin{aligned}
\Phi_i(\beta,t) &= h(\beta,t)^T A^{-1}\Psi_{1i}(\beta)+\Psi_{2i}(\beta,t),\\
h(\beta,t) &= -\sum_{k=1}^{K} w_k\int_0^t \bar{z}_k(\beta,s)\theta_k d\mu_0(s),\\
\Psi_{1i}(\beta) &= \sum_{k=1}^{K}\int_0^\tau \{Z_i(t)-\bar{z}_k(\beta,t)\}dM_{ik}(t)\\
\Psi_{2i}(\beta,t) &= \sum_{k=1}^{K} w_k\int_0^t s_k^{(0)}(\beta,s)^{-1}dM_{ik}(\beta,s),\\
M_{ik}(\beta,t) &= N_{ik}(t)-\int_0^t Y_{ik}(s)\exp\{\beta^T Z_i(s)\}\theta_k d\mu_0(s).
\end{aligned}$$

In the Appendix, we prove that $\xi(s,t)$ can be consistently estimated by $\widehat{\xi}(s,t)$, where $\widehat{\xi}(s,t)=n^{-1}\sum_{i=1}^{n}\{\widehat{\Phi}_i(\widehat{\beta},s)\widehat{\Phi}_i(\widehat{\beta},t)\}$ with $\widehat{\Phi}_i(\widehat{\beta},t)$ obtained by replacing all limiting values in $\Phi_i(\beta_0,t)$ with their empirical counterparts and replacing $\theta_k\mu_0(t)$ with $\widehat{\mu}_{0k}(\widehat{\beta},t)$.

Next, we establish the asymptotic property for $\widehat{\theta}_k$.

Theorem IV.2. *For* $k=1,\ldots,K$, *under conditions* $(a)-(h)$, $\widehat{\theta}_k\to\theta_k$ *almost surely as* $n\to\infty$; $n^{1/2}(\widehat{\theta}_k-\theta_k)$ *is asymptotically normally distributed with mean zero and covariance* $\Sigma_k^P=\mathcal{E}\{\Gamma_{ki}(\beta_0)^2\}$, *where*

$$\begin{aligned}
\Gamma_{ki}(\beta_0) &= P_{ki}^1(\beta_0)+P_{ki}^2(\beta_0)+P_{ki}^3(\beta_0),\\
P_{ki}^1(\beta_0) &= -\theta_k^2 o_k^{-1}\left\{\int_0^\tau s_k^{(1)}(\beta_0,t)d\mu_0(t)+\int_0^\tau s_k^{(0)}(\beta_0,t)dh(\beta_0,t)\right\}^T A^{-1}\Psi_{1i}(\beta_0),\\
P_{ki}^2(\beta_0) &= -\theta_k^2 o_k^{-1}\int_0^\tau s_k^{(0)}(\beta_0,t)d\Psi_{2i}(\beta_0,t),\\
P_{ki}^3(\beta_0) &= \theta_k o_k^{-1}\int_0^\tau dM_{ik}(\beta_0,t)
\end{aligned}$$

with $o_k=\lim_{n\to\infty}\bar{N}_k(\tau)$ *and* $\bar{N}_k(\tau)=n^{-1}N_k(\tau)$.

In the Appendix, we prove that Σ_k can be consistently estimated by $\widehat{\Sigma}_k$, where $\widehat{\Sigma}_k = n^{-1}\sum_{i=1}^{n}\widehat{\Gamma}_{ki}(\widehat{\beta})^2$ with $\widehat{\Gamma}_{ki}(\widehat{\beta})$ obtained by replacing all limiting values in $\Gamma_{ki}(\beta_0)$ with their empirical counterparts and replacing $\theta_k\mu_0(t)$ with $\widehat{\mu}_{0k}(\widehat{\beta}, t)$.

The consistency of $\widetilde{\theta}_k$ follows from the Slutsky's theorem and the known large sample properties of the regression parameters under the proportional rates model (Lin et al., 2000). Its asymptotic distribution can be obtained using the delta method. Specifically, let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^T$ and $\widetilde{\boldsymbol{\theta}}$ be the corresponding estimator. It can be easily shown that $n^{1/2}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})$ converges to a K-variate normal vector with mean zero and a covariance matrix, which can be estimated as $\widetilde{\Sigma} = \widetilde{Q}^T\widetilde{\Omega}\widetilde{Q}$, where $\widetilde{Q}$ is a $K \times (K-1)$ matrix with the k-th row equal to the partial derivative of $f_k(\widetilde{\boldsymbol{\eta}}^r)$ (omitting the r-th argument) and $\widetilde{\Omega}$ is the consistent estimator of the asymptotic covariance matrix for $\widetilde{\boldsymbol{\eta}}$, where $\widetilde{\boldsymbol{\eta}}$ is a $(K-1)$-vector without the r-th element in $\widetilde{\boldsymbol{\eta}}^r$.

Comparison between center effects is another point of interest in FEM. If one wants to compare center effects θ_k and θ_l, the null hypothesis is $H_0 : \theta_k = \theta_l$. We consider three test statistics with one based on the proposed method (PM) and two based on the commonly used method (CM). For PM, one can first obtain the asymptotic distribution for $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, where $\widehat{\boldsymbol{\theta}}$ denotes the estimators for $\boldsymbol{\theta}$ from PM. According to Theorem IV.2, $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ converges to a K-variate normal vector with mean zero and covariance matrix $\Sigma = \mathcal{E}\{\Gamma_1(\beta_0)^2\}$, where Γ_1 is a K-vector with the kth element equal to $\Gamma_{k1}(\beta_0)$. It can be easily shown that $\widehat{\Sigma}$, a consistent estimator for Σ, can be obtained similarly to $\widehat{\Sigma}_k$ for Σ_k. The test statistic is then $T_1 = e_{kl}^T(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})/(n^{-1}e_{kl}^T\ \widehat{\Sigma}\ e_{kl})^{1/2}$ for the one-sided test and T_1^2 for the two-sided, where e_{kl} is a K-vector with the k-th element equal to 1, the l-th element equal to -1 and the rest $K-2$ elements equal to 0. For CM, two test statistics are available. One is based on the transformed estimator $\widetilde{\boldsymbol{\theta}}$, and is $T_2 = e_{kl}^T(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})/(n^{-1}e_{kl}^T\ \widetilde{\Sigma}\ e_{kl})^{1/2}$

for the one-sided and T_2^2 for the two-sided. The other one is the Wald statistic for $\widetilde{\eta}_k$, the estimator for $\eta_k - \eta_l$ obtained from CM with reference center l. This yields $T_3 = e_k^T(\widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta})/(n^{-1}e_k^T\ \widetilde{\Omega}\ e_k)^{1/2}$ for the one-sided test and T_3^2 for the two-sided, where $\boldsymbol{\eta}$ is the true value of $\widetilde{\boldsymbol{\eta}}$ and e_k is a $(K-1)$-vector with k-th element equal to 1 and the rest equal to 0.

4.5 Simulation

In this section, we assess the finite sample properties of the proposed estimator $\widehat{\theta}_k$ through a series of simulation studies. Recurrent event data are simulated from the conditional Poisson model

$$E\{dN_{ik}^*(t)|G_i = k, Z_i, D_i \geq t, W_i\} = W_i \exp\{\beta_0^T Z_i\}\theta_k d\mu_0(t), \tag{4.10}$$

where W_i follows a gamma distribution with mean 1 and variance 0.5, Z_i follows a Bernoulli distribution with $P(Z_i = 1) = 0.5$, $\beta_0 = 0.5$, and $\mu_0(t) = 0.5t$. The censoring time is fixed at 3 and the death time $D_i \sim U(0, 9)$. For the first 12 centers, we consider the combination of $\theta_k = 0.5, 1, 1.5$ and $n_k = 20, 50, 100, 200$. The remaining centers have center size of 50 and center effect alternating among 0.5, 1, and 1.5. The total number of centers is always a multiple of three so that the weighted averaged center effect equals to 1 with weights $w_k = 1/K$. Obviously, model (4.10) satisfies model (4.2). We consider $K = 30, 60, 150$.

We then compare the two estimators (4.8) and (4.9) from PM and CM respectively. The results shown in Table 4.1 are based on 1000 replicates. We report the bias of the sampling mean of the estimators (BIAS), the averaged standard error estimators (ASE), the empirical standard deviation of the estimators (ESD) and the 95% coverage probability (CP) for both $\widehat{\theta}_k$ and $\widetilde{\theta}_k$, $k = 1, \ldots, 12$. In the last column, we report the asymptotic relative efficiency of $\widetilde{\theta}_k$ with respect to $\widehat{\theta}_k$ (ARE), which

is calculated as the ratio of the squared ESD($\widehat{\theta}_k$) to the squared ESD($\widetilde{\theta}_k$). There is very little difference between the two estimators at least in the cases considered. Both are unbiased and CPs become accurate as the center size increases. The results with $K = 150$ are similar to $K = 60$ and are not shown. As expected, the proposed estimator does not overperform the commonly used estimator since the partial likelihood estimator is semiparametric efficient. However, the proposed estimator is considerably more efficient in terms of computational time, which is important when the number of centers is very large. The times to compute 1000 replicates for PM and CM are respectively 1.5 hours and 4 hours when $K = 30$, 2.5 hours and 17.33 hours when $K = 60$, and 8 hours and 150 hours when $K = 150$. The advantage of PM becomes increasingly pronounced as the number of centers increases (note: the ESRD database for the motivating example has >5000 centers). This is due to the fact that for PM, the dimension of the parameter space does not change with the number of centers, while it does change under CM. Another important advantage of the PM estimates is its observed-to-expected structure, which is simple to describe to non-statisticians.

In the second simulation study, we compare the two estimators when the recurrent events and the terminating event are correlated. Specifically, we simulate $D_i = W_i D_i^1$ with $D_i^1 \sim U(0, 9)$. We keep the rest settings the same as the first simulation study. The results based on 1000 replicates are shown in Table 4.2. Similar to the first simulation study, both estimators are unbiased with closed ESDs and ASEs. The CPs becomes closer to the nominal value as center size increases. Again, there is little difference between the two estimators except that PM has remarkably shorter computation time.

Table 4.1: Results of the first simulation study with $\beta = 0.5$, $\mu_{0k}(t) = 0.5t$ and 1000 replicates comparing $\widehat{\theta}_k$ and $\widetilde{\theta}_k$.

			PM($\widehat{\theta}_k$)				CM($\widetilde{\theta}_k$)				
K	n_k	θ_k	BIAS	ASE	ESD	CP	BIAS	ASE	ESD	CP	ARE
30	20	0.5	-0.002	0.140	0.154	0.892	-0.002	0.140	0.154	0.895	1.000
		1	-0.002	0.221	0.236	0.907	-0.002	0.222	0.236	0.908	0.991
		1.5	-0.012	0.295	0.323	0.903	-0.012	0.296	0.323	0.903	0.993
	50	0.5	-0.001	0.092	0.095	0.931	-0.001	0.092	0.095	0.929	1.000
		1	-0.004	0.146	0.146	0.930	-0.004	0.146	0.146	0.930	1.000
		1.5	-0.001	0.198	0.208	0.926	-0.001	0.199	0.208	0.925	0.990
	100	0.5	0.002	0.067	0.067	0.950	0.002	0.067	0.067	0.951	1.000
		1	0.004	0.108	0.102	0.959	0.004	0.108	0.102	0.957	1.000
		1.5	-0.010	0.145	0.150	0.935	-0.010	0.145	0.150	0.935	1.000
	200	0.5	0.001	0.049	0.047	0.946	0.000	0.049	0.047	0.946	1.000
		1	0.008	0.080	0.085	0.942	0.008	0.080	0.085	0.942	1.000
		1.5	0.000	0.108	0.107	0.947	0.000	0.108	0.107	0.946	1.000
60	20	0.5	-0.002	0.141	0.155	0.892	-0.002	0.141	0.155	0.893	1.000
		1	-0.002	0.225	0.240	0.908	-0.002	0.226	0.240	0.910	0.991
		1.5	-0.011	0.301	0.329	0.909	-0.011	0.302	0.329	0.910	0.993
	50	0.5	-0.002	0.092	0.095	0.930	-0.001	0.092	0.095	0.930	1.000
		1	-0.004	0.147	0.148	0.929	-0.004	0.147	0.148	0.929	1.000
		1.5	-0.001	0.201	0.213	0.922	-0.001	0.201	0.213	0.923	1.000
	100	0.5	0.002	0.067	0.067	0.945	0.002	0.067	0.067	0.946	1.000
		1	0.003	0.108	0.102	0.961	0.003	0.108	0.102	0.960	1.000
		1.5	-0.011	0.146	0.150	0.939	-0.011	0.146	0.150	0.938	1.000
	200	0.5	0.000	0.048	0.046	0.952	0.000	0.048	0.046	0.952	1.000
		1	0.007	0.078	0.084	0.947	0.007	0.078	0.084	0.945	1.000
		1.5	-0.001	0.106	0.105	0.944	-0.001	0.106	0.105	0.946	1.000

Table 4.2: Results of the second simulation study with $\beta = 0.5$, $\mu_{0k}(t) = 0.5t$ and 1000 replicates comparing $\widehat{\theta}_k$ and $\widetilde{\theta}_k$.

			PM($\widehat{\theta}_k$)				CM($\widetilde{\theta}_k$)				
K	n_k	θ_k	BIAS	ASE	ESD	CP	BIAS	ASE	ESD	CP	ARE
30	20	0.5	-0.001	0.135	0.149	0.893	-0.001	0.136	0.149	0.892	0.985
		1	-0.000	0.210	0.228	0.906	0.001	0.211	0.228	0.907	0.991
		1.5	-0.018	0.279	0.304	0.896	-0.016	0.280	0.305	0.898	0.993
	50	0.5	-0.003	0.089	0.093	0.924	-0.003	0.089	0.093	0.925	1.000
		1	-0.005	0.139	0.137	0.941	-0.003	0.139	0.138	0.939	1.000
		1.5	-0.005	0.185	0.199	0.919	-0.003	0.186	0.199	0.919	0.989
	100	0.5	0.005	0.065	0.066	0.934	0.006	0.065	0.066	0.933	1.000
		1	0.004	0.104	0.103	0.943	0.006	0.104	0.104	0.943	1.000
		1.5	-0.006	0.138	0.138	0.948	-0.004	0.138	0.138	0.951	1.000
	200	0.5	0.001	0.047	0.047	0.942	0.002	0.047	0.047	0.940	1.000
		1	0.008	0.076	0.078	0.942	0.009	0.076	0.078	0.940	1.000
		1.5	-0.004	0.102	0.102	0.952	-0.002	0.102	0.102	0.953	1.000
60	20	0.5	-0.001	0.136	0.150	0.892	-0.001	0.136	0.150	0.893	1.000
		1	0.000	0.213	0.232	0.907	0.001	0.214	0.232	0.911	0.991
		1.5	-0.017	0.285	0.309	0.895	-0.015	0.287	0.310	0.896	0.986
	50	0.5	-0.004	0.089	0.093	0.921	-0.003	0.089	0.093	0.923	1.000
		1	-0.005	0.140	0.139	0.940	-0.004	0.140	0.139	0.939	1.000
		1.5	-0.005	0.188	0.204	0.905	-0.003	0.188	0.204	0.907	1.000
	100	0.5	0.005	0.065	0.066	0.932	0.006	0.065	0.066	0.933	1.000
		1	0.004	0.103	0.104	0.943	0.005	0.104	0.104	0.943	0.981
		1.5	-0.006	0.138	0.140	0.948	-0.004	0.138	0.141	0.947	1.000
	200	0.5	0.001	0.047	0.046	0.944	0.002	0.047	0.046	0.944	1.000
		1	0.007	0.075	0.076	0.939	0.008	0.075	0.077	0.939	1.000
		1.5	-0.005	0.100	0.100	0.948	-0.003	0.101	0.100	0.945	1.000

4.6 Application

In this section, we apply the fixed center effects model to the dialysis patients hospital admission study and estimate the center effects using the proposed method. The hospital admission records are obtained using both national end-stage renal disease (ESRD) registry data and that from the Centers for Medicare and Medicard Services (CMS). Patients are included in the study on day 91 since diagnosis of ESRD in order to assure that most patients are eligible for Medicare insurance either as their primary or secondary insurer. For illustrative purposes, we analyze hospital admission from dialysis patients treated in Michigan dialysis facilities in 2008. Therefore, the data are subject to left truncation (Jan. 1, 2008), a terminating event (death), and right censoring (Dec. 31, 2008 and loss to follow-up). In 2008, there were 8,204 dialysis patients treated in 173 Medicare-certified dialysis facilities in Michigan, with facility sizes ranging from 1 to 187 patients. Times of hospital admissions are recorded for each patient and are measured in days. The average number of hospital admissions per patient was 2.7.

We adjust for the following patient characteristics: age, race, gender, ethnicity and diabetes status. Patient age is categorized as 0-14, 15-24, 25-44, 45-59, 60-74, and $\geq$75. Race is classified as African-American, Caucasian, Asian, Native-American and other. Ethnicity includes 3 levels, Hispanic, non-Hispanic and missing/unknown. Similarly, diabetes status was represented by three levels: diabetic, non-diabetic and missing/unknown. Under the proposed method, we estimate the center effects relative to the weighted average of center effect with the weights defined as the sampling proportion of each facility (n_k/n). In figure 4.1, the estimated effects and the p-value from the one-sided test $H_0 : \theta_k > 1.2$ are plotted. There are 9 (5%)

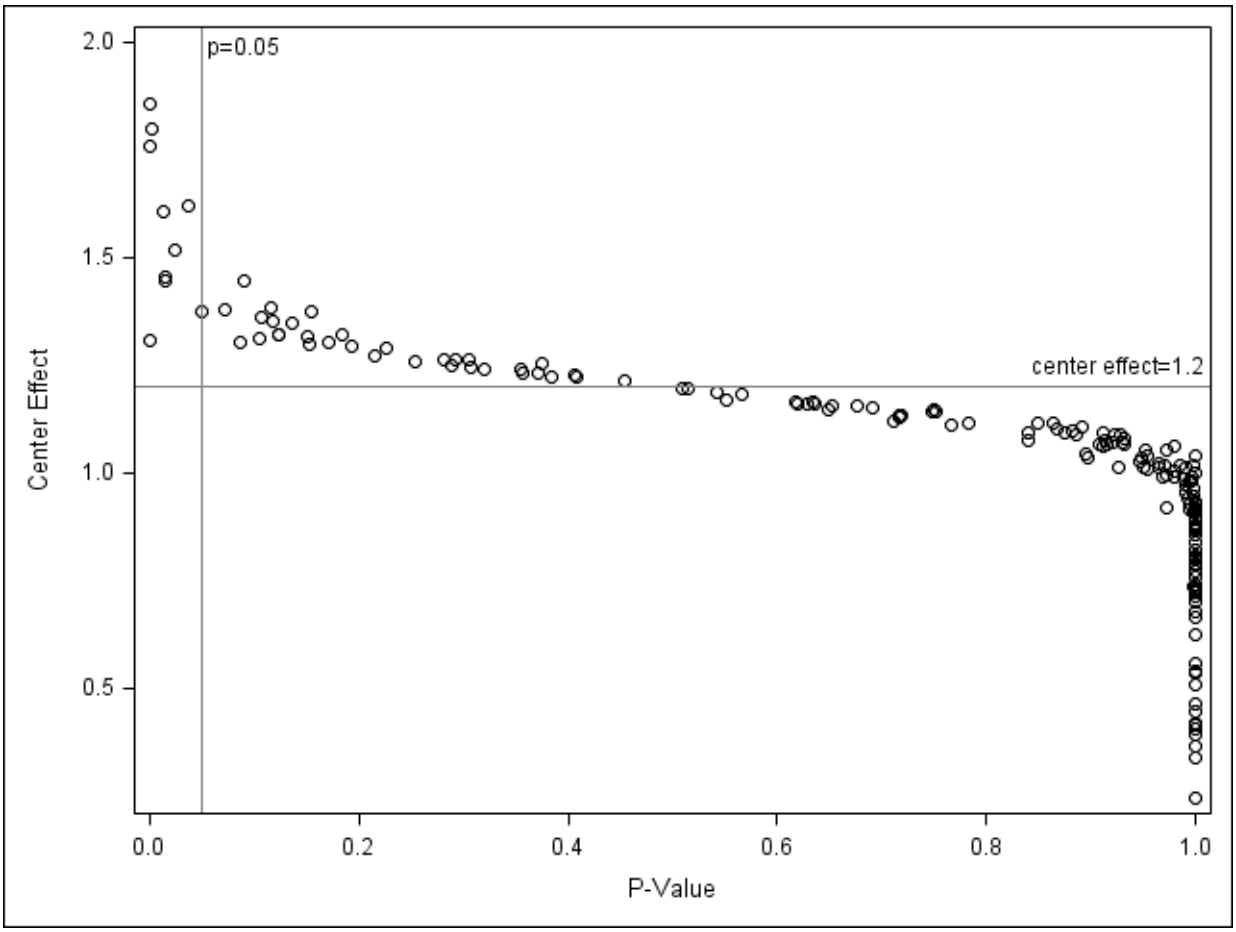

Figure 4.1: Scatter plot of estimated center effect from PM and the corresponding p-value for $H_0 : \theta_k > 1.2$.

centers flagged with p-value less than 0.05.

4.7 Discussion

In this chapter, we proposed an alternative estimator for center effects in the fixed center effects proportional rates model for recurrent event data. Both the absence and the presence of a terminal event are considered. The new estimator is the ratio of the observed recurrent events and the expected quantity based on a weighted average center effect and the same patient mix. The new estimator performs as well as the

commonly used estimator which is obtained by including parameter for centers in the regression parameters and transforming accordingly. The proposed estimation procedure is computationally more efficient, since the dimension of the regression parameter space only depends on the number of adjusting covariates. When the number of centers is large, such as large registry studies, the commonly used method may not be computationally feasible. In this case, particularly, one may consider the usage of the proposed estimator for evaluation of center effects. In addition, the standardization with respect to this multiplicative center effect θ_k instead of the additive center effect η_k within the exponential form avoids the problems when $\eta_k = -\infty$.

It should be noted that the interpretation of the center effect θ_k is relative to the weighted average center effect, which is constrained to be 1 in Section 4.2 for the purpose of identifiability. When the weight changes, the interpretation should also changes accordingly.

Another possible estimation method is based on the population averaged model, which can be obtained from (4.2) by integrating out G as

$$E\{dN^*(t)|Z(t), D \geq t\} \;=\; \exp\{\beta_0^T Z(t)\} d\mu_0(t) \sum_{k=1}^{K} \theta_k P\{G = k|Z(t), D \geq t\}.$$

If the distribution of G does not depend on patient characteristics and whether the patient is alive, i.e. $P\{G = k|Z(t), D \geq t\} = P\{G = k\}$, the population averaged model still keeps the proportional rates format as

$$E\{dN^*(t)|Z(t), D \geq t\} \;=\; \exp\{\beta_0^T Z(t)\} d\mu_0^*(t), \tag{4.11}$$

where $d\mu_0^*(t) = d\mu_0(t) \sum_{k=1}^{K} \theta_k^* P\{G = k\} = d\mu_0(t)$ under the constraint that $\sum_{k=1}^{K} \theta_k P\{G = k\} = 1$. An estimator for $\mu_0(t)$ can be obtained from the marginal model with $\widehat{\beta}^T Z(t)$ as an offset. An estimator for θ_k can then be obtained similar

to $\widehat{\theta}_k$. Under this condition, $d\mu_0(t)$ can be interpreted as the population averaged baseline rate function, and θ_k can be interpreted as the center effect relative to the population level. If, in practice, the distribution of G does not depend on $Z(t)$ and is a constant over time, this estimator could be used without specifying the weight terms. For the motivating study, however, this condition may not be satisfied. For example, pediatric dialysis patients might more likely receive treatment from pediatric dialysis units.

The proposed method can be easily extended to the fixed center effects proportional hazards model for failure time data. Similarly, it can also be extended to other semiparametric models with appropriate constraint on the fixed center effects.

4.8 Appendix

4.8.1 Proof of Theorem IV.1

Consistency of $\widehat{\mu}_0(\widehat{\beta}, t)$

Let $\phi_0(t) = \widehat{\mu}_0(\widehat{\beta}, t) - \mu_0(t) = \phi_1(t) + \phi_2(t)$, where $\phi_1(t) = \widehat{\mu}_0(\widehat{\beta}, t) - \widehat{\mu}_0(\beta, t)$ and $\phi_2(t) = \widehat{\mu}_0(\beta_0, t) - \mu_0(t)$. By the triangle inequality,

$$\sup_{t\in[0,\tau]} |\phi_0(t)| \leq \sup_{t\in[0,\tau]} |\phi_1(t)| + \sup_{t\in[0,\tau]} |\phi_2(t)|. \tag{4.12}$$

A Taylor series expansion of

$$\phi_1(t) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} w_k \left\{ \int_0^t \frac{dN_{ik}(s)}{S_k^{(0)}(\widehat{\beta}, s)} - \int_0^t \frac{dN_{ik}(s)}{S_k^{(0)}(\beta_0, s)} \right\}$$

yields

$$\phi_1(t) = H(\beta^\dagger, t)^T (\widehat{\beta} - \beta_0) + o_p(n^{-1/2}),$$

where $\beta^\dagger$ lies in the line segment between $\widehat{\beta}$ and β_0 and $H(\beta, t) = n^{-1} \sum_{i=1}^{n} H_i(\beta, t)$ with

$$H_i(\beta, t) = -\sum_{k=1}^{K} w_k \int_0^t \frac{\bar{Z}_k(\beta, s)}{S_k^{(0)}(\beta, s)} dN_{ik}(s).$$

Since $N_{ik}(s)$, $S_k^{(1)}(\beta, s)$ are bounded and $S_k^{(0)}(\beta, s)$ is bounded away from 0 for $\beta \in \mathcal{B}$ and $s \in [0, \tau]$ as $n \to \infty$, it folows that $H(\beta, t) = n^{-1} \sum_{i=1}^{n} H_i(\beta, t)$ is bounded for sufficiently large n. Since $\widehat{\beta} \overset{a.s.}{\to} \beta_0$ as $n \to \infty$, it follows that

$$\sup_{t \in [0,\tau]} |\phi_1(t)| \overset{a.s.}{\to} 0. \tag{4.13}$$

Further, since $\sum_{k=1}^{K} w_k \theta_k = 1$, we find

$$\begin{aligned}
\phi_2(t) &= \sum_{k=1}^{K} w_k \left\{ \sum_{i=1}^{n} \int_0^t \frac{dN_{ik}(s)}{nS_k^{(0)}(\beta_0, s)} - \theta_k \mu_0(t) \right\} \\
&= \sum_{k=1}^{K} w_k \sum_{i=1}^{n} \int_0^t \frac{dN_{ik}(s) - Y_{ik}(s) \exp\{\beta_0^T Z_i(s)\} \theta_k d\mu_0(s)}{nS_k^{(0)}(\beta_0, s)} \\
&= n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} w_k \int_0^t \frac{dM_{ik}(\beta_0, s)}{S_k^{(0)}(\beta_0, s)}.
\end{aligned}$$

Since $S_k^{(0)}(\beta_0, s)$ is bounded away from 0 and $n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} M_{ik}(\beta_0, s) \overset{a.s.}{\to} 0$ by the Strong Law of Large Numbers (SLLN) as $n \to \infty$ for $s \in [0, \tau]$, we have that

$$\sup_{t \in [0,\tau]} |\phi_2(t)| \overset{a.s.}{\to} 0. \tag{4.14}$$

The uniform consistency of $\widehat{\mu}_0(\widehat{\beta}, t)$ follows from (4.12), (4.13) and (4.14).

Weak Convergence of $\widehat{\mu}_0(\widehat{\beta}, t)$

We now consider the process $n^{1/2}\phi_0(t) = n^{1/2}\phi_1(t) + n^{1/2}\phi_2(t)$. We see that

$$\begin{aligned}
n^{1/2}\phi_1(t) &= H(\beta_0, t)^T n^{1/2}(\widehat{\beta} - \beta_0) + o_p(1) \\
n^{1/2}\phi_2(t) &= n^{-1/2} \sum_{i=1}^{n} \widehat{\Psi}_{2i}(\beta_0, t),
\end{aligned}$$

where $\widehat{\Psi}_{2i}(\beta_0, t) = \sum_{k=1}^{K} w_k \int_0^t S_k^{(0)}(\beta_0, s)^{-1} dM_{ik}(\beta_0, s)$.

Under conditions $(a) - (g)$, it follows that $\sup_{t \in [0,\tau]} |\widehat{\mu}_{0k}(\widehat{\beta}, t) - \mu_{0k}(t)| \overset{a.s.}{\to} 0$, $k = 1, \ldots, K$, and $H(\beta_0, t) \overset{a.s.}{\to} h(\beta_0, t)$ for $t \in [0, \tau]$.

The partial likelihood score equation (4.6) can be written as

$$\begin{aligned} U(\beta) &= \sum_{i=1}^{n} \widehat{\Psi}_i(\beta), \\ \widehat{\Psi}_{1i}(\beta) &= \sum_{k=1}^{K} \int_0^\tau \{Z_i(t) - \bar{Z}_k(\beta, t)\} dM_{ik}(t). \end{aligned}$$

Under condition (e), $\widehat{\Psi}_{1i}(\beta)$ converges to $\Psi_{1i}(\beta)$ for $i = 1, \ldots, n$. A Taylor series expansion of the score equation at $\beta = \widehat{\beta}$ around β_0 yields

$$\begin{aligned} U(\beta_0) &= \widehat{I}(\beta_0)(\widehat{\beta} - \beta_0) + o_p(n^{-1/2}), \\ \widehat{I}(\beta_0) &= \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^\tau V_{ik}(t) dN_{ik}(t). \end{aligned}$$

As $n \to \infty$, $n^{-1}\widehat{I}(\beta_0) \xrightarrow{a.s.} A$, where A is the positive definite matrix defined in condition (f). We can see that

$$n^{1/2}\phi_1(t) = h(\beta_0, t)^T A^{-1} n^{-1/2} \sum_{i=1}^{n} \Psi_{1i}(\beta_0) + o_p(1). \tag{4.15}$$

Under condition (e), $\widehat{\Psi}_{2i}(\beta_0, t)$ converges to $\Psi_{2i}(\beta_0, t)$ for $i = 1, \ldots, n$. Hence

$$n^{1/2}\phi_2(t) = \sum_{i=1}^{n} \Psi_{2i}(\beta_0, t) + o_p(1). \tag{4.16}$$

Combining (4.15) and (4.16), we see that $n^{1/2}\phi_0(t) = n^{-1/2} \sum_{i=1}^{n} \Phi_i(\beta_0, t) + o_p(1)$, which converges weakly to a Gaussian Process with covariance function ξ.

4.8.2 Proof of Theorem IV.2

Consistency of $\widehat{\theta}_k$

From (4.8),

$$\widehat{\theta}_k - \theta_k = \frac{n^{-1} \sum_{i=1}^{n} \int_0^\tau \left[dN_{ik}(t) - Y_{ik}(t) \exp\{\widehat{\beta}^T Z_i(t)\} \theta_k d\widehat{\mu}_0(\widehat{\beta}, t) \right]}{\int_0^\tau S_k^{(0)}(\widehat{\beta}, t) d\widehat{\mu}_0(\widehat{\beta}, t)}. \tag{4.17}$$

With the uniform consistency of $\widehat{\beta}$ and $\widehat{\mu}_0(\widehat{\beta}, t)$ as $n \to \infty$,

$$n^{-1} \sum_{i=1}^{n} \left\{ \int_0^\tau Y_{ik}(t) \exp\{\widehat{\beta}^T Z_i(t)\} \theta_k d\widehat{\mu}_0(\widehat{\beta}, t) - \int_0^\tau Y_{ik}(t) \exp\{\beta_0^T Z_i(t)\} \theta_k d\mu_0(\beta, t) \right\} \xrightarrow{a.s.} 0.$$

In addition,

$$n^{-1}\sum_{i=1}^{n}\int_0^\tau \left[dN_{ik}(t) - Y_{ik}(t)\exp\{\beta_0^T Z_i(t)\}\theta_k d\mu_0(\beta,t)\right] = n^{-1}\sum_{i=1}^{n}\int_0^\tau dM_{ik}(\beta_0,t) \overset{a.s.}{\to} 0$$

by the SLLN as $n \to \infty$. Hence the numerator in (4.17) converges almost surely to 0 as $n \to \infty$. With the boundness condition (c) and (e), the denominator in (4.17) is also bounded. Hence,

$$\widehat{\theta}_k - \theta_k \overset{a.s.}{\to} 0.$$

Weak convergence of $\widehat{\theta}_k$

Let $\zeta_k = \theta_k^{-1}$ and $\widehat{\zeta}_k = \widehat{\theta}_k^{-1}$ for $k = 1, \ldots, K$. In the following, we will first show the asymptotic approximation of $n^{1/2}(\widehat{\zeta}_k - \eta_k)$ and then use the delta method to obtain the asymptotic properties for $n^{1/2}(\widehat{\theta}_k - \theta_k)$. Since

$$\begin{aligned} n^{1/2}(\widehat{\eta}_k - \eta_k) &= n^{1/2}\left\{\bar{N}_k(\tau)^{-1}\int_0^\tau S_k^{(0)}(\widehat{\beta},t)d\widehat{\mu}_0(\widehat{\beta},t) - \theta_k^{-1}\right\} \\ &= n^{1/2}\bar{N}_k(\tau)^{-1}\theta_k^{-1}\int_0^\tau \left\{S_k^{(0)}(\widehat{\beta},t)\theta_k d\widehat{\mu}_0(\widehat{\beta},t) - d\bar{N}_k(t)\right\}, \end{aligned}$$

$n^{1/2}(\widehat{\eta}_k - \eta_k)$ can be decomposed into three terms as

$$\begin{aligned} n^{1/2}(\widehat{\eta}_k - \eta_k) &= n^{1/2}\bar{N}_k(\tau)^{-1}\int_0^\tau \left\{S_k^{(0)}(\widehat{\beta},t)d\widehat{\mu}_0(\widehat{\beta},t) - S_k^{(0)}(\beta_0,t)d\widehat{\mu}_0(\beta_0,t)\right\} \\ &+ n^{1/2}\bar{N}_k(\tau)^{-1}\int_0^\tau S_k^{(0)}(\beta_0,t)\left\{d\widehat{\mu}_0(\beta_0,t) - d\mu_0(t)\right\} \\ &+ n^{1/2}\bar{N}_k(\tau)^{-1}\theta_k^{-1}\int_0^\tau \left\{S_k^{(0)}(\beta_0,t)\theta_k d\mu_0(t) - d\bar{N}_k(t)\right\}. \end{aligned} \tag{4.18}$$

A Taylor series expansion of the first term of the right-hand side of equation (4.18) equals

$$n^{1/2}\bar{N}_k(\tau)^{-1}\left\{\int_0^\tau S_k^{(1)}(\beta^\dagger,t)d\widehat{\mu}_0(\beta^\dagger,t) + \int_0^\tau S_k^{(0)}(\beta^\dagger,t)dH(\beta^\dagger,t)\right\}^T(\widehat{\beta} - \beta_0),$$

where $\beta^\dagger$ is on the line segment between $\widehat{\beta}$ and β_0. With the results in 4.8.1, condition (e) and the uniform consistency of $\widehat{\beta}$ and $\widehat{\mu}_0(\widehat{\beta},t)$, the first term of the right-hand side of equation (4.18) equals

$$(4.19)\quad n^{-1/2}o_k^{-1}\left\{\int_0^\tau s_k^{(1)}(\beta_0,t)d\mu_0(t)+\int_0^\tau s_k^{(0)}(\beta_0,t)dh(\beta_0,t)\right\}^T \times A^{-1}\sum_{i=1}^n \Psi_{1i}(\beta_0)+o_p(1).$$

With the results in (4.8.1) and condition (e), the second term of the right-hand side of equation (4.18) equals

$$(4.20)\qquad n^{-1/2}o_k^{-1}\sum_{i=1}^n\int_0^\tau s_k^{(0)}(\beta_0,t)d\Psi_{2i}(\beta_0,t)+o_p(1).$$

The third term of the right-hand side of equation (4.18) equals

$$(4.21)\qquad -n^{-1/2}o_k^{-1}\theta_k^{-1}\sum_{i=1}^n\int_0^\tau dM_{ik}(\beta_0,t)+o_p(1).$$

Combining (4.18), (4.19), (4.20) and (4.21), we see that

$$n^{1/2}(\widehat{\eta}_k-\eta_k)=n^{-1/2}\sum_{i=1}^n -\zeta_k^2\Gamma_{ki}(\beta_0)+o_p(1).$$

Therefore, $n^{1/2}(\widehat{\eta}_k-\eta_k)$ converges in distribution to a normal variable with mean 0 and variance is asymptotically $\mathcal{E}(\zeta^4\Gamma_{ki}(\beta_0)^2)$. Since $\widehat{\theta}_k=1/\widehat{\zeta}_k$, applying the delta method, we have that $n^{1/2}(\widehat{\theta}_k-\theta_k)$ is asymptotically normally distributed with mean 0 and covariance $\Sigma_k=\mathcal{E}(\Gamma_{ki}(\beta_0)^2)$.

CHAPTER V

Conclusion

This dissertation proposes three novel methods for clustered event history data, under the setting that the number of clusters are large relative to the sample size. Motivated by the real problems in kidney transplant data and hospitalization of dialysis patients data in the U.S., the three papers not only address statistical problems, but also contribute in meeting specific needs of answering clinical questions. Chapter II investigates cluster effects on the variability of failure times under the proportional hazards model. Chapter III proposes a computationally efficient proportional rates model for clustered recurrent event data with cluster-specific piecewise-constant baseline rate function, which can be fitted using standard statistical software. Chapter IV addresses the estimation of fixed center effects under a proportional rates model for recurrent event data when the number of centers is large, where the traditional estimation method might not be feasible.

The consideration of covariate-dependent frailty in Chapter II is novel in that it advocates cluster effects on the heterogeneity of the failure time distribution. Through the use of a stratified Cox model and the "bridge" property of the positive stable distribution, the proposed estimation procedure avoids the calculation of the conditional expectation of the frailties, which may not be feasible when cluster

sizes are large. Chapter III and Chapter IV target the computation difficulties for recurrent event data from a large registry study. Chapter III considers the situation when the number of distinct recurrent event times is large, i.e. high occurrence rates. By measuring the time at risk and the counts of the recurrent events within certain pre-specified intervals, calculations are carried out only at the cut-points of those intervals. This interval-grouping method reduces the data storage volume and makes the estimation feasible for recurrent event data with high occurrence rates. Chapter IV considers the situation when the number of centers is large and proposes a new estimator for fixed center effects, which performs as well as the traditional estimator, but is much more efficient in terms of computation.

Applying the method in Chapter II on the national kidney transplant data, we found that lower percentage of female patients in a facility is associated with greater heterogeneity in facility performance. For the study of hospitalization days of Medicare dialysis patients in the U.S., we collect the number of hospital days and the time at risk within intervals defined by 6 months, 1 year, 2 year2, 3 years, and 5 years since 90 days of diagnosis ESRD for each Medicare dialysis patients. Using the method in Chapter III with facility-specific baseline rate functions, we obtained the effects of gender, race, ethnicity, diabetes, BMI and whether patients stayed in nursing home in the previous calendar year on the hospitalization rates. For the same study with hospital admission as the measure of interests, we apply the method in Chapter IV and evaluate center effects for 173 dialysis facilities in Michigan.

There are several possible extensions of the methods in this dissertation. For method in Chapter II, one could develop models based on other frailty distribution. Combining the methods in Chapter III and Chapter IV, we could develop methods evaluating center effects for recurrent event data with both large number of centers

and high occurrence rates. Another possible extension is to consider the marginal model instead of partial marginal model in Chapter IV in the presence of a terminal event. The corresponding fixed center effects then evaluate centers' influence on the occurrence rate averaging over mortality experience.

BIBLIOGRAPHY

BIBLIOGRAPHY

Agresti, A. and Hartzel, J. (2000). Strategies for comparing treatment on a binary response with multicentre data. *Statistics in Medicine* **19,** 1115–1139.

Allison, P. D. (2010). *Survival Analysis Using SAS: A Practical Guide*. SAS Publishing, 2nd edition.

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

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

Chambers, J. M., Mallows, C. L., and Stuck, B. W. (1976). A method for simulating stable random variables. *Journal of the American Statistical Association* **71,** 340–344.

Cook, R. and Lawless, J. (1997). Marginal analysis of recurrent events and a terminal event. *Statistics in Medicine* **16,** 911–924.

Cook, R. and Lawless, J. (2007). *The Statistical Analysis of Recurrent events*. New York: Springer.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

Fine, J. P., Glidden, D. V., and Lee, K. E. (2003). A simple estimator for a shared frailty regression model. *Journal of the Royal Statistical Society Series B* **65,** 317–329.

Ghosh, D. and Lin, D. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica* **12,** 663–688.

Glidden, D. V. (2000). A two-stage estimator of the dependence parameter for the Clayton-Oakes model. *Lifetime Data Analysis* **6,** 141–156.

Glidden, D. V. and Self, S. (1999). Semiparametric likelihood estimation in the Clayton-Oakes failure time model. *Scandinavian Journal of Statistics* **26,** 363–372.

Glidden, D. V. and Vittinghoff, E. (2004). Modelling clustered survival data from multicentre clinical trials. *Statistics in Medicine* **23,** 369–388.

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *Annals of Statistics* **9,** 861–869.

Gould, A. (1998). Multi-centre trial analysis revisited. *Statistics in Medicine* **17,** 1779–1797.

Heagerty, P. J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics* **55,** 688–698.

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

Jacobsen, M. (1989). Existence and unicity of mles in discrete exponential family distribution. *Scandinavian Journal of Statistics* **16,** 335–349.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. New York: Wiley, 2nd edition.

Lawless, J. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *technometrics* **37,** 158–168.

Lawless, J. and Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *The Canadian Journal of Statistics* **26,** 549–565.

Lee, E. W., Wei, L. J., and Amato, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In Klein, J. P. and Goel, P. K., editors, *Survival Analysis: State of the Art*. Dordrecht: Kluwer Academic Publishers.

Lin, D., Wei, L., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society Series B* **62,** 771–730.

Lin, X., Raz, J., and Harlow, S. (1997). Linear mixed models with heterogeneous within-cluster variances. *Biometrics* **53,** 910–923.

Liu, L. and Huang, X. (2007). The use of Gaussian quadrature for estimation in frailty proportional hazards models. *Statistics in Medicine* **27,** 2665–2683.

Localio, A., Berlin, J., Ten Have, T., and Kimmel, S. (2001). Adjustments for centre in multicentre studies: an overview. *Annals of Interval Medicine* **135,** 112–123.

Martinussen, T. and Pipper, C. B. (2005). Estimation in the positive stable shared frailty cox proportional hazards model. *Lifetime Data Analysis* **11,** 99–115.

Pipper, C. B. and Martinussen, T. (2003). A likelihood based estimating equation for the Clayton-Oakes model with marginal proportional hazards. *Scandinavian Journal of Statistics* **30,** 509–522.

Prentice, R. L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association* **81,** 321–327.

Prentice, R. L. and Self, S. G. (1983). Asymptotic distribution theory for cox-type regression models with general relative risk form. *Annals of Statistics* **81,** 804–813.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton: Printceton University Press.

Schaubel, D. and Cai, J. (2005). Semiparametric methods for clustered recurrentevent data. *Lifetime Data Analysis* **11,** 405–425.

Schaubel, D. and Zhang, M. (2010). Estimating treatment effects on the marginal recurrentevent mean in the presence of a terminating event. *Lifetime Data Analysis* **16,** 451–477.

Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* **93,** 1164–1175.

Sun, J. and Kalbfleisch, J. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica* **5,** 279–290.

Sun, J. and Wei, L. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society Series B* **62,** 293–302.

Thall, P. and Lachin, J. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of the American Statistical Association* **83,** 339–347.

Wang, S. T., Klein, J. P., and Moeschberger, M. L. (1995). Semiparametric estimation of covariate effects using the positive stable frailty model. *Applied Stochastic Models and Data Analysis* **11,** 121–133.

Wang, Z. and Louis, T. A. (2003). Matching conditional and marginal shapes in binary mixed-effect models using a bridge distribution function. *Biometrika* **90,** 765–775.

Wang, Z. and Louis, T. A. (2004). Marginalized binary mixed-effects models with covariate-dependent random effects and likelihood inference. *Biometrics* **60,** 884–891.

Wassell, J. T., Kulczycki, G. W., and Moyer, E. S. (1995). Frailty models of manufacturing effects. *Lifetime data Analysis* **1,** 161–170.

Wassell, J. T. and Moeschberger, M. L. (1993). A bivariate survival model with modified gamma frailty for assessing the impact of interventions. *Statistics in Medicine* **12,** 241–248.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84,** 1065–1073.

Wellner, J. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *The Annals of statistics* **28,** 779–814.

Ye, Y., Kalbfleisch, J., and Schaubel, D. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* **63,** 78–87.