2011

New Semiparametric Methods for Recurrent Events Data

Yu Gu

THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

NEW SEMIPARAMETRIC METHODS FOR RECURRENT EVENTS DATA

By

YU GU

A Thesis submitted to the
Department of Statistics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Spring Semester, 2011

The members of the committee approve the thesis of Yu Gu defended on March 25, 2011.

Debajyoti Sinha
Professor Directing Thesis

Isaac W. Eberstein
University Representative

Dan McGee
Committee Member

Xufeng Niu
Committee Member

Approved:

Dan McGee, Chair, Department of Statistics

Joseph Travis, Dean, College of Arts and Sciences

The Graduate School has verified and approved the above-named committee members.

This dissertation is a dedication to my parents Wenxue and Jianming.

ACKNOWLEDGMENTS

Firstly I would like to thank my advisor, Dr. D. Sinha, for his endless guidance on my research, his encouragement and his support on my dissertation work at FSU. I would also like to thank Dr. S.R. Lipsitz from Harvard medical school and Dr. E.H. Slate from Medical University of South Carolina for their suggestions and help on my dissertation work. Special thanks goes to my other committee members, Dr. D. McGee, Dr. X. Niu and Dr. I. W. Eberstein.

I am also grateful for all the faculty members and all the staff in the department, for their countless support during my time at FSU.

My special thanks goes to my parents, Wenxue and Jianming, for their invaluable emotional support over the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Recurrent events data are rising in all areas of biomedical research. We present a model for recurrent events data with the same link for the intensity and mean functions. Simple interpretations of the covariate effects on both the intensity and mean functions lead to a better understanding of the covariate effects on the recurrent events process. We use partial likelihood and empirical Bayes methods for inference and provide theoretical justifications and as well as relationships between these methods. We also show the asymptotic properties of the empirical Bayes estimators. We illustrate the computational convenience and implementation of our methods with the analysis of a heart transplant study.

We also propose an additive regression model and associated empirical Bayes method for the risk of a new event given the history of the recurrent events. Both the cumulative mean and rate functions have closed form expressions for our model. Our inference method for the simiparametric model is based on maximizing a finite dimensional integrated likelihood obtained by integrating over the nonparametric cumulative baseline hazard function. Our method can accommodate time-varying covariates and is easier to implement computationally instead of iterative algorithm based full Bayes methods. The asymptotic properties of our estimates give the large-sample justifications from a frequentist stand point. We apply our method on a study of heart transplant patients to illustrate the computational convenience and other advantages of our method.

CHAPTER 1

INTRODUCTION AND MOTIVATING EXAMPLE

1.1 Introduction

In all kinds of areas of science and technology, a lot of research involves processes that generate events repeatedly over time. This kind of processes is often called recurrent events processes and the data that are generated by this process are called recurrent events data. Recurrent events data arise in medical studies frequently, for example, the occurrence of asthma attacks in respirology trials, mammary tumors in a carcinogenicity study, and pulmonary exacerbations in cystic fibrosis. In this circumstance, we have information from different subjects about the clinical events that have been experienced by them repeatedly over time. Usually, the number of events in each process is relatively smaller than that in the process from other areas, such as the incidence of injuries in manufacturing plants.

The special features of recurrent event processes lead to different ways to model this kind of data. Event counts and waiting (or gap) time can both be used to analyse the data. Some commonly used models can be found in [7]. In many settings, it is of interest to characterize the incidence of recurrent events associated with an event which terminates the recurrent event process. Many examples can be found especially in settings that involve patients with a serious disease which is related both recurrent complications and high mortality. For example, in oncology, interest may lie in characterizing the use of health services following diagnosis of cancer, but use of this kind of services terminates upon death. This will complicate the problem and for simplicity termination can be thought of as a second type of event. This type of recurrent event data can be handled with intensity-based model relatively easily.

However, analyses are less straightforward when the termination is dependent on the recurrent events.

1.2 Heart Transplant Data

The heart transplant data from the Medical University of South Carolina contains patients who received heart transplant between January 1992 and May 2007. Cardiac transplant recipients may experience non-fatal graft rejections after the transplant. Usually they can be treated by drug therapy. However, a high frequency of graft rejections can be related to the underlying morbidity of the patient and also will affect the quality of the patients' lives. Endomyocardial biopsy (EMB) is very important to detect clinically significant rejections early for patients after cardiac transplantation ([17]). On the other hand, there are procedural related complications, increased risk of infection and patient discomfort that accompany the biopsies ([40]). The opinion for frequent EMB beyond first year and long term surveillance EMB still remains in question. By modeling the intensity of new rejection through covariate x such as race and gender and observed/past rejection history, we can explain the effect of covariate x on recurrent events $N(t)$, which will help us identify patient group with high risk of rejection. We also want to understand the risk of graft rejections, and give prediction of a patient's risk of graft rejection given the history of rejections. Knowing this will help the clinicians to schedule patients' biopsy given the history of graft rejections. Another goal of this study is to find out the effect of the covariates on the risk of recurrent graft rejections.

The terminating event is either death or acute rejection (which is soon followed by death). The termination (or can be called "death") is censored by non-informative events, for example drop-out. There are about 65% of these patients died over the follow-up period. The covariates we are interested in are the race of the patients (0 for Caucasian and 1 for African American) and the gender of the patients (0 for male and 1 for female). Table 1.1 describes the distribution of the four groups in the data set. The number of female patients or African American patients is relatively small comparing to the number of white male patients, so in the analysis we consider combining some of the groups based on the their similar characteristics to get a more accurate estimate.

Figure 1.1 provides a graphical description of the recurrent events and termination

Table 1.1: Description of the data set.

	male	female	Total
white	68	13	81
African American	16	8	24
Total	84	21	105

time for different races. The crosses are the graft rejections for each patient. Different symbols are used to denote the termination times for patients with different covariates. The circles are for white patients and the triangles are for African American patients. For those patients who experienced death, the symbols are solid and for those who were censored, the symbols are hollow.

69 out of the total 105 patients have at least one rejection, and 40 out of these 105 patients have at least two rejections. We plotted the Kaplan-Meier estimate of the survival function of the time to the first rejection and the time to the second rejection for groups with different covariate values in Figure 1.2 and Figure 1.3. Based on both figures, white male patients have longer survival time for both first and second rejections than the other three groups, which means their risk of getting new events is lower. We confirmed this observation with our analysis results later.

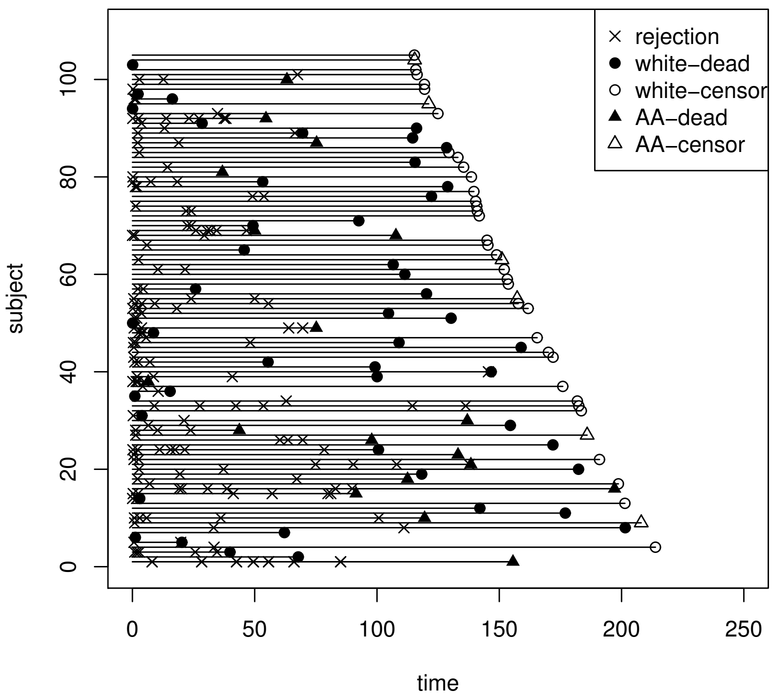

Figure 1.1: Plot of the recurrent events and termination times for the heart transplant data.

1.3 Existing methods

Recurrent events data has been found in many areas of applications in medical studies, including the recurrence of asthma attacks in respiratory trials, recurrent infection episodes in HIV studies and multiple recurrences of tumors. Usually, researchers are interested in investigating the effects of an external covariate x on the recurrent events process $\{N(\cdot)\}$, where $N(t)$ is the cumulative number of recurrent events to a subject by time t and $dN(t) = N(t + dt) - N(t)$ is the increment of $N(t)$ in the small interval $(t, t + dt)$. For recurrent events data, the covariate effects are usually explained via the regression functions of essentially three quantities: (1) mean number of recurrences $\mu(t|x) = E[N(t)|x]$ (e.g., [26]), (2) rate function $r(t|x)dt \simeq P[dN(t) = 1|x]$ of the recurrence unconditional on the observed history $\mathcal{N}(t-) = \{N(u) : u < t\}$ (e.g., [30], [24], [26], [32] and [31]), (3) intensity/risk function $h(t|\mathcal{N}(t-), x) \simeq P[dN(t) = 1|\mathcal{N}(t-), x]$ of a new event given history $\{N(\cdot)\}$

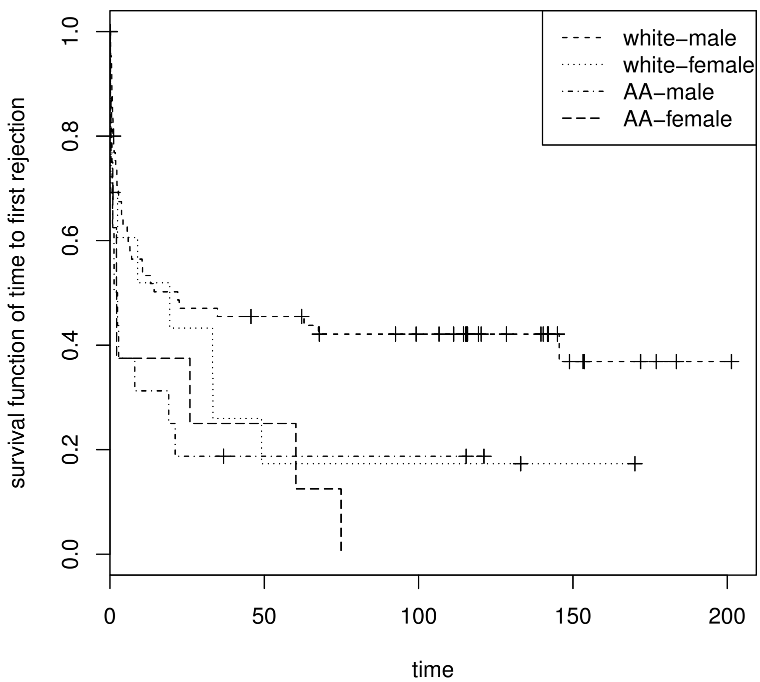

Figure 1.2: Plot of the Kaplan-meier survival curve of the time to the first rejection for heart transplant patient groups with different covariate values.

(e.g., [4], [7]). For fixed covariate x, we further have $r(t|x) = \mu'(t|x) = \frac{d}{dt}\mu(t|x)$.

The present frequentist literature on recurrent events data has put high emphasis on partially specified models where the interest is restricted to robust estimation of the covariate effect on the cumulative rate (or on the mean function) which is not conditioned on event history. Let $N_i(t)$ denote the number of rejections by time point t. Let $\mathcal{N}(t)$ denote the history of point process of rejections $N_i(t)$ and $\mathcal{H}_{T_i}(t-)$ denote the history of termination (or death). Based on the assumption that the termination is independent of recurrent events, several papers proposed marginal methods to model the recurrent event process $\{N(t)\}$ conditional on part of the combined history $\mathcal{H}_p(t-) \subseteq \mathcal{H}^* = (\mathcal{N}(t-), \mathcal{H}_T(t-); x)$. [42] proposed a proportional hazard structure to estimate $E[dN_{(k)}(t)|\mathcal{N}_{(k)}(t-); x]$, the intensity function of k^{th} rejection time $Y_{(k)}$, given the partial history of whether $Y_{(k)} < t$. Under the assumption of these intensity functions, it is possible for a subject to be at risk of experiencing the k^{th} event without experiencing the $(k-1)^{th}$ event. There are some other models based on

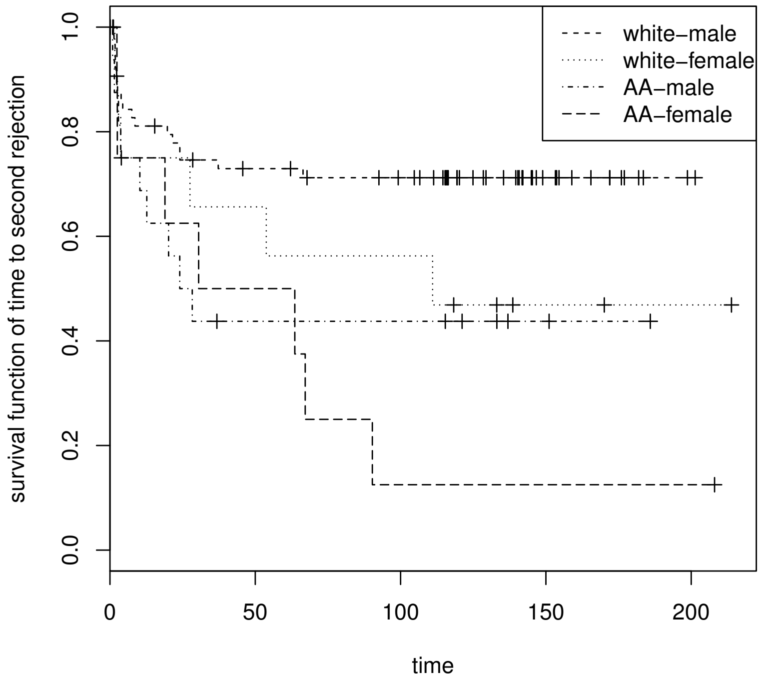

Figure 1.3: Plot of the Kaplan-meier survival curve of the time to the second rejection for heart transplant patient groups with different covariate values.

partial history are proposed later, for example in [30], $E[dN_{(k)}(t)|N(t-) = k-1; x]$ was estimated through semiparametric method for the regression parameter, and in [26], $E[dN(t)|T > t; x]$ was estimated by maximizing the partial likelihood function. Similar method was established by [25] by modeling the mean function $E[N(t)|T > t; x]$ and obtaining point estimate using Poisson models and robust variance estimate. These analyses using partially specified models lack transparency because they do not give the physical description of the underlying process, and they cannot be used for direct subject-specific prediction of the recurrent events, which is an important goal of some studies. Because these models are only based on partial history, the regression models are different for different $\mathcal{H}_p(t-)$. Therefore, the parameter estimates will be quite different for different models, which will make the interpretation even more difficult. Because of all the differences, the compatibility of these models is another important issue for researchers analysing the same data set. With only part of the history considered in the model, any prediction using the parameter estimates from the model would only use part of the information that the data contains, which will introduce bias into the prediction that is based on these models.

When the regression model is based on either the rate function $\mu'(t|x)$ or the cumulative mean function $\mu(t|x)$, which is unconditional on the observed events history $\{N(\cdot)\}$, the model cannot be used for prediction and dynamic assessment of the risk of recurrence. On the other hand, the interpretation of the covariate effect only on the intensity $h(t|\mathcal{N}(t-), x)$ (conditional on the history of past events) may not clearly explain the effects of x because it affects the history $\mathcal{N}(t-)$ as well as the future event $\{dN(t) = 1\}$. Additionally, an interpretable regression function defined on one quantity, say the mean function $\mu(t|x)$, may not give a closed form and interpretable regression function for the intensity function $h(t|\mathcal{N}(t-), x)$ (or other functions of interest). See the recent book by [7] for a very up to date review about the goals, limitations and challenges of regression modeling of recurrent events process.

One popular approach for the recurrent event process is the proportional intensity model $h(t|\mathcal{N}(t-), x) = h_0(t) \exp(\beta x)$ of [4], an extension of Cox's proportional hazard model ([9]). Under this model, the events that occur to the same subject at different time points are independent of each other after adjusting for the effects of the covariate vector $x(t)$. Also, this model does not allow for heterogeneity among subjects. [29] among others introduced subject-specific frailty random effects to Andersen-Gill's model (essentially a non-homogeneous Poisson process model) to account for the variation among subjects and the effects of number of past events $N(t-)$. A proportional intensity function with subject-specific random effect (frailty) w takes the form

$$\lambda(t|x(t), \mathcal{H}(t-), w) = \lambda_0(t) w e^{\beta x(t)}, \tag{1.1}$$

where $\mathcal{H}(t-) = \{N(s), x(s) : s < t\}$ is the history of number of recurrent events $N(s)$. The variance of w represents the variation among the subjects and the intra-subject association among event times is taken into consideration through the subject-specific frailty w. [34] presented the Bayesian analysis of recurrent events data using the proportional intensity model with frailty. However, the unknown frailty in these models can lead to computational difficulties (for semiparametric maximum likelihood and Bayes estimations). Empirical Bayes methods have potential for computational convenience, however, such methods encounter theoretical difficulties, such as difficulty in proving the asymptotic properties for frailty models.

Another approach is to extend the Anderson-Gill model with either specifying the conditional intensity (hereafter, just 'intensity' for short) $h(t|\mathcal{N}(t-), x) = h(t|z(t))$, where $z(t)$ is a function of x and $\mathcal{N}(t-)$ ([1], [28]). An important example of such a model is the modulated intensity model by [29] with $h(t|\mathcal{N}(t-), x) = h(t|N(t-) = j, x) = (1 + \eta j) \exp(\beta x) h_0(t)$. Such a model has interpretational problems for understanding the cause-effect relationship between x and $dN(t)$ because $z(t)$ is not an 'exogenous' covariate. One way to 'isolate' the effects of x from its effects via history is to also consider the mean function $\mu(t|x)$ corresponding to the assumed intensity $h(t|\mathcal{N}(t-), x)$. However, there is no guarantee of a simple interpretation of marginal effects of x on $\mu(t|x)$ corresponding to every $h(t|\mathcal{N}(t-), x)$. For example, the intensity model in [29] has the corresponding $\mu(t|x) = \eta^{-1}[\exp\{\eta H_0(t) \exp(\beta x)\} - 1]$, where $H_0(t) = \int_0^t h_0(u) du$ ([25]). The same problem exists for regression functions of the 'conditional' rate function $P[dN(t) = 1|N(t-) = j, x] \simeq \mu'(t|j, x)dt$. The difficulty in understanding the covariate effects while focusing on the regression parameters of only one of such functions have been recognized by multiple authors (e.g., [7]).

The multiplicative assumption of the covariate effect may not be satisfied in some practical situations. Some alternative non-multiplicative models are needed to analyze recurrent events data. For the semiparametric transformation model with random effects of [46], the cumulative intensity function for $N(t)$, the number of events that has been experienced by a subject by time t, is given as

$$E[N(t)|X(t), Z(t), b] = \Lambda(t|X(t), Z(t); b) = G\Big(\int_0^t \lambda(s) e^{\beta x(s) + b z(s)} ds\Big),$$

where β is the unknown regression parameters, b is the vector of random effects, $X(t) = \{x(s) : s \leqslant t\}$ and $Z(t) = \{z(s) : s \leqslant t\}$ are the entire paths of the corresponding covariate processes $x(s)$ and $z(s)$ respectively, $\lambda(\cdot)$ is an arbitrary positive function and $G(\cdot)$ is a transformation function with $G(0) = 0$ and $G(\infty) = \infty$. When $G(x) = x$, it is a proportional cumulative intensity model with random effects. This model is focused on estimating the cumulative intensity instead of the risk of a new event the given history $\mathcal{H}(t)$. The covariate effects on the risk of a new event given the history $\mathcal{H}(t)$ do not have a clear interpretation for this transformation model. Misspecification of the transformation would cause incorrect characterization of the covariate effect and biased prediction of the occurrence of the events over time.

Additive rate models have been studied by [32] and [31] using semiparametric and nonparametric methods. However, models based on the rate function have their own disadvantages. Since they are not fully specified, it is very difficult to understand the risk of a new event given the history.

1.4 Goals for Modeling

The data set with recurrent events contains a lot of information about the process $\{N(t)\}$. Therefore, our goals for modeling this kind of data set should also take all the aspects into consideration. To understand the effect of all the covariates and past history on mean number of future events, and to predict the future occurrences of events, we want to propose a method that can provide estimates of all of these functions (some are unconditional and others are conditional on history of events). $N(t)$ is a crucial measure of quality of life and health, so explanation of covariate effect on recurrent events $N(t)$ and termination time T is very important. Given the estimates, we also want to predict the risk of recurrent events and use this information for future prediction. Prediction of the cumulative mean function given past history can be very helpful for the patients and clinicians to adjust treatment. We proposed a semiparameter empirical Bayes model in Chapter 2. Closed form expression is available for the cumulative mean function, which also provides a nice and simple interpretation for the covariate effects.

The model we proposed in Chapter 2, as well as most of the existing models for recurrent events data, has a multiplicative structure for either the intensity function or the mean/rate function. We need an alternative class of models when this assumption can not be satisfied. Therefore, we proposed a nonmultiplicative intensity model for recurrent events in Chapter 3, which also has a closed form expression for cumulative mean function. We showed the asymptotic properties of the empricial Bayes estimators for both of our models. Both our models can be extended to accommodate time-varying covariates to some extend.

With wide range of models existing for recurrent events data, model comparisons and diagnostics methods are a very important issue. For this kind of data, model comparison is very complicated because many factors need to be considered, for example the frequency and the timing of the events. We developed a model comparison method for survival data and extended it to the recurrent events data.

CHAPTER 2

MODEL FOR RECURRENT EVENTS DATA WITH SAME LINK FOR INTENSITY AND MEAN FUNCTIONS

2.1 Proposed Model

For the time being, we only consider time-constant covariate x. We relax this assumption in Section 2.5. Our model assumes the process $\{N(t)\}$ has an intensity function

$$\lim_{dt \to 0} \frac{P[dN(t) = 1|\mathcal{N}(t-), x_i]}{dt} = h_{j+1}(t|x_i, N(t-) = j) = \lambda_0(t)(j\eta + \theta(x_i\beta)), \quad (2.1)$$

where $\theta(\cdot)$ is the link function of the covariate x with regression parameter β, $\lambda_0(t)$ is the baseline hazard function, and η is the effect of the previous number of events $N(t-) = j$ on the risk of a new event $\{dN(t) = 1\}$. This model allows a straightforward interpretation of the effects of $N(t-) = j$ and x on the prediction of $dN(t)$. It assumes Cox's relative risk model $h(t|x) = \lambda_0(t)\theta(\beta x)$ for the time to the first event. For any subsequent past recurrence, it adds an additional intercept term η to the risk of $dN(t)$; we discuss at length later in this section that η can be positive or negative. For any subject, the risk of new event increases by $\lambda_0(t)\eta$ amount for every past recurrence to the same subject. Unlike previous models with exponentiated link for x, we are not assuming at this point any particular link for x. Some examples of $\theta(x)$ can be linear relative risk with $\theta(x) = 1 + \beta x$ and exponentiated risk with $\theta(x) = \exp(\beta x)$.

Our new regression model of (2.1) has an expression similar to $h(t|\mathcal{N}(t-), x) =$

$\lambda_0(t)(1 + \eta N(t-)) \exp(\beta x)$ of [29]. However, the formulation of our model ensures some advantages for our model compared to Oakes' model. Based on the conditional expectation $E[dN(t)|\mathcal{N}(t-); x] = d\Lambda_0(t)(N(t-)\eta + \theta)$, we obtain a differential equation for the cumulative mean function $\mu(t|x) = E[N(t)|x]$. Solution of the differential equation is

$$\mu(t|x) = \frac{\theta}{\eta}(e^{\eta\Lambda_0(t)} - 1) \tag{2.2}$$

where $\theta = \theta(\beta x)$. The equation in (2.2) has a simple closed-form expression with a multiplicative structure for the covariate effect on the mean function $\mu(t|x)$. Again, the link θ of x has many choices including additive with $\theta(\beta x) = 1+\beta x$ and exponentiatial with $\theta(\beta x) = \exp(\beta x)$. The cumulative mean function $\mu(t|x)$ has form similar to the hazard of the first event except with different baseline function. When η equals to 0, our model is similar to the Anderson-Gill model with $E[N(t)|x] = \theta\Lambda_0(t)$ and with no effect of previous number of events on the risk of $dN(t) = 1$. By differentiating $\mu(t|x)$, we can obtain the rate function

$$\mu'(t|x) = \theta\lambda_0(t)\ e^{\eta\Lambda_0(t)}\ , \tag{2.3}$$

which also has a multiplicative structure. Our model allows the β and η to be negative as long as $\theta(\beta x) + \eta N(t-) > 0$ for all possible values of x and $N(t-)$. For example, with $\theta(\beta x) = \exp(\beta x)$, a scalar bounded $x \in (-1, +1)$ and maximum allowable number of recurrences $N(t) < K$ for all t, we have $\eta \in (-e^{-|\beta_0|}/K, \infty)$, where β_0 is the true value of β. We can get a similar bound for β for any known true value of $\eta = \eta_0$. This places the pair $(\eta = 0, \beta = 0)$ inside the parameter space of (β, η). Later this will allow us to develop the asymptotic properties which are valid even when $(\eta, \beta) = (0, 0)$.

Our model ensures that the corresponding regression model for mean function $\mu(t|x)$ (as a consequence $\mu'(t|x)$) has the interpretation of the effects of x similar to the relative risk of the first event. For our example discussed in Section 4, a heart transplant study, we are interested in the effect of, say, gender on the process of recurrent graft-versus-host disease (GVHD). When our model is appropriate for such a study, the effect of gender on the risk of first GVHD is the same as the effect of gender on the mean number of GVHD since the origin/entry.

One method of analysis for such a semiprametric model is through a Bayesian paradigm using a prior process, say, a gamma process of [23], with possibly unknown hyperparameters, for the unknown baseline function $\Lambda_0(t)$. In section 2.2, we derive the finite-dimensional integrated likelihood of this model after integrating out the unknown function $\Lambda_0(t)$. We also show that this finite-dimensional integrated likelihood can be used either for estimation via empirical Bayes method or for deriving a partial likelihood as a limiting case of the non-informative prior belief. Both of these methods are very easy to implement computationally in statistical or mathematical software. In section 2.3, we present the empirical Bayes estimate and the standard error of the cumulative mean function. A study of recurrent minor graft-versus-host (GVH) episodes for a group of heart transplant patients is analyzed to illustrate the convenience of our method. In the end, we show the asymptotic properties of the semiparametric empirical Bayes estimates.

2.2 Integrated Likelihood

Throughout this paper, we assume that the recurrent event processes are subject to only non-informative censoring/termination. Let $N_i(t)$ denote the number of recurrent events by time t for subject $i = 1, \ldots, n$. Let $0 = y_0 < y_1 < \ldots < y_M$ denote the ordered distinct recurrent event times and censoring times (maximum time under observation) for all subjects. For the sake of convenience, we first assume there are no ties among all the observed event times across all the subjects. Let N_{ij} be the indicator variable for the occurrence of an event for subject i at time y_j. Based on our assumption, $\sum_{i=1}^{n} N_{ij} = 1$ when there is an event occurring at y_j. Let $R_j = R(y_j)$ be the set of all the subjects under observation at time point y_j. The observed data is given by $D = \{N_i(t) : 0 < t \leqslant T_i;\ x_i;\ T_i : i = 1, \ldots, n\}$, where T_i is the censoring/termination time for subject i.

One method for inference about (β, η) is based on the partial likelihood (similar to Cox's partial likelihood ([10])), which is

$$L_{PL}(\beta, \eta | D) \propto \prod_{i=1}^{n} \prod_{j=1}^{M} \Big\{ \frac{n_{ij}\eta + \theta_i}{\sum_{l \in R_j} (n_{lj}\eta + \theta_l)} \Big\}^{N_{ij}}, \tag{2.4}$$

where $\theta_i = \theta(x_i\beta)$ and $n_{ij} = N_i(y_j-)$ is the number of events that have occurred to subject i before y_j. We will later show that this partial likelihood of (2.4) is a limiting marginal posterior after integrating out the cumulative baseline function $\Lambda_0(t)$ from the joint posterior under a very diffuse/non-informative prior. This gives a theoretical justification as well as an empirical Bayes (or an integrated likelihood) justification of (2.4).

The full likelihood of (β, η, Λ_0) based on D is

$$L(\beta, \eta, \Lambda_0|D) = \prod_{i=1}^{n}\prod_{j=1}^{M} e^{[-\delta_{ij}\{\Lambda_{0j}(n_{ij}\eta+\theta_i)\}]}\{d\Lambda_0(y_j)(n_{ij}\eta + \theta_i)\}^{N_{ij}}, \qquad (2.5)$$

where $\Lambda_{0j} = \Lambda_0(y_j) - \Lambda_0(y_{j-1})$ and , and $\delta_{ij} = 1$ when $i \in R_j$. There are two factors in the likelihood of (2.5) for each observed event time y_j. The first is the likelihood contribution from the interval $[y_j, y_j + dy_j)$. When $\sum_{i=1}^{n} N_{ij} = 1$, this likelihood contribution is

$$\sum_{i=1}^{n} N_{ij}\{(n_{ij}\eta + \theta_i)d\Lambda_0(y_j)\}\exp[-\sum_{i\in R_j}\{(n_{ij}\eta + \theta_i)d\Lambda_0(y_j)\}].$$

Otherwise, the likelihood contribution is 1. The second factor is the likelihood contribution from the interval (y_{j-1}, y_j) and it is approximately equal to

$$\prod_{i=1}^{n}\exp[-\delta_{ij}\{(n_{ij}\eta + \theta_i)\Lambda_{0j}\}].$$

A full Bayesian procedure uses the full likelihood (2.5) as well as the prior $\pi(\beta, \eta, \Lambda_0)$. For semiparametric Bayesian analysis (see [21]), one uses a prior process $\pi_1(\Lambda_0)$ for the nonparametric baseline function and an independent parametric prior $\pi_2(\beta, \eta)$ for the parameters. One option is to assign a Gamma process $GP(\Lambda^*(t), \alpha)$ of [23] as the prior $\pi_1(\Lambda_0(t)|\Lambda^*, \alpha)$. The mean of the prior process is $\Lambda^*(t) = \int_0^t \lambda^*(u)du$, and the known prior precision parameter is α. A joint posterior density $p(\beta, \eta, \Lambda_0|D, \alpha, \lambda^*)$, which is a product of the likelihood of (2.5) along with the joint prior $\pi(\beta, \eta, \Lambda_0) \propto \pi_1(\Lambda_0|\Lambda^*, \alpha)\pi_2(\beta, \eta)$, is used for full Bayesian analysis of recurrent events data. A full Bayesian procedure requires a full specification of the mean function $\Lambda^*(t)$. As a very practical alternative, an empirical Bayesian procedure (hereafter, integrated likelihood method) only requires the mean of the prior process $\Lambda^*(t|\lambda^*)$ to be parametric with some unknown hyper-parameter λ^*. The precision parameter α describes our prior belief of how close the unknown $\Lambda_0(\cdot)$ is to the prior guess, the parametric function $\Lambda^*(\cdot)$. The method is based on the finite dimensional integrated likelihood $L(\beta, \eta, \lambda^*|D)$ derived via integrating the joint posterior $p(\beta, \eta, \Lambda_0|D, \alpha, \lambda^*)$ with respect to the unknown Λ_0. See [37] and the references therein for integrated likelihood method for additive hazards model for univariate survival data.

After integrating out the increments $d\Lambda_0(t)$ with respect to their independent $Ga(\alpha\Lambda^*, \alpha)$ prior densities, the marginal posterior based on the likelihood of (2.5) is proportional to

$$
\begin{aligned}
L(\beta, \eta, \Lambda^*|D) &= \int L(\beta, \eta, \Lambda_0|D)\pi(\Lambda_0|\Lambda^*, \alpha)d\Lambda_0 \\
= & \prod_{j=1}^{M}\left[\Big(\frac{\alpha}{\alpha + B_{2j}\eta + B_{1j}}\Big)^{\alpha\Lambda_j}\prod_{i=1}^{n}\Big\{\frac{(n_{ij}\eta + \theta_i)\lambda^*(y_j)\alpha}{\alpha + B_{2j}\eta + B_{1j}}\Big\}^{N_{ij}}\right], \qquad (2.6)
\end{aligned}
$$

where $\Lambda_j = \Lambda^*(y_j) - \Lambda^*(y_{j-1})$, $B_{1j} = \sum_{i\in R_j}\theta_i$, $B_{2j} = \sum_{i\in R_j} n_{ij}$, and $\lambda^*(t) = \frac{d}{dt}\Lambda^*(t)$.

Since we have

$$
\lim_{\alpha\to 0}\Big(\frac{1}{\alpha + B_{2j}\eta + B_{1j}}\Big)^{\alpha\Lambda_j} = 1, \qquad (2.7)
$$

we can get

$$
\lim_{\alpha\to 0}\frac{L(\beta, \eta, \Lambda^*|D)}{\alpha^{\alpha\Lambda_j + \sum_{i,j}N_{ij}}\lambda^*} = \prod_{i=1}^{n}\prod_{j=1}^{M}\Big\{\frac{n_{ij}\eta + \theta_i}{\sum_{l\in R_j}(n_{lj}\eta + \theta_l)}\Big\}^{N_{ij}}, \qquad (2.8)
$$

where $\lambda^* = \prod_{i,j}\lambda^*(y_j)^{N_{ij}}$ is free of β and η. The right-hand side of equation (2.8) gives us the partial likelihood in (2.4) of the conditional intensity model (2.1). This provides us the empirical Bayes justification of the partial likelihood in (2.4) as a marginal posterior of (β, η) under a very non-informative prior process (as a limit) for Λ_0 and a non-integrable 'flat' prior for (β, η). This is an extension of the works of [23] and [35] for univariate survival data. This also justifies the use of the product of (2.4) and the prior $\pi_2(\beta, \eta)$ of the parameters as a marginal posterior (as a limit of the prior precision going to zero) of only the finite-dimensional parameters (β, η). This can allow a Bayesian to avoid difficult elicitation of the prior process when there is a lack of meaningful prior opinion about the baseline function.

When we let $\alpha \to \infty$, we have

$$
\lim_{\alpha\to\infty}\left(\frac{\alpha}{\alpha + B_{2j}\eta + B_{1j}}\right)^{\alpha\Lambda_j} = \exp\{-(B_{2j}\eta + B_{1j})\Lambda_j\}, \qquad (2.9)
$$

and therefore

$$\lim_{\alpha\to\infty} L(\beta,\eta,\Lambda^*|D) = \prod_{j=1}^{M}\Big[\exp\{-(B_{2j}\eta + B_{1j})\Lambda_j\}\prod_{i=1}^{n}\Big\{(n_{ij}\eta+\theta_i)\lambda^*(y_j)\Big\}^{N_{ij}}\Big]. \quad (2.10)$$

We can see that the right-hand side of equation (2.10) is the parametric likelihood function of (β, η, Λ^*) based on the conditional intensity model (2.1) with $\lambda_0(t) = \lambda^*(t)$ (parametric baseline function).

We can obtain both an empirical Bayes estimate and partial likelihood estimate. Assuming an unknown constant λ as the prior mean of the baseline function, the empirical Bayes estimates $(\hat{\beta}, \hat{\eta}, \hat{\lambda})$ can be obtained by maximizing the integrated likelihood function in equation (2.6). The estimated variance-covariance matrix V of $(\hat{\beta}, \hat{\eta}, \hat{\lambda})$ is the inverse of the observed information matrix based on the integrated likelihood. To obtain the partial likelihood estimates $(\hat{\beta}_{PL}, \hat{\eta}_{PL})$, we can maximize the partial likelihood function on the right-hand side of equation (2.8). The estimated variance-covariance matrix V of $(\hat{\beta}_{PL}, \hat{\eta}_{PL})$ is the inverse of the observed information matrix based on the partial likelihood. The standard error of the estimates can be used to conduct the hypothesis testing for β and η and obtain interval estimates.

2.3 The Posterior Process and Estimator

We can obtain the empirical Bayes estimator of the cumulative mean function $\mu(t|x) = E[N(t)|x]$ in equation (2.2) through the posterior process $p(\Lambda_0|\beta, \eta, \lambda, D)$. A closed form expression of the Laplace transform of the posterior process $p(\Lambda_0|\beta, \eta, \lambda, D)$ can be derived, which can be used to obtain the empirical Bayes estimator of the cumulative mean function. In particular, we discuss the empirical Bayes estimator

$$E\{\mu(t|x)|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\} == \frac{\hat{\theta}}{\hat{\eta}}(E[\exp\{\eta\Lambda_0(t)\}|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}] - 1)\ .$$

First of all, we put a grid $0 = t_1 < t_2 < \ldots < t_{M+1} < \infty$ on the time so that $t_j \in (y_{j-1}, y_j)$. As discussed earlier, $0 = y_0 < y_1 < y_2 < \ldots < y_M < \infty$ are the ordered distinct recurrent event times and censoring times for all subjects. We let $u_{j1} = \Lambda_0(y_j) - \Lambda_0(t_j)$, $u_{j2} = \Lambda_0(t_{j+1}) - \Lambda_0(y_j)$, and $u_j = \Lambda_0(t_{j+1}) - \Lambda_0(t_j) = u_{j1} + u_{j2}$ denote the increments of $\Lambda_0(t)$ generated by the grid and y_j's. Also, we let $\lambda_{j1} = \Lambda^*(y_j) - \Lambda^*(t_j)$, $\lambda_{j2} = \Lambda^*(t_{j+1}) - \Lambda^*(y_j)$, and $\lambda_j = \Lambda^*(t_{j+1}) - \Lambda^*(t_j) = \lambda_{j1} + \lambda_{j2}$ denote the increments of the prior mean $\Lambda^*(t)$. Given this new notation, we can write the full likelihood function of $(\beta, \eta, u_1^*, u_2^*)$ as

$$L(\beta, \eta, u_1^*, u_2^*|D) =$$
$$\prod_{j=1}^{M} \left[e^{-\{u_{j1}(B_{2j}\eta + B_{1j}) + u_{j2}(B_{2,j+1}\eta + B_{1,j+1})\}} \prod_{i=1}^{n} \{d\Lambda_0(y_j)(n_{ij}\eta + \theta_i)\}^{N_{ij}} \right],$$

where $\theta_i = \theta(\beta x_i)$, $u_1^* = (u_{11}, \ldots, u_{M1})$ and $u_2^* = (u_{12}, \ldots, u_{M2})$. The joint distribution of the observed data vector $D = \{N_i(t) : 0 < t \leqslant T_i;\ x_i;\ T_i : i = 1, \ldots, n\}$ and $u^* = (u_1, u_2, \ldots, u_{M+1})$ is given by

$$L(\beta, \eta|u^*, D) = E_{u_1^*, u_2^*}\{L(\beta, \eta, u_1^*, u_2^*|D)\},$$

where the expectation is taken over the prior distribution of u_{j1}, u_{j2}, for $j = 1, \ldots, M$. Note that u_{j1} and u_{j2} have independent gamma distribution based on the Gamma prior process.

We can obtain the expression of $E\{\exp(-\sum_{j=1}^{M} h_j u_j)|D, \beta, \eta, \lambda\}$ which is given below:

$$E\{\exp(-\sum_{j=1}^{M} h_j u_j)|D, \beta, \eta, \lambda\} \tag{2.11}$$
$$= \prod_{j=1}^{M} \Big\{ \Big(\frac{\alpha}{\alpha + B_{2j}\eta + B_{1j} + h_j} \Big)^{\alpha\lambda_{j1}} \Big(\frac{\alpha}{\alpha + B_{2,j+1}\eta + B_{1,j+1} + h_j} \Big)^{\alpha\lambda_{j2}} \Big\} \times$$
$$\prod_{j=1}^{M} \prod_{i=1}^{n} \left\{ \frac{\lambda^*(y_j)(n_{ij}\eta + \theta_i)\alpha}{\alpha + B_{2j}\eta + B_{1j} + h_j} \right\}^{N_{ij}}.$$

We can obtain the Laplace transform of the posterior process $p(\Lambda_0|\beta, \eta, \lambda, D)$ by dividing $E\{\exp(-\sum_{j=1}^{M} h_j u_j)|D, \beta, \eta, \lambda\}$ by the marginal density of the observed data in equation (2.6). It can be shown that the Laplace transform of the posterior process has the following closed form expression,

$$E\{\exp(-\textstyle\sum_{j=1}^{M} h_j u_j)|D\} = \prod_{j=1}^{M} \Big\{ \Big(\frac{\alpha + B_{2j}\eta + B_{1j}}{\alpha + B_{2j}\eta + B_{1j} + h_j} \Big)^{\alpha\lambda_{j1} + \sum_i N_{ij}} \Big(\frac{\alpha + B_{2,j+1}\eta + B_{1,j+1}}{\alpha + B_{2,j+1}\eta + B_{1,j+1} + h_j} \Big)^{\alpha\lambda_{j2}} \Big\}. \tag{2.12}$$

Given the expression in (2.12), we can obtain the Laplace transform of the posterior process for any $t \in (y_k, y_{k+1})$ as

$$E[\exp\{-h\Lambda_0(t)\}|D] = \left(\frac{\alpha + B_{2,k+1}\eta + B_{1,k+1}}{\alpha + B_{2,k+1}\eta + B_{1,k+1} + h}\right)^{\alpha\lambda_{0k}} \prod_{j=1}^{k}\left\{\left(\frac{\alpha + B_{2j}\eta + B_{1j}}{\alpha + B_{2j}\eta + B_{1j} + h}\right)^{\alpha\lambda_{yj}+\sum_i N_{ij}}\right\}, \quad (2.13)$$

where $\lambda_{0k} = \Lambda^*(t) - \Lambda^*(y_k)$ and $\lambda_{yj} = \Lambda^*(y_j) - \Lambda^*(y_{j-1})$.

The posterior moments of functions of $\Lambda_0(t)$ can be found using equation (2.13) as follows,

$$E[\exp\{\eta\Lambda_0(t)\}|D, \beta, \eta, \lambda] = \left(\frac{\rho_{k+1}}{\rho_{k+1} - \eta}\right)^{\alpha\lambda_{0k}} \prod_{j=1}^{k}\left(\frac{\rho_j}{\rho_j - \eta}\right)^{\alpha\lambda_{yj}+\sum_i N_{ij}}, \quad (2.14)$$

$$E[\exp\{\eta\Lambda_0(t)\}^2|D, \beta, \eta, \lambda] = \left(\frac{\rho_{k+1}}{\rho_{k+1} - 2\eta}\right)^{\alpha\lambda_{0k}} \prod_{j=1}^{k}\left(\frac{\rho_j}{\rho_j - 2\eta}\right)^{\alpha\lambda_{yj}+\sum_i N_{ij}}, \quad (2.15)$$

where $\theta = \theta(\beta x)$, $B_{1j} = \sum_{i \in R_j} \theta_i$, and $\rho_j = \alpha + B_{2j}\eta + B_{1j}$. The formula above can be used to compute the empirical Bayes estimator of the cumulative mean function $\mu(t|x) = E\{N(t)|x\} = \frac{\theta}{\eta}(e^{\eta\Lambda_0(t)} - 1)$ and $\mu(t|x)^2$ as

$$E\{\mu(t|x)|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\} = \frac{\hat{\theta}}{\hat{\eta}}(E[\exp\{\eta\Lambda_0(t)\}|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}] - 1), \quad (2.16)$$

$$E\{\mu(t|x)^2|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\} = \left(\frac{\hat{\theta}}{\hat{\eta}}\right)^2(E[\exp\{\eta\Lambda_0(t)\}^2|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}] - 2E[\exp\{\eta\Lambda_0(t)\}|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}] + 1), \quad (2.17)$$

where the terms $E[\exp\{\eta\Lambda_0(t)\}|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}]$ and $E[\exp\{\eta\Lambda_0(t)\}|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}]$ in (2.16) and (2.17) are obtained by replacing (β, η, λ) in (2.14) and (2.15) with empirical Bayes estimators $(\hat{\beta}, \hat{\eta}, \hat{\lambda})$. We can estimate the variance of the empirical Bayes estimator $\mu(\hat{t}|x) = E\{\mu(t|x)|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\}$ using

$$var\{\hat{\mu}(t|x)\} = E^*[var\{\mu(t|x)|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\}] + var^*[E\{\mu(t|x)|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\}], \quad (2.18)$$

where $E(\cdot)$ and $var(\cdot)$ are taken with respect to the posterior density when (β, η, λ) is assumed to be equal to $(\hat{\beta}, \hat{\eta}, \hat{\lambda})$ and $E^*(\cdot)$ and $var^*(\cdot)$ are taken with respect to the asymptotic density of $(\hat{\beta}, \hat{\eta}, \hat{\lambda})$. We can obtain an empirical Bayes estimator of $E^{*}[var\{\mu(t|x)|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\}]$ as

$$var\{\mu(t|x)|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\} = E\{\mu(t|x)^2|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\} - E\{\mu(t|x)|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\}^2. \qquad (2.19)$$

The second term in (2.18) can be computed from (2.16) using the Delta method.

2.4 Data Analysis

We analyzed the motivating example from Chapter 1, the study of heart transplant patients from the Medical University of South Carolina who received a heart transplant between January 1992 and May 2007. Cardiac transplant recipients may experience repeated non-fatal graft rejections (NFGR) after the transplant. These rejections can usually be treated with drugs if they are not very severe. For 105 patients in the data set, the maximum number of recurrent NFGR during the follow-up period is 7. We used two covariates–race ($x_1 = 0$ for Caucasian and 1 for African American) and Gender ($x_2 = 0$ for male and 1 for female), with additive link $\theta(x) = 1 + \beta_1 x_1 + \beta_2 x_2$ and exponentiated link $\theta(x) = exp(\beta_1 x_1 + \beta_2 x_2)$.

Table 2.1: Partial likelihood estimate (PL) and integrated likelihood estimates for different α values using additive link, $\theta(x) = 1 + \beta_1 x_1 + \beta_2 x_2$.

α	$\hat{\beta}_1(SE)$	$\hat{\beta}_2(SE)$	$\eta(SE)$
PL	1.3350(0.5324)	0.9117(0.5036)	0.7641(0.2788)
0.001	1.3228(0.5309)	0.8981(0.5016)	0.7498(0.2765)
0.01	0.9325(0.3791)	0.4473(0.3473)	0.3134(0.1543)
0.1	0.8445(0.3460)	0.3418(0.3134)	0.2235(0.1300)
1	0.7220(0.2970)	0.2106(0.2655)	0.1054(0.0984)
10	0.6293(0.2561)	0.1390(0.2296)	-0.0041(0.074)

We used both integrated likelihood and partial likelihood methods for inference and the estimates of (β, η) and their standard errors using the linear link function are shown in Table 2.1. The integrated likelihood estimates of β increases when the value of α decreases. For different values of α, the interval estimate (based on $\hat{\beta}_1$) of the coefficient of race are above zero, suggesting a higher risk of experiencing a new event for an African American patient compared to a Caucasian. On the other hand, the interval estimate of the coefficient of gender (β_2) do overlap with zero. Our

Table 2.2: Partial likelihood estimate (PL) and integrated likelihood estimates for different α values using exponentiatial link, $\theta(x) = exp(\beta_1 x_1 + \beta_2 x_2)$.

α	$\hat{\beta}_1(SE)$	$\hat{\beta}_2(SE)$	$\eta(SE)$
PL	0.7471(0.2461)	0.5120(0.2710)	0.7482(0.2729)
0.001	0.7458(0.2155)	0.5093(0.2249)	0.7413(0.2706)
0.01	0.6179(0.1887)	0.3290(0.2016)	0.3096(0.1532)
0.1	0.5824(0.1808)	0.2763(0.1945)	0.2203(0.1293)
1	0.5257(0.1676)	0.2023(0.1815)	0.1023(0.0981)

model allows η to take negative values, which means $\eta = 0$ is within the interior of the parameter space. Therefore a hypothesis test about $H_0 : \eta = 0$ vs $H_a : \eta \neq 0$ can be formed based on the asymptotic distribution of $(\hat{\beta}, \hat{\eta})$ under both the integrated and partial likelihood methods. We have found that tests based on the interval-estimate of η to be statistically significant only when α is relatively small for the integrated likelihood. This shows that if we assume the model to be too parametric (moderate to large $\alpha > 0.1$), the estimates and conclusion may be affected by the model. Since the partial likelihood is the limiting function of the integrated likelihood when α goes to zero, we can see the estimates of (β_1, β_2) and η based on partial likelihood are very similar to those from integrated likelihood with $\alpha < 0.01$. The integrated likelihood estimates have smaller standard error than that from partial likelihood. From the result in Table 2.2 we see, using the exponentiatial link leads to the same conclusions as the additive link. The estimates for β_1 and β_2 show the same trend and the estimate for η is very close to that from the additive link.

The plots of the cumulative mean functions of female and male patients are shown in Figure 2.1 using both empirical Bayes method and partial likelihood under the additive link. The plot on the left is for female patients and the one on the right is for male patients. The solid line and dashed line are for African American and Caucasian patients, respectively, using empirical Bayes method. The dotted line and dash-dot line are based on the partial likelihood. The cumulative baseline function is estimated with a similar nonparametric estimator in [4], which has the form as

$$\hat{\Lambda}(t) = \sum_{j=1}^{M} \frac{\sum_{i=1}^{n} N_{ij}}{\sum_{k \in R_j} (n_{kj}\hat{\eta} + \hat{\theta}_k)} \ ,$$

where $\theta_k = 1 + \beta_1 x_1 + \beta_2 x_2$. The mean function curves from these two methods have a similar shape. We can also see that African American patients (within any gender group) have a larger estimated number of rejections over time under all the methods we considered for estimation. The plots of estimated cumulative mean functions under exponentiated link are shown in Figure 2.2. The curves show a similar estimate as those under additive link in Figure 2.1.

The choice of link functions and inference procedures can be made based on the given situations and convenience of implementation. If there is no prior information on this matter, we developed a goodness-of-fit statistic to compare different link functions and different inference procedures. This goodness-of-fit statistic will be introduced in Chapter 4.

2.5 Time-varying covariates

We can extend our model when time-dependent covariates exist in the covariate vector. With time-dependent covariates, the cumulative mean function of our model is given as

$$E[N(t)|X(t)] = e^{\Lambda_0(t)\eta} \int_0^t e^{-\Lambda_0(u)\eta} \lambda_0(u)\theta(\beta x(u))du.$$

The rate function can be obtained by taking the derivative of the cumulative mean function as follows

$$r(t|X(t)) = \lambda_0(t)\theta(\beta x(t)) + \lambda_0(t)\eta e^{\Lambda_0(t)\eta} \int_0^t e^{-\Lambda_0(u)\eta} \lambda_0(u)\theta(\beta x(u))du.$$

Our rate function consists of two parts. The first part is proportional to the link function given the current state of the time-varying covariates and the second part involves the whole history of the time-varying covariates up to t. With the existence of time-dependent covariates, both the current state and the past history should have an effect on the rate of the recurrence. When the time-dependent covariates can be well approximated by piece-wise constant functions, the cumulative mean and rate function can be computed easily.

The partial likelihood function for time-varying covariates has a similar form as the partial likelihood in (2.4), which is

$$L_{PL}(\beta,\eta|D) \propto \prod_{i=1}^{n}\prod_{j=1}^{M}\Big\{\frac{n_{ij}\eta+\theta(\beta x_i(y_j))}{\sum_{l\in R_j}(n_{lj}\eta+\theta(\beta x_l(y_j))}\Big\}^{N_{ij}}. \tag{2.20}$$

2.6 Asymptotic Properties

For the semiparametric empirical Bayes estimators, there has not been any rigorous proof of the asymptotic properties in a general setting available in the literature. [37] gave an example for the additive hazard model with univariate survival response. To show the asymptotic normality of the empirical Bayes estimator $(\hat{\beta},\hat{\eta},\hat{\lambda})$, we first we will show the asymptotic normality of the score function, with respect to the marginal density of the data given (β,η,λ), where $\Lambda^*(t|\lambda)=\lambda t$ is the prior mean/guess for unknown $\Lambda_0(t)$. We assume that the prior guess for $\lambda_0(t)$ is a constant. We also assume the random censoring variable $C_i = min(C_{1i}, C_{2i})$, where C_{1i} is distributed independently of the recurrent events process and covariates, and C_{2i} equals the K_{th} recurrent event time for subject i. By making this assumption, an upper bound is put on the number of recurrent events a patient can have during the entire follow-up period. This assumption can be satisfied in many practical situations. For example, in a transplant study, due to the nature of the human body, the number of rejections one patient can take will not be big. Hence, there exists a big enough number as the upper limit of the number of rejections one patient can have. The non-informative censoring condition can be satisfied under our assumptions. The link function can be any monotone function. This condition holds for most common link functions, for example additive and exponentiated link function. Without loss of generality, we can assume that $x_i's$ are bounded between 0 and 1. The proof below holds for all $\beta\in[-a_\beta,\infty)$ and $\eta\in[-a_\eta,\infty)$ such that $-Ka_\eta+\theta(-a_\beta)>0$ and $-Ka_\eta+\theta(0)>0$. As mentioned before, under this assumption, $(0,0)$ is an interior point of the parameter space for (β,η). Therefore the hypothesis of $H_0:\eta=0$ vs $H_a:\eta\neq 0$ and $H_0:\beta=0$ vs $H_a:\beta\neq 0$ based on the asymptotic distribution can be performed. Proofs with more general form of prior mean $\Lambda^*(t)$ and censoring assumptions are omitted for the sake of brevity.

Based on the integrated likelihood function in (2.6), we can write the corresponding score functions as,

$$
\begin{aligned}
-\frac{\partial l}{\partial \beta} &= \sum_{j=1}^{M}\Big[\frac{\alpha\lambda\Delta_j B_{3j}}{\alpha+B_{2j}\eta+B_{1j}} - \sum_{i=1}^{n} N_{ij}\Big\{\frac{x_i}{n_{ij}\eta+\theta_i} - \frac{B_{3j}}{\alpha+B_{2j}\eta+B_{1j}}\Big\}\Big];\\
-\frac{\partial l}{\partial \eta} &= \sum_{j=1}^{M}\Big[\frac{\alpha\lambda\Delta_j B_{2j}}{\alpha+B_{2j}\eta+B_{1j}} - \sum_{i=1}^{n} N_{ij}\Big\{\frac{n_{ij}}{n_{ij}\eta+\theta_i} - \frac{B_{2j}}{\alpha+B_{2j}\eta+B_{1j}}\Big\}\Big];\\
-\frac{\partial l}{\partial \lambda} &= \sum_{j=1}^{M}\Big\{-\alpha\Delta_j \log\frac{\alpha}{\alpha+B_{2j}\eta+B_{1j}} - \sum_{i=1}^{n}\frac{N_{ij}}{\lambda}\Big\},
\end{aligned}
$$

where $\theta_i = \beta x_i$, $B_{1j} = \sum_{i\in R_j}\theta_i$, $\Delta_j = y_j - y_{j-1}$ and $0 = y_0 < y_1 < \ldots < y_M$ are ordered recurrent events times and censoring times for all subjects.

Let $G_{1i} = \sum_{j\in H_i}\frac{\lambda\alpha\Delta_j B_{3j}}{\alpha+B_{2j}\eta+B_{1j}} = \sum_{j\in H_i} G_{1ij}$, $G_{2i} = \sum_{j\in H_i}\frac{\lambda\alpha\Delta_j B_{2j}}{\alpha+B_{2j}\eta+B_{1j}} = \sum_{j\in H_i} G_{2ij}$, and $G_{3i} = \sum_{j\in H_i}\alpha\Delta_j log\frac{\alpha+B_{2j}\eta+B_{1j}}{\alpha}$, where $H_i = \{j : N_{ij} = 1,\ or\ y_j = C_i\}$. We can rewrite the score functions as

$$
\sum_{i=1}^{n}\Big(G_{1i}I_1 + G_{2i}I_2 + G_{3i}I_3 - \sum_{j=1}^{M} N_{ij}w_{ij}\Big), \tag{2.21}
$$

where $I_1 = (1,0,0)^T$, $I_2 = (0,1,0)^T$, $I_3 = (0,0,1)^T$, $w_{ij} = (w_{ij1}, w_{ij2}, w_{ij3},)^T$, $w_{ij1} = \frac{x_i}{n_{ij}\eta+\theta_i} - \frac{B_{3j}}{\alpha+B_{2j}\eta+B_{1j}}$, $w_{ij2} = \frac{n_{ij}}{n_{ij}\eta+\theta_i} - \frac{B_{2j}}{\alpha+B_{2j}\eta+B_{1j}}$, and $w_{ij2} = \frac{1}{\lambda}$.

Based on the integrated likelihood function, we can prove that $(G_{1i}, G_{2i}, G_{3i}, N_{ij}, j = 1,\ldots,M)$ for $i = 1,\ldots,n$ are independent random variables where G_{1i}, G_{2i}, and G_{3i} are defined as above, with Δ_j follows exponential distribution with parameter $\lambda_\Delta = \alpha\lambda\log\frac{\alpha+B_{2j}\eta+B_{1j}}{\alpha}$, $G_{2i} \sim Ga(d_i+1,\lambda)$ with parameter d_i as the number of recurrent events for subject i, and $\{N_{ij}, j = 1,\ldots,M\}$ follows a multinomial distribution with parameter d_i and $p_i = (p_{i1}, p_{i1},\ldots,p_{iM})$ where $p_{ij} = \frac{D_{ij}}{\sum_{j'=1}^{M} D_{ij'}}$ and $D_{ij} = \frac{n_{ij}\eta+\theta_i}{\alpha+B_{2j}\eta+B_{1j}}$.

According to the Cramér-Wold theorem, we can show the asymptotic normality of the score vector (2.21) by showing the asymptotic normality of $\sum_{i=1}^{n}\Big(G_{1i}I_1 + G_{2i}I_2 + G_{3i}I_3 - \sum_{j=1}^{M} N_{ij}w_{ij}\Big)^T a$ for all nonzero vector $a = (a_1, a_2, a_3)^T$. We suppose that there exists a fixed constant a_0, such that $\max(|a_1|, |a_2|, |a_3|) = a_0$ for any a. Under some mild conditions, using C_Δ-inequality, it can be shown that for any $\Delta > 0$,

$$\begin{aligned}
&\sum_{i=1}^{n} E|(G_{1i}I_1 + G_{2i}I_2 + G_{3i}I_3 - \sum_{j=1}^{M} N_{ij}w_{ij})^T a|^{2+\Delta} \\
\leqslant\ & (4a_0)^{2+\Delta}\Big\{\Gamma(3+\Delta)A_1(d_i+1) + \frac{\Gamma(d_i+\Delta+3)}{\Gamma(d_i+1)\lambda^{\Delta+2}} + A_2 d_i\Big\} = O_e(n),
\end{aligned}$$

where A_1 and A_2 are constants related to β, η, a_η and a_β.

We can also show that

$$\begin{aligned}
&var\Big\{\sum_{i=1}^{n}(C_iI_1 + G_{1i}I_2 + G_{2i}I_3 - \sum_{j=1}^{M} N_{ij}w_{ij})^T a\Big\} \\
\leqslant\ & a_0^2 \sum_{i=1}^{n}((d_i+1)A_3 + d_iA_4) = O_e(n),
\end{aligned}$$

where A_1 and A_2 are constants related to β, η, λ, a_η and a_β.

Then the two results above can give us

$$\frac{\sum_{i=1}^{n} E|(G_{1i}I_1 + G_{2i}I_2 + G_{3i}I_3 - \sum_{j=1}^{M} N_{ij}w_{ij})^T a|^{2+\Delta}}{\Big[var\Big\{\sum_{i=1}^{n}(C_iI_1 + G_{1i}I_2 + G_{2i}I_3 - \sum_{j=1}^{M} N_{ij}w_{ij})^T a\Big\}\Big]^{\Delta/2}} = O_e(n^{-\Delta/2}) \to 0. \tag{2.22}$$

The Liapounov condition for the asymptotic normality of $\sum_{i=1}^{n}(G_{1i}I_1+G_{2i}I_2+G_{3i}I_3-\sum_{j=1}^{M} N_{ij}w_{ij})^T a$ is validated by (2.22). After applying a Taylor series expansion we can obtain the proof of the asymptotic normality of the empirical Bayes estimator $(\hat{\beta}, \hat{\eta}, \hat{\lambda})$ based on the proof of the asymptotic normality of the score function.

2.7 Conclusion

We present a new regression model of (2.1) for intensity $h(t|\mathcal{N}(t-), x)$. Previous model of [29] has the intensity function $h(t|\mathcal{N}(t-), x) = \lambda_0(t)(1 + \eta N(t-))\exp(\beta x)$, related to the expression of $h(t|\mathcal{N}(t-), x)$ of our model. However, the formulation of our model ensures some advantages for our model compared to Oakes' model. Our model has similar interpretable regression functions for covariate effects on intensity as well as rate function. This property allows a more comprehensive understanding of the covariate effects on the recurrent event process compared to other competing models. Our intensity (and rate function) model allows for different link functions including exponentiated as well as additive covariate effects.

Our estimation method based on the integrated likelihood of (2.6) has a close relationship with a full semiparametric Bayesian analysis. We can regard the later as a special case of our method when the class of hyperpriors is assumed to have only one member prior, fully specified by the user. [16] and [15], among others, called this Type-2 likelihood method. One topic of interest, from a frequentist point of view, is the bias and the Mean-Square-Error (MSE) of the Type-2 MLE with respect to the "true" unknown process $\Pi = (\mu(\cdot), \Lambda_0(\cdot))$. Theoretical investigation of the MSE of integrated likelihood under "true" model is beyond the scope of this paper. However, via the following simulation study, we investigate the finite-sample sampling bias and MSE of our estimates under a known simulation model.

Another major concern is the sensitivity of the estimate of the regression effect when the link function is misspecified. A simulation study is carried out to show the performance of our estimate when the link function is misspecified. A single binary covariate x is considered, which takes values 0 and 1 with probability 1/2 for each value. The simulated data has sample size $n = 100$ and the true regression parameter is $\beta_0 = 2$. The true value of the effect of previous number of events is $\eta_0 = 2$. Since the model is built on the non-informative termination assumption, we assume an independent termination time C generated by a Uniform distribution between 0 and 1. We use a linear function as the true baseline intensity function $\lambda_0(t) = \lambda_{10}t + \lambda_{20}$ where $\beta_{10} = 2$ and $\beta_{20} = 0.1$. 3000 data sets are generated using exponentiated link function.

The average number of recurrent events for all the subjects is about 15. About 10% of the subjects don't have any recurrent events. It will not be fair to compare the estimated β and true β directly because the parameter has completely different meanings under two different link functions. However, to evaluate the performance of the estimation under wrong link, we can focus on how does the estimate of the hazard ratio $1+\hat{\beta}$ under additive link function compare to the hazard ratio $\exp(\beta)$ under the true model. The hazard ratio column in Table 2.3 is the hazard ratio of two subjects with covariate values 1 and 0 and no previous events and the value under the true model is $exp(\beta_0) = 7.3891$. We evaluate the performances of the estimated hazard ratio using partial likelihood and using integrated likelihood method with α smaller than 0.01. For the parameter for the effect of the number of previous events η, we have found that the estimate is not affected much for incorrectly specified link function.

Both partial likelihood method and integrated likelihood method with $\alpha < 0.01$ have very small bias. These biases are much smaller than the bias of a MLE obtained from a (wrongly specified) parammetric model with constant baseline intensity, i.e. $\lambda_0(t) = \lambda$. The bias and sampling standard deviation for integrated likelihood method (as well as for partial likelihood) are comparable to those obtained using MLE with correct parametric model. The integrated likelihood method with moderate α performs better than partial likelihood. We have also found that, for moderate values of α, the integrated likelihood method has smaller standard errors of η than that of the partial likelihood method. Therefore, in practice, when the functional form of the baseline and the link function are unknown, an excellent estimate of the regression effects can be obtained from our partial likelihood method and integrated likelihood method with α small.

Table 2.3: Parameter estimates (HR), bias and Mean Standard Error (MSE) from the simulation results: parameter of interest is the hazard ratio $\exp(\beta_0) \simeq 7.389$ between two subjects with no previous events and covariate values 1 and 0).

	HR	bias(%)	η SE	MSE
PL	7.6167	0.47	0.4027	0.4077
$\alpha = 0.001$	7.5817	0.85	0.4048	0.4288
$\alpha = 0.01$	7.5461	-0.08	0.3991	0.403
Correct Parametric MLE	7.4512	-0.25	0.0751	0.0734
Incorrect Parametric MLE	5.8686	114.8	0.8326	6.1616

We conducted another simulation study to show the performance of our model when the assumption about the baseline intensity is correct. We use the same setting as the previous simulation study and assume the true baseline intensity used to generate the data sets is a constant, i.e. $\lambda_0(t) = \lambda$. We compared the partial likelihood method, integrated likelihood method with different α values, maximum likelihood method with the correct assumption and with the incorrect assumption as a linear baseline intensity, i.e. $\lambda_0(t) = \beta_{10}t + \beta_{20}$. The bias of the parameter estimates and the standard error for both regression parameter β and effect of the previous number of events η are listed in Table 2.4. When we know the baseline intensity function, our methods give a similar relatively small bias as using maximum likelihood estimate (MLE) with the correct assumption. With moderate α value, we even have a smaller bias and standard error comparing to other methods. MLE failed completely when the assumption about the baseline intensity is not correct.

Table 2.4: Parameter estimates and standard errors from the simulation results when the assumption about the baseline is correct.

	β		η	
	bias(%)	SE	bias(%)	SE
PL	4.365	0.6241	0.655	0.3954
$\alpha = 0.001$	4.295	0.6043	1.32	0.4012
$\alpha = 0.01$	3.505	0.5962	0.59	0.3863
Correct Parametric MLE	3.945	0.5909	0.645	0.3464
Incorrect Parametric MLE	526.4	1.2096	77.05	0.1733

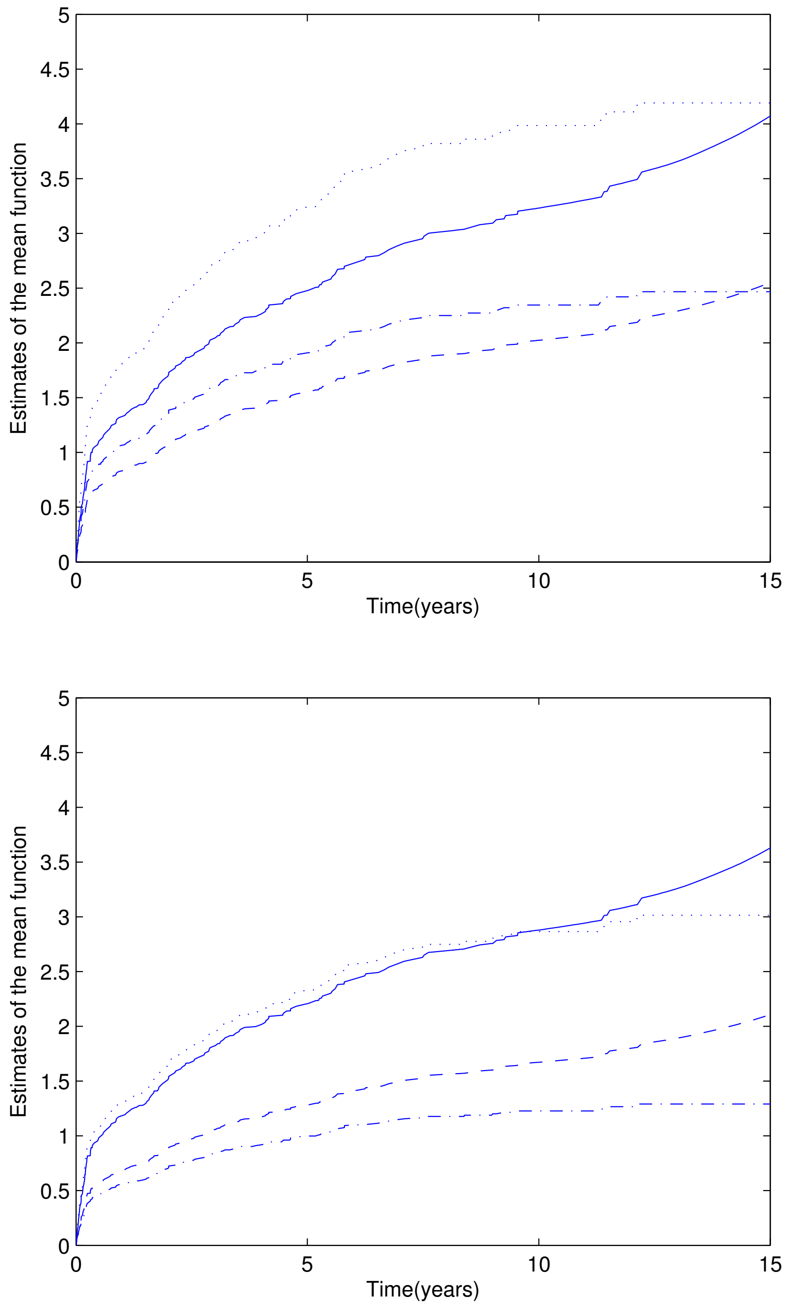

Figure 2.1: Plots of estimated mean functions under additive link of female (top panel) and male (bottom panel) patients: solid line= African American using empirical Bayes method; Dashed line= White using empirical Bayes method; Dotted line= African American using partial likelihood; Dash-dotted line= White using partial likelihood.

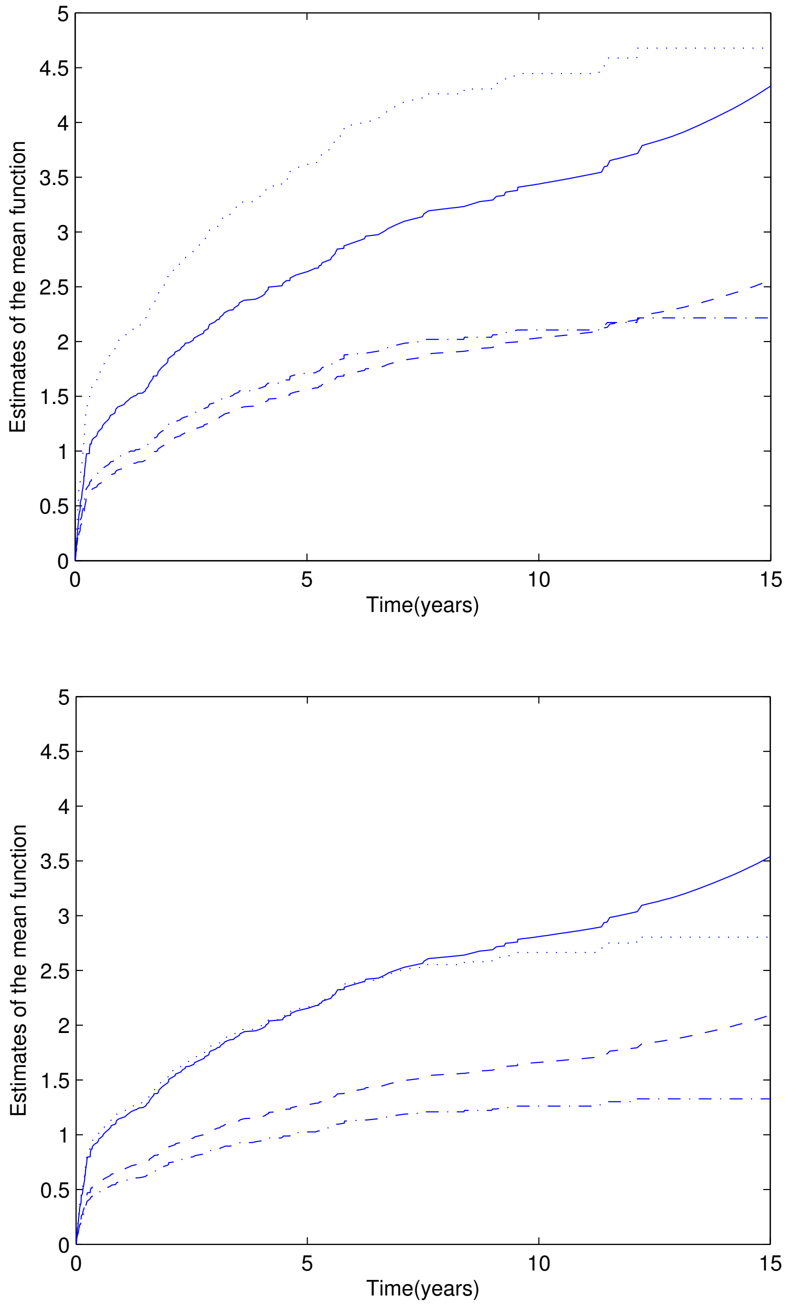

Figure 2.2: Plots of estimated mean functions under exponentiated link of female (left panel) and male (right panel) patients: solid line= African American using empirical Bayes method; Dashed line= White using empirical Bayes method; Dotted line= African American using partial likelihood; Dash-dotted line= White using partial likelihood.

CHAPTER 3

ADDITIVE INTENSITY MODELS FOR RECURRENT EVENTS DATA

In this chapter, we present a nonmultiplicative model for the risk of a new event for recurrent events data with covariate effects incorporated through an additive structure. Both the mean and rate functions of our model have closed form expressions.

To the best of our knowledge, there is no Bayesian method available for analyzing recurrent events data with non-multiplicative covariate effects. The empirical Bayes method used here combines the advantages of prediction and convenience of moderate sample justifications with computational advantages of a non Markov Chain Monte Carlo tool and large-sample justifications of a frequentist procedure. Unlike a full Bayes methods requiring problem specific programs and full specification and elicitation of the prior process of the baseline function, our method requires users to only specify the functional form of the 'prior guess' of the nonparametric part of the model, instead of eliciting all the hyperparameters of the prior process.

We obtain the empirical Bayes estimators of the regression parameters and of the mean function. We show the asymptotic properties of the semiparametric empirical Bayes estimates and illustrate the conveniences of our method by analyzing a study of recurrent minor graft-versus-host (GVH) episodes for a group of heart transplant patients.

3.1 Models and Integrated Likelihood

We consider the case where the recurrent events are subject to non-informative censoring/termination, which means the risk of termination time T at any time point t is independent of the recurrent event process $\{N(t)\}$, given the history of covariates

$X(t)$ and the history of recurrent events $\mathcal{N}(t-)$; that is

$$h_T(t|\mathcal{N}(t-), dN(t), X(t)) = h_T(t|\mathcal{N}(t-), X(t)). \tag{3.1}$$

From this point forward, we will assume fixed covariates x_i for all the derivations. The case with time-varying covariates $x_i(t)$ will be discussed later in Section 3.4. We assume the process $\{N_i(t)\}$ has a conditional intensity function

$$\lim_{dt\to 0} \frac{P[Y_{j+1} \in (t, t+dt)|\mathcal{N}(t-), x_i]}{dt} = h_{j+1}(t|x_i, N(t-) = j) = \lambda_0(t)(1 + j\eta) + x_i\beta, \tag{3.2}$$

where $\lambda_0(t)$ is the baseline hazard function, β is the additive covariate effect on the risk of a new event, and η is the effect of the current number of events $N(t-)$. For two subjects with covariate values x_1 and x_2 and with equal number of events $N(t-) = j$ by time t, the difference in risk for a new event is $h_{j+1}(t|x_2, N(t-) = j) - h_{j+1}(t|x_1, N(t-) = j) = \beta(x_2 - x_1)$, which can be evaluated using only regression parameter β.

For the frailty model (1.1) with $Ga(\eta^{-1}, \eta^{-1})$ frailty density of w (gamma density with unit mean and variance η), [29] showed that

$$\begin{aligned} P[dN(t) = 1|N(t-); x(t)] &\simeq h_{j+1}(t|N(t-); x(t))\, dt \\ &= (1 + N(t-)\eta)h_1(t|N(t-) = 0; x(t))\, dt. \end{aligned} \tag{3.3}$$

Our model of (3.2) is motivated by the Gamma frailty result of (3.3). However, we assume that the covariate effect on the conditional risk $h_{j+1}(t|N(t-) = j; x(t))$ acts additively to the risk $h_{j+1}(t|N(t-) = j; x(t) = 0) = h_0(t)(1 + j\eta)$.

For the model in (3.2), the conditional expectation $E[dN(t)|N(t-); x] = (1 + N(t-)\eta)d\Lambda_0(t) + \beta x dt$ gives a differential equation for the cumulative mean function $E[N(t)|x]$. By solving the differential equation, we can obtain

$$E[N(t)|x] = \frac{1}{\eta}(e^{\eta\Lambda_0(t)} - 1) + \theta e^{\eta\Lambda_0(t)} \int_0^t e^{-\eta\Lambda_0(u)} du. \tag{3.4}$$

Differentiating the cumulate mean function gives us the rate function

$$E[dN(t)|x] = e^{\eta\Lambda_0(t)}\lambda_0(t) + \theta + \theta\eta\lambda_0(t)e^{\eta\Lambda_0(t)} \int_0^t e^{-\eta\Lambda_0(u)} du.$$

Unlike the case of the risk of a new event conditional on the history $\mathcal{H}(t-)$, there is no simple interpretation of covariate effects on the mean and rate functions. The closed form expression in (3.4) can help predict the expected number of events for any given covariate value x.

Suppose $N_i(t)$ denotes the number of recurrent events by time t for subject i, $i = 1, \ldots, n$. Let $0 = y_0 < y_1 < \ldots < y_M$ be the ordered distinct event times and censoring times (maximum time under observation) for all subjects, and let N_{ij} be the indicator variable for occurrence of an event to subject i at time y_j. For ease of presentation, we first assume that the observed event times across all n subjects are distinct. According to our assumption of distinct event times, $\sum_{i=1}^{n} N_{ij} = 1$ when y_j is an observed event time. Let $I_j = (y_{j-1}, y_j]$, $x_i(t)$ be the covariate for the subject i, $R_j = R(y_j)$ be the set of subjects under observation at y_j, and $\delta_{ij} = 1$ when $i \in R_j$. The observed data is given by $D = \{N_i(t) : 0 < t \leqslant C_i;\ x_i;\ C_i : i = 1, \ldots, n\}$, where C_i is the censoring/termination time for subject i.

The full likelihood of (β, η, Λ_0) based on the observed recurrent events data D is given by

$$L(\beta, \eta, \Lambda_0 | D) = \prod_{i=1}^{n} \prod_{j=1}^{M} e^{[-\delta_{ij}\{\Lambda_{0j}(1+n_{ij}\eta)+\beta x_i \Delta_j\}]} \{d\Lambda_0(y_j)(1 + n_{ij}\eta) + \theta_i dy_j\}^{N_{ij}}, \quad (3.5)$$

where $\theta_i = x_i\beta$, $\Lambda_{0j} = \Lambda_0(y_j) - \Lambda_0(y_{j-1})$, $\Delta_j = y_j - y_{j-1}$ and $n_{ij} = N_i(y_j-)$ is the number of events that have occurred to subject i before y_j.

The likelihood in (3.5) can be interpreted only in a limiting sense as $dy_j \to 0$. For any observed event time y_j, the likelihood has two factors. The first factor is from the interval $[y_j, y_j + dy_j)$ and when $\sum_{i=1}^{n} N_{ij} = 1$ the likelihood contribution is

$$\sum_{i=1}^{n} N_{ij}\{d\Lambda_0(y_j)(1 + n_{ij}\eta) + \theta_i dy_j\} \exp[-\sum_{i \in R_j} \{d\Lambda_0(y_j)(1 + n_{ij}\eta) + \theta_i dy_j\}].$$

Otherwise, the likelihood contribution is 1. The second factor is from the interval (y_{j-1}, y_j) and it is approximately equal to

$$\prod_{i=1}^{n} \exp[-\delta_{ij}\{\Lambda_{0j}(1 + n_{ij}\eta) + \theta_i \Delta_j\}].$$

We assume $\Lambda_0(t)$ has a Gamma process prior ([23]) $\Lambda_0(t) \sim GP(\Lambda^*(t), \alpha)$ with prior mean $\Lambda^*(t) = \int_0^t \lambda^*(u)\,du$ with some unknown hyper-parameter λ^* and known prior precision α. The prior precision α specifies our degree of prior belief about unknown $\Lambda_0(\cdot)$ being close to the prior "guess" $\Lambda^*(\cdot)$. The prior "guess" (mean function) $\Lambda^*(\cdot)$ of the unknown $\Lambda_0(\cdot)$ is generally assumed to be a known parametric function $\Lambda^*(t|\lambda)$ indexed by the unknown hyperparameter λ. One example is $\Lambda^*(t|\lambda) = \lambda t$, where the prior "guess" is the constant baseline intensity with unknown hyperparameter λ. Either a literature survey or a pooling of experts' opinion are used to determine α. By integrating out the process $\Lambda_0(t)$ with respect to the prior process $\pi(\Lambda_0|\Lambda^*, \alpha)$ (gamma process), we get

$$\begin{aligned} L(\beta, \eta, \Lambda^*|D) &= \int L(\beta, \eta, \Lambda_0|D)\pi(\Lambda_0|\Lambda^*, \alpha)d\Lambda_0 \\ = \prod_{j=1}^{M} &\left[e^{-B_{1j}\Delta_j} \left(\frac{\alpha}{\alpha + r_j + B_{2j}\eta} \right)^{\alpha\Lambda_j} \prod_{i=1}^{n} \left\{ \frac{(1+n_{ij}\eta)\lambda^*(y_j)\alpha}{\alpha + r_j + B_{2j}\eta} + \theta_i \right\}^{N_{ij}} \right], \end{aligned} \tag{3.6}$$

where $\Lambda_j = \Lambda^*(y_j) - \Lambda^*(y_{j-1})$, $B_{1j} = \sum_{i \in R_j} \theta_i$, $B_{2j} = \sum_{i \in R_j} n_{ij}$, $\Delta_j = y_j - y_{j-1}$, $\lambda^*(t) = \frac{d}{dt}\Lambda^*(t)$, and $r_j = \sum_i \delta_{ij}$, that is, the number of subjects at risk at y_j, for $j = 1, \ldots, M$.

The integrated likelihood can be approximated when ties exist for the recurrent event times. For those recurrent events that occur at y_j, we can consider that they occur in very small consecutive intervals right after y_j. Therefore, in limiting sense we can use the same integrated likelihood in equation (3.6) to approximate.

We assume that the prior mean of the baseline function is an unknown constant, that is $\lambda^*(t) = \lambda$. For β, η and λ, the empirical Bayes estimates can be obtained by maximizing the integrated likelihood function of (3.6). Newton-Raphson method is used to maximize the finite dimensional integrated likelihood function. The corresponding score equations of the log-likelihood function $l(\beta, \eta, \lambda)$ for β, η and λ are given below:

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \sum_{j=1}^{M} \left[-\Delta_j B_{3j} + \sum_{i=1}^{n} N_{ij} \frac{x_i(\alpha + r_j + B_{2j}\eta)}{\alpha\lambda(1+n_{ij}\eta) + \theta_i(\alpha + r_j + B_{2j}\eta)} \right]; \\ \frac{\partial l}{\partial \eta} &= \sum_{j=1}^{M} \left[-\frac{\alpha\Lambda_j B_{2j}}{\alpha + r_j + B_{2j}\eta} + \sum_{i=1}^{n} N_{ij} \frac{\alpha\lambda\{n_{ij}(\alpha + r_j) - B_{2j}\}}{\{\alpha\lambda(1+n_{ij}\eta) + \theta_i(\alpha + r_j + B_{2j}\eta)\}(\alpha + r_j + B_{2j}\eta)} \right]; \\ \frac{\partial l}{\partial \lambda} &= \sum_{j=1}^{M} \left[\alpha\Delta_j log \frac{\alpha}{\alpha + r_j + B_{2j}\eta} + \sum_{i=1}^{n} N_{ij} \frac{\alpha(1+n_{ij}\eta)}{\alpha\lambda(1+n_{ij}\eta) + \theta_i(\alpha + r_j + B_{2j}\eta)} \right], \end{aligned}$$

where $B_{3j} = \sum_{i \in R_j} x_i$ and $\theta_i = x_i\beta$. The estimated variance-covariance matrix V of $(\hat{\beta}, \hat{\eta}, \hat{\lambda})$ is the inverse of the observed information matrix. The standard error obtained from V can be used to conduct the hypothesis testing involving (β, η) and to find the interval estimates. All the second partial derivatives are given in the Appendix.

3.2 The Posterior Process and Estimator

To obtain the empirical Bayes estimator of the mean function $E[N(t)|x]$ in equation (3.4), we will start with the posterior process $p(\Lambda_0|\beta, \eta, \lambda, D)$. We will first show that the Laplace transform of $p(\Lambda_0|\beta, \eta, \lambda, D)$ has a closed form, which will be subsequently used to obtain the empirical Bayes estimator of the cumulative mean function .

First a grid $0 = t_1 < t_2 < \ldots < t_{M+1} < \infty$ is put so that $t_j \in (y_{j-1}, y_j)$, where $0 = y_0 < y_1 < y_2 < \ldots < y_M < \infty$ are the ordered recurrent event times and censoring times for subject i, $i = 1, \ldots, n$. The increment of the baseline intensity function $\Lambda_0(t)$ generated by the grid and $y_j's$ can be denoted by $u_{j1} = \Lambda_0(y_j) - \Lambda_0(t_j)$, $u_{j2} = \Lambda_0(t_{j+1}) - \Lambda_0(y_j)$, and $u_j = \Lambda_0(t_{j+1}) - \Lambda_0(t_j) = u_{j1} + u_{j2}$. Also the increments of the prior mean $\Lambda^*(t)$ can be denoted by $\lambda_{j1} = \Lambda^*(y_j) - \Lambda^*(t_j)$, $\lambda_{j2} = \Lambda^*(t_{j+1}) - \Lambda^*(y_j)$, and $\lambda_j = \Lambda^*(t_{j+1}) - \Lambda^*(t_j) = \lambda_{j1} + \lambda_{j2}$. The likelihood function of $(\beta, \eta, u_1^*, u_2^*)$ can be written as

$$L(\beta, \eta, u_1^*, u_2^*|D) =$$
$$\prod_{j=1}^{M} \left[e^{-\{B_{1j}\Delta_j + u_{j1}(r_j + B_{2j}\eta) + u_{j2}(r_{j+1} + B_{2,j+1}\eta)\}} \prod_{i=1}^{n} \{d\Lambda_0(y_j)(1 + n_{ij}\eta) + \theta_i dy_j\}^{N_{ij}} \right],$$

where $\theta_i = \beta x_i$, $u_1^* = (u_{11}, \ldots, u_{M1})$ and $u_2^* = (u_{12}, \ldots, u_{M2})$. The joint distribution of the observed data vector $D = \{N_i(t) : 0 < t \leqslant T_i;\ x_i;\ T_i : i = 1, \ldots, n\}$ and $u^* = (u_1, u_2, \ldots, u_{M+1})$ is given by

$$L(\beta, \eta|u^*, D) = E_{u_1^*, u_2^*}[L(\beta, \eta, u_1^*, u_2^*|D)],$$

where the expectation is taken over the prior distribution of u_{j1}, u_{j2}, for $j = 1, \ldots, M$, which is an independent gamma distribution resulting from the Gamma prior process of $\Lambda_0(t)$.

Based on this joint distribution, we can get the expectation $E[\exp(-\sum_{j=1}^{M} h_j u_j)|D, \beta, \eta, \lambda]$ which is given by

$$
\begin{aligned}
& E[\exp(-\sum_{j=1}^{M} h_j u_j)|D, \beta, \eta, \lambda] && (3.7) \\
= & \prod_{j=1}^{M} \left\{ e^{-B_{1j}\Delta_j} \left(\frac{\alpha}{\alpha + r_j + B_{2j}\eta + h_j} \right)^{\alpha\lambda_{j1}} \left(\frac{\alpha}{\alpha + r_{j+1} + B_{2,j+1}\eta + h_j} \right)^{\alpha\lambda_{j2}} \right\} \times \\
& \prod_{j=1}^{M} \prod_{i=1}^{n} \left\{ \frac{\lambda^*(y_j)\alpha(1 + n_{ij}\eta)}{\alpha + r_j + B_{2j}\eta + h_j} + \theta_i \right\}^{N_{ij}} .
\end{aligned}
$$

The Laplace transform of $p[\Lambda_0(t)|D]$ can be obtained through the posterior expectation of $\exp(-\sum_{j=1}^{M} h_j u_j)$, which is given by dividing $E[\exp(-\sum_{j=1}^{M} h_j u_j)|D, \beta, \eta, \lambda]$ by the marginal density of the observed data, $L(\beta, \Lambda^*|D, \alpha)$ in equation (3.6). It can be shown that the Laplace transform has the expression as

$$
\begin{aligned}
& E[\exp(-\sum_{j=1}^{M} h_j u_j)|D] && (3.8) \\
= & \prod_{j=1}^{M} \left\{ \left(\frac{\alpha + r_j + B_{2j}\eta}{\alpha + r_j + B_{2j}\eta + h_j} \right)^{\alpha\lambda_{j1}} \left(\frac{\alpha + r_{j+1} + B_{2,j+1}\eta}{\alpha + r_{j+1} + B_{2,j+1}\eta + h_j} \right)^{\alpha\lambda_{j2}} \right\} \times \\
& \prod_{j=1}^{M} \prod_{i=1}^{n} \left[\frac{\frac{\lambda^*(y_j)\alpha(1+n_{ij}\eta)}{\alpha+r_j+B_{2j}\eta+h_j} + \theta_i}{\frac{\lambda^*(y_j)\alpha(1+n_{ij}\eta)}{\alpha+r_j+B_{2j}\eta} + \theta_i} \right]^{N_{ij}} .
\end{aligned}
$$

For any $t \in (y_k, y_{k+1})$, the Laplace transform of the posterior density $p[\Lambda_0(t)|D]$ is

$$
\begin{aligned}
& E[\exp(-h\Lambda_0(t))|D] && (3.9) \\
= & \left(\frac{\alpha + r_{k+1} + B_{2,k+1}\eta}{\alpha + r_{k+1} + B_{2,k+1}\eta + h} \right)^{\alpha\lambda_{0k}} \prod_{j=1}^{k} \left\{ \left(\frac{\alpha + r_j + B_{2j}\eta}{\alpha + r_j + B_{2j}\eta + h} \right)^{\alpha\lambda_{yj}} \right\} \times \\
& \prod_{j=1}^{k} \prod_{i=1}^{n} \left[\frac{\frac{\lambda^*(y_j)\alpha(1+n_{ij}\eta)}{\alpha+r_j+B_{2j}\eta+h} + \theta_i}{\frac{\lambda^*(y_j)\alpha(1+n_{ij}\eta)}{\alpha+r_j+B_{2j}\eta} + \theta_i} \right]^{N_{ij}} ,
\end{aligned}
$$

where $\lambda_{0k} = \Lambda^*(t) - \Lambda^*(y_k)$ and $\lambda_{yj} = \Lambda^*(y_j) - \Lambda^*(y_{j-1})$.

Equation (3.9) can be used to obtain the posterior estimate of the cumulative mean function $\mu(t|x) = E[N(t)|x]$ in the following way

$$
\begin{aligned}
& E[\mu(t|x)|D, \hat{\beta}, \hat{\lambda}, \hat{\eta}] \qquad (3.10)\\
= \; & \frac{1}{\hat{\eta}}(A^*(t) - 1) + \hat{\theta} A^*(t) \frac{1}{\alpha \hat{\lambda}} \times \\
& \Big[\sum_{l=1}^{k} \frac{A(y_l)}{log(\frac{\alpha + r_l + B_{2l}\hat{\eta}}{\alpha + r_l + B_{2l}\hat{\eta} + \hat{\eta}})} \Big\{ 1 - \Big(\frac{\alpha + r_l + B_{2l}\hat{\eta}}{\alpha + r_l + B_{2l}\hat{\eta} + \hat{\eta}} \Big)^{-\alpha\hat{\lambda}(y_l - y_{l-1})} \Big\} + \\
& \frac{A(t)}{log(\frac{\alpha + r_{k+1} + B_{2,k+1}\hat{\eta}}{\alpha + r_{k+1} + B_{2,k+1}\hat{\eta} + \hat{\eta}})} \Big\{ 1 - \Big(\frac{\alpha + r_k + B_{2k}\hat{\eta}}{\alpha + r_k + B_{2k}\hat{\eta} + \hat{\eta}} \Big)^{-\alpha\hat{\lambda}(t - y_k)} \Big\} \Big],
\end{aligned}
$$

where

$$
\begin{aligned}
A(t) &= E[e^{-\hat{\eta}\Lambda_0(t)}|D, \hat{\beta}, \hat{\lambda}, \hat{\eta}] \\
&= \left(\frac{\alpha + r_{k+1} + B_{2,k+1}\hat{\eta}}{\alpha + r_{k+1} + B_{2,k+1}\hat{\eta} + \hat{\eta}} \right)^{\alpha\lambda_{0k}} \prod_{j=1}^{k} \left\{ \left(\frac{\alpha + r_j + B_{2j}\hat{\eta}}{\alpha + r_j + B_{2j}\hat{\eta} + \hat{\eta}} \right)^{\alpha\lambda_{yj}} \right\} \times \\
& \prod_{j=1}^{k} \prod_{i=1}^{n} \left\{ \frac{\frac{\hat{\lambda}(y_j)\alpha(1 + n_{ij}\hat{\eta})}{\alpha + r_j + B_{2j}\hat{\eta} + \hat{\eta}} + \hat{\theta}_i}{\frac{\hat{\lambda}(y_j)\alpha(1 + n_{ij}\hat{\eta})}{\alpha + r_j + B_{2j}\hat{\eta}} + \hat{\theta}_i} \right\}^{N_{ij}} ;
\end{aligned}
$$

$$
\begin{aligned}
A^*(t) &= E[e^{\hat{\eta}\Lambda_0(t)}|D, \hat{\beta}, \hat{\lambda}, \hat{\eta}] \\
&= \left(\frac{\alpha + r_{k+1} + B_{2,k+1}\hat{\eta}}{\alpha + r_{k+1} + B_{2,k+1}\hat{\eta} - \hat{\eta}} \right)^{\alpha\lambda_{0k}} \prod_{j=1}^{k} \left\{ \left(\frac{\alpha + r_j + B_{2j}\hat{\eta}}{\alpha + r_j + B_{2j}\hat{\eta} - \hat{\eta}} \right)^{\alpha\lambda_{yj}} \right\} \times \\
& \prod_{j=1}^{k} \prod_{i=1}^{n} \left\{ \frac{\frac{\hat{\lambda}(y_j)\alpha(1 + n_{ij}\hat{\eta})}{\alpha + r_j + B_{2j}\hat{\eta} - \hat{\eta}} + \hat{\theta}_i}{\frac{\hat{\lambda}(y_j)\alpha(1 + n_{ij}\hat{\eta})}{\alpha + r_j + B_{2j}\hat{\eta}} + \hat{\theta}_i} \right\}^{N_{ij}} .
\end{aligned}
$$

These point estimates of $\mu(t|x) = E[N(t)|x]$ are obtained by inserting $\hat{\beta}$, $\hat{\eta}$, and $\hat{\lambda}$ into equation (3.10). The estimated variance of the empirical Bayes estimate $\mu(\hat{t}|x) = E[\mu(t|x)|D, \hat{\beta}, \hat{\lambda}, \hat{\eta}]$ can be computed from

$$
var\{\hat{\mu}(t|x)\} = E^*[var\{\mu(t|x)|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\}] + var^*[E\{\mu(t|x)|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\}], \qquad (3.11)
$$

where $E^*(\cdot)$ and $var^*(\cdot)$ are taken with respect to the asymptotic density of $(\hat{\beta}, \hat{\eta}, \hat{\lambda})$ and $E(\cdot)$ and $var(\cdot)$ are taken with respect to the posterior density when (β, η, λ)

is assumed to be equal to $(\hat{\beta}, \hat{\eta}, \hat{\lambda})$. The estimated variance consists of two parts, the variability from the posterior process and from estimating the hyperparameters. Using estimate from only one source will lead to an underestimated variance. The first term can be obtained from $var\{\mu(t|x)|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\}$, which takes form

$$var\{\mu(t|x)|D, \hat{\beta}, \hat{\eta}, \hat{\lambda}\} = \frac{\{A_2^*(t) - A^*(t)^2\}}{\hat{\eta}^2} E_1 + \hat{\theta}^2 E_2,$$

where $A_2^*(t) = E[e^{2\hat{\eta}\Lambda_0(t)}|D, \hat{\beta}, \hat{\lambda}, \hat{\eta}]$, $\hat{\theta} = x\hat{\beta}$, $E_1 = \left\{1 + 2\hat{\theta}\hat{\eta}\int_0^t A(u)du\right\}$ and $E_2 = \left[A_2^*(t)E\left\{\int_0^t e^{-\hat{\eta}\Lambda_0(u)}du\right\}^2 - A^*(t)^2\left\{\int_0^t A(u)du\right\}\right]$. All these terms can be computed using the Laplace transform of the posterior process in equation (3.9). The second term in equation (3.11) can be obtained from equation (3.10) using the Delta method.

3.3 Data Analysis

We applied the proposed method to the motivating example, the study of heart transplant patients from the Medical University of South Carolina receiving a heart transplant between January 1992 and May 2007. Cardiac transplant recipients may experience repeated non-fatal graft rejections (NFGR) after the transplant, which can usually be treated by drug therapy. There are 105 patients in this data set. The number of recurrent NFGR of a patient during the follow-up period is between 0 and 7 and the average is 1.6. Two covariates, Race (0 for Caucasian and 1 for African American) and Gender (0 for male and 1 for female), are considered for our data analysis to explore their effects on the recurrent event process.

Table 3.1: Estimates using non-multiplicative intensity model with different α values

α	$\hat{\beta}_1(SE)$	$\hat{\beta}_2(SE)$	$\eta(SE)$
0.1	0.2624(0.0381)	0.1710(0.0352)	0.0399(0.1121)
1	0.2589(0.0383)	0.1668(0.0354)	-0.0262(0.0842)
10	0.2485(0.0385)	0.1544(0.0357)	-0.0815(0.0565)
100	0.2192(0.0393)	0.1216(0.0368)	-0.0828(0.0513)
1000	0.1910(0.0410)	0.0937(0.0386)	-0.0441(0.0615)

Table 3.1 gives the point estimates of (β, η) using empirical Bayes method of Section 3.1. The estimates $\hat{\beta}$ change as α decreases and the values of $\hat{\beta}$ stabilize when

α is less than 100. There is a physical interpretation of α as the information content of the prior based on observing effectively α patients at a prior study ([20]). Having a value of $\alpha \geqslant 100$ implies too much prior precision around prior guess (especially when data has only 105 patients). This means that although the parameters of the prior process are estimated from the data, the estimates of covariate effects are not affected too much by the prior model under some threshold value of α. However, when α is too small, the variance of the the Gamma prior process will be too large, which will produce an unrealistic prior for the increment of the cumulative baseline hazard function $d\Lambda_0(t)$. The estimate of λ, hyperparameter of the prior guess/mean of the unknown $\lambda_0(t)$, is somewhat affected by α. One advantage of our model in (3.2) is that it allows the range of η to be $(-\frac{1}{K}, +\infty)$ where K is the pre-specified maximum number of events under observation period. We later discuss this in more detail. Because the case $\eta = 0$ is within the interior of the parameter space, we can actually test the hypothesis $H_0 : \eta = 0$ vs $H_a : \eta \neq 0$ based on the asymptotic distribution of $(\hat{\beta}, \hat{\eta})$. The estimate of η is not statistically significant no matter what values α takes, suggesting a lack of evidence for the effect of the number of past NFGR on the risk of a new NFGR.

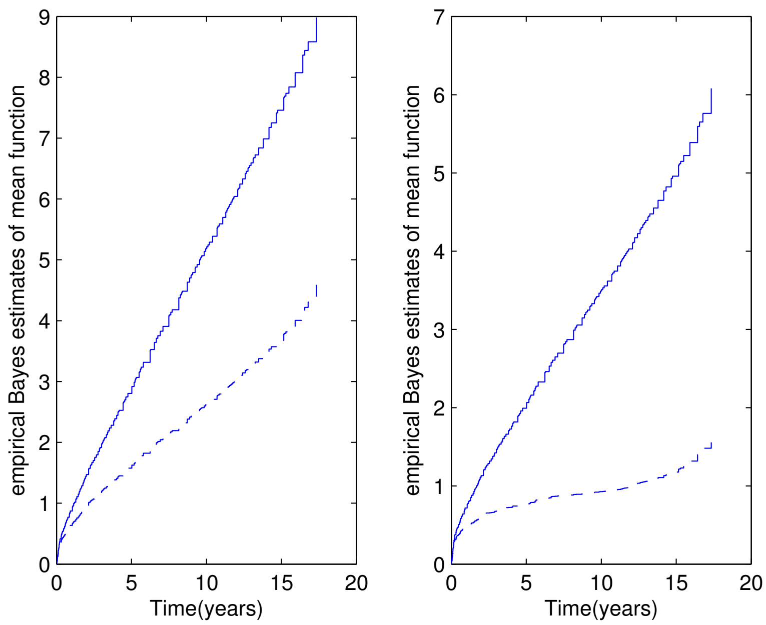

Figure 3.1: Plots of empirical Bayes estimates of mean functions of female and male patients (black, solid and white, dashed).

Figure 3.1 shows the empirical Bayes estimates of the mean function of groups with different covariate values when $\alpha = 1$. The two curves in the plot on the left are the estimate mean functions for white female patients and black female patients and the two curves in the plot on the right are those for white male patients and black male patients. In both male and female patients, black patients are estimated to have a higher number of events (rejections) than white patients.

3.4 Time-varying Covariates

We can extend the empirical Bayes method to the case where the covariate vector $x(t)$ is time-dependent. One extra assumption needed for this case is that the sample path of each component of $x(t)$ is either continuous or their discontinuity points are different from the ordered distinct recurrent events times and censoring times. For an additive conditional hazard function with time-varying covariates, the cumulative mean function can be written as

$$E[N(t)|X(t)] = \frac{1}{\eta}(e^{\eta\Lambda_0(t)} - 1) + \beta e^{\eta\Lambda_0(t)} \int_0^t e^{-\eta\Lambda_0(u)} x(u)du. \tag{3.12}$$

The integrated likelihood function can be obtained using similar method as in Section 3.1, which takes the form

$$\begin{aligned} & L(\beta, \eta, \Lambda|D) \\ = & \prod_{j=1}^{M} \left[e^{-B_{3j}\Delta_j} \left(\frac{\alpha}{\alpha + r_j + B_{2j}\eta} \right)^{\alpha\Lambda_j} \prod_{i=1}^{n} \left\{ \frac{(1 + n_{ij}\eta)\lambda^*(y_j)\alpha}{\alpha + r_j + B_{2j}\eta} + \theta_i(y_j) \right\}^{N_{ij}} \right], \end{aligned} \tag{3.13}$$

where $\Lambda_j = \Lambda^*(y_j) - \Lambda^*(y_{j-1})$, $B_{3j} = \sum_{i \in R_j} \beta z_{ij}$, $z_{ij} = \int_{y_{j-1}}^{y_j} x_i(u)du$, $B_{2j} = \sum_{i \in R_j} n_{ij}$, $\Delta_j = y_j - y_{j-1}$, $\theta_i(t) = \beta x_i(t)$, and $r_j = \sum_i \delta_{ij}$, i.e. the number of subjects at risk at y_j, for $j = 1, \ldots, M$. This integrated likelihood function is in similar form as the integrated likelihood function for time-independent covariates in equation (3.6) except for some different notations. We can also derive the Laplace transform of the posterior density of $p(\Lambda_0|\beta, \eta, \lambda, D)$ for the time-varying covariates in the similar steps in section 3.2, which is given by

$$
\begin{aligned}
& E[exp(-h\Lambda_0(t))|D] \qquad (3.14)\\
= & \left(\frac{\alpha + r_{k+1} + B_{2,k+1}\eta}{\alpha + r_{k+1} + B_{2,k+1}\eta + w}\right)^{\alpha\lambda_{0k}} \prod_{j=1}^{k} \left\{ \left(\frac{\alpha + r_j + B_{2j}\eta}{\alpha + r_j + B_{2j}\eta + h}\right)^{\alpha\lambda_{yj}} \right\} \times \\
& \prod_{j=1}^{k}\prod_{i=1}^{n} \left\{ \frac{\frac{\lambda^*(y_j)\alpha(1+n_{ij}\eta)}{\alpha+r_j+B_{2j}\eta+h} + \theta_i(y_j)}{\frac{\lambda^*(y_j)\alpha(1+n_{ij}\eta)}{\alpha+r_j+B_{2j}\eta} + \theta_i(y_j)} \right\}^{N_{ij}},
\end{aligned}
$$

where $\lambda_{0k} = \Lambda^*(t) - \Lambda^*(y_k)$ and $\lambda_{yj} = \Lambda^*(y_j) - \Lambda^*(y_{j-1})$.

The cumulative mean function in equation (3.12) can also be calculated based on the Laplace transform in (3.14). This will be similar as equation (3.10) in Section 3.2. We will omit the details for the sake of brevity.

3.5 Asymptotic Properties

There are very few results available in the literature about asymptotic properties of semiparametric empirical Bayes estimators. An exception is the proof of the consistency and asymptotic normality of the regression estimates for additive hazards model for univariate survival response ([37]). At first, we prove the asymptotic normality of the score function, with respect to the marginal density of the data given $(\beta, \eta, \lambda, \alpha)$, where $\Lambda^*(t|\lambda) = \lambda t$ is the prior mean/guess for unknown $\Lambda_0(t)$. No assumption about the function forms of $\Lambda_0(t)$ is made except its prior mean/guess. We also make the assumption about the random censoring variable $C_i = min(C_{1i}, C_{2i})$, where C_{1i} is distributed with finite moments, and independent of the recurrent events process and covariates, and C_{2i} equals the K_{th} recurrent event time for subject i. This assumption puts an upper bound on the number of recurrent events one patient can have during the entire observational period. This assumption can be satisfied in most situations in practice. For example in a transplant study, a big enough number for K can be chosen as the upper limit for the number of rejections one patient can have. This censoring process also satisfies the non-informative censoring condition of (3.1). The assumption on C_1 is a very minor assumption and true for any bounded censoring variable. Proofs with more general form of prior mean $\Lambda^*(t)$ and censoring assumptions are omitted for the sake of brevity.

The integrated log-likelihood has the expression

$$l = \sum_{j=1}^{M} \left\{ -B_{1j}\Delta_j + \alpha\Lambda_j \log\left(\frac{\alpha}{\alpha + r_j + B_{2j}\eta}\right) + \sum_{i=1}^{n} N_{ij} log\left(\frac{(1+n_{ij}\eta)\lambda^*(y_j)\alpha}{\alpha + r_j + B_{2j}} + \theta_i\right) \right\},$$

where $\theta_i = \beta x_i$, $B_{1j} = \sum_{i\in R_j}\theta_i$, $\Delta_j = y_j - y_{j-1}$ and $0 = y_0 < y_1 < \ldots < y_M$ are ordered recurrent events times and censoring times for all subjects. It can be further simplified as

$$l = -\sum_{i=1}^{n}\theta_i C_i + \sum_{j=1}^{M}\alpha\lambda\Delta_j log\frac{\alpha}{\alpha + r_j + B_{2j}\eta} + \sum_{i=1}^{n}\sum_{j=1}^{M} N_{ij} log\{\frac{\alpha\lambda(1+n_{ij}\eta)}{\alpha + r_j + B_{2j}\eta} + \theta_i\}.$$

The corresponding score functions can be rewritten as

$$\begin{aligned}
-\frac{\partial l}{\partial\beta} &= \sum_{i=1}^{n} x_i C_i - \sum_{i=1}^{n}\sum_{j=1}^{M} N_{ij}\frac{x_i}{\frac{\alpha\lambda(1+n_{ij}\eta)}{\alpha+r_j+B_{2j}\eta} + \theta_i};\\
-\frac{\partial l}{\partial\eta} &= \sum_{j=1}^{M}\alpha\lambda\Delta_j\frac{B_{2j}}{\alpha + r_j + B_{2j}\eta} - \sum_{i=1}^{n}\sum_{j=1}^{M} N_{ij}\frac{\alpha\lambda\frac{n_{ij}(\alpha+r_j)-B_{2j}}{(\alpha+r_j+B_{2j}\eta)^2}}{\frac{\alpha\lambda(1+n_{ij}\eta)}{\alpha+r_j+B_{2j}\eta} + \theta_i};\\
-\frac{\partial l}{\partial\lambda} &= \sum_{j=1}^{M}\alpha\Delta_j log\frac{\alpha + r_j + B_{2j}\eta}{\alpha} - \sum_{i=1}^{n}\sum_{j=1}^{M} N_{ij}\frac{\frac{\alpha(1+n_{ij}\eta)}{\alpha+r_j+B_{2j}\eta}}{\frac{\alpha\lambda(1+n_{ij}\eta)}{\alpha+r_j+B_{2j}\eta} + \theta_i}.
\end{aligned}$$

Let $G_{1i} = \sum_{j\in H_i}\lambda\alpha\Delta_j\frac{B_{2j}}{\alpha+r_j+B_{2j}\eta} = \sum_{j\in H_i} G_{1ij}$ and $G_{2i} = \sum_{j\in H_i}\alpha\Delta_j log\frac{\alpha+r_j+B_{2j}\eta}{\alpha}$ where $H_i = \{j : N_{ij} = 1, or \quad y_j = C_i\}$. The score functions can be rewritten as

$$\sum_{i=1}^{n}\left[C_i I_1 + G_{1i}I_2 + G_{2i}I_3 - \sum_{j=1}^{M} N_{ij}w_{ij}\right], \tag{3.15}$$

where $I_1 = (1,0,0)^T$, $I_2 = (0,1,0)^T$, $I_3 = (0,0,1)^T$, and $w_{ij} = (w_{1ij}, w_{2ij}, w_{3ij},)^T$, $w_{1ij} = \frac{x_i}{D_{ij}}$, $w_{2ij} = \frac{\alpha\lambda\frac{n_{ij}(\alpha+r_j)-B_{2j}}{(\alpha+r_j+B_{2j}\eta)^2}}{D_{ij}}$, $w_{3ij} = \frac{\frac{\alpha(1+n_{ij}\eta)}{\alpha+r_j+B_{2j}\eta}}{D_{ij}}$, $D_{ij} = \frac{\alpha\lambda(1+n_{ij}\eta)}{\alpha+r_j+B_{2j}\eta} + \theta_i$. Based on the integrated likelihood, we can show that $(C_i, G_{1i}, G_{2i}, N_{ij}, j = 1, \ldots, M)$ for $i = 1, \ldots, n$ are independent random variables with $G_{1i} = \sum_{j\in H_i} G_{1ij}$ where $G_{1ij} \sim Exp(\lambda_j)$ with $\lambda_j = \frac{B_{2j}}{(\alpha+r_j+B_{2j}\eta)\log\frac{\alpha+r_j+B_{2j}\eta}{\alpha}}$, $G_{2i} \sim Ga(d_i + 1, \lambda)$ where d_i is the number of recurrent events for subject i and $\{N_{ij}, j = 1, \ldots, M\}$ distributed as multinomial with parameter d_i and $p_i = (p_{i1}, p_{i1}, \ldots, p_{iM})$ where $p_{ij} = \frac{D_{ij}}{\sum_{j'=1}^{M} D_{ij'}}$. Based on our assumption, $d_i \leqslant K$ for $i = 1, \ldots, n$.

According to the Cramér-Wold theorem, to show the asymptotic normality of the score vector (3.15), it suffices to show the asymptotic normality of $\sum_{i=1}^{n}[C_iI_1+G_{1i}I_2+G_{2i}I_3-\sum_{j=1}^{M}N_{ij}w_{ij}]^Ta$ for all nonzero vector $a=(a_1,a_2,a_3)^T$. We can assume that there exists a fixed constant a_0 such that $\max(|a_1|,|a_2|,|a_3|)=a_0$ for any a. Under some mild assumptions, such as x_i is bounded, using C_Δ-inequality, it can be shown that for any $\Delta>0$,

$$\begin{aligned}
&\sum_{i=1}^{n} E|(C_iI_1+G_{1i}I_2+G_{2i}I_3-\sum_{j=1}^{M}N_{ij}w_{ij})^Ta|^{2+\Delta} \\
\leqslant\;& 2^{1+\Delta}a_0^{2+\Delta}\Big\{\left(\sum_{i=1}^{n}|x_i|^{2+\Delta}\right)E(C_{1i}^{2+\Delta})+M\Gamma(3+\Delta)\left(\frac{2}{\eta}\right)^{1+\Delta} \\
&\quad+\sum_{i=1}^{n}\frac{\Gamma(d_i+2+\Delta)}{\Gamma(d_i)\lambda^{2+\Delta}}+B_{w1}\sum_{i=1}^{n}d_i\Big\}=O_e(n),
\end{aligned}$$

where B_{w1} is a constant and $|w_{1ij}+w_{2ij}+w_{3ij}|^{2+\Delta}\leqslant B_{w1}$ for all i,j.

Further it can be shown that

$$\begin{aligned}
&var\Big\{\sum_{i=1}^{n}(C_iI_1+G_{1i}I_2+G_{2i}I_3-\sum_{j=1}^{M}N_{ij}w_{ij})^Ta\Big\} \\
\leqslant\;& a_0^2\left(\sum_{i=1}^{n}E_i+\frac{M}{\eta^2}+B_{w1}\sum_{i=1}^{n}d_i\right)=O_e(n),
\end{aligned}$$

where $E_i=\{var(C_{1i})+2E(C_{1i})^2+\frac{(d_i+1)}{\lambda^2}(1+E(C_{1i})\lambda)+\frac{d_iE(C_{1i})}{\eta}\}$, B_{w2} is a constant and $(w_{ij1}+w_{ij2})^2\leqslant B_{w2}$ for all i,j.

Then the two results above can give us

$$\frac{\sum_{i=1}^{n}E|(C_iI_1+G_{1i}I_2+G_{2i}I_3-\sum_{j=1}^{M}N_{ij}w_{ij})^Ta|^{2+\Delta}}{\left[var\{\sum_{i=1}^{n}(C_iI_1+G_{1i}I_2+G_{2i}I_3-\sum_{j=1}^{M}N_{ij}w_{ij})^Ta\}\right]^{1+\Delta/2}}=O_e(n^{-\Delta/2})\to 0. \tag{3.16}$$

The Liapounov condition for the asymptotic normality of $\sum_{i=1}^{n}(C_iI_1+G_{1i}I_2+G_{2i}I_3-\sum_{j=1}^{M}N_{ij}w_{ij})^Ta$ is validated by (3.16). The proof of the asymptotic normality of the empirical Bayes estimator $(\hat\beta,\hat\eta,\hat\lambda)$ can be obtained by an application of a Taylor series expansion based on the proof of the asymptotic normality of the score function.

3.6 Discussion

We have presented a very computationally convenient procedure for estimating regression effects on an extension of Aalen's additive hazards model ([2]) to the intensity function of recurrent events. There are some similarities and close relationships between our method and a fully Bayesian semiparametric method. Our method uses a family of priors for $(\beta, \eta, \Lambda_0(t))$ that are characterized by a hyperparameter α, which is the precision parameter of the gamma process prior on $\Lambda_0(t)$. (β, η) has a constant as prior with the support that will make the joint prior appropriate. Similar as the relationship between an empirical Bayes semiparametric approach and a full semiparametric Bayesian method [34] for multiplicative intensity, one can regard a full semiparametric Bayesian method as a special case of our method when the set of hyperparameters of the prior process for $(\beta, \eta, \Lambda_0(t))$ is assumed to have only one member, fully specified by the user. From a subjective Bayesian point of view, the result of our analysis can be used as a benchmark for studying the sensitivity of a full Bayesian procedure.

The noninformative censoring assumption may not be valid. For example, when the censoring/termination is caused by death, the risk of death depends on the risk of NFGR. In this case, the relationship between the recurrent events times and censoring/termination times should be incorporated in the model. There have been some models in the literature for informative termination ([36], [28]). One topic for our future work is to investigate the empirical Bayes method for informative censoring. The data set we used for analysis is only a subset of the actual study and the actual study subjects were followed until death. Models with informative censoring should give a better fit for our data set.

CHAPTER 4

MODEL COMPARISON

With the broad range of models that the above Bayesian formulation includes, model selection becomes an important question. Although several models may provide adequate fit to the data, each model for cure-rate survival data represents a hypothesis (or a set of hypotheses) and it is beneficial to have a framework for choosing between these models.

4.1 Existing Methods

One decision-theoretic criteria proceeds from a posterior predictive loss paradigm (Gelfand and Ghosh, 1998). The posterior predictive distribution for subject i is given by

$$p(t_i^*|D) = \int p\left(t_i^*|D, \Omega\right) p(\Omega|D) d\Omega, \tag{4.1}$$

$p\left(t_i^*|D, \Omega\right)$ further simplifies to the likelihood $L(\Omega; D)$ due to the assumed independence between t_i^* and D conditional upon Ω being known. Sampling from (4.1) proceeds using composition sampling: given samples $\{\Omega_{(j)}\}_{j=1}^G$ from the posterior distribution, we sample $t_{i(j)}^*$ from $p(t_i^*|D, \Omega = \Omega_{(j)})$ for $i = 1, \ldots, n$ and $j = 1, \ldots G$. The collection $\{t_{ij}^*\}_{j=1}^G$ are samples from the posterior predictive distribution of the i-th subject, which is $p(t_i^*|D)$. Preferred models will perform well under a decision-theoretic *balanced loss function* that yields a model selection metric called the L-measure (e.g. [22]) that can be calculated (using Monte Carlo approximation) as

$$\begin{aligned} L_M &= E\left[\sum_{i=1}^{n}(\log(t_i) - \log(t_i^*))^2|D\right] \qquad (4.2) \\ &\approx \frac{1}{G}\sum_{j=1}^{G}\sum_{i=1}^{n}(\log(t_i) - \log(t_{i(j))}^*)^2. \end{aligned}$$

If there is no cure-rate and no right-censoring, the sampling of $t_{i(j)}^*$ given Ω_j (MCMC sample from posterior) can be easily done. For the cure-rate survival model subject to censoring, the response consists of $\min(T_i, C_i)$ and censoring indicator $\delta_i = I_{(T_i \leq C_i)}$, where T_i is the survival time under cure-rate model and C_i is the non-informative censoring variable. Therefore, we need a new definition of t_i^*.

Deviance Information Criterion (DIC, [38]) is a Bayesian criterion for model comparison. This method is based on identifying models that can fit the observed data the best and likely minimize the uncertainty about observations generated in the same way. It has a similar format as some of the classical model selection methods, such as Akaike Information Criterion (AIC, [3]) and Bayesian Information Criterion (BIC, [33]).

For a Bayesian parametric statistical model, it includes a probability model for the likelihood, $p(y|\theta)$, $\theta \in \Theta$, and a prior distribution , $p(\theta)$, specified on the parameter θ. A Bayesian measure of model complexity can be written as

$$p_D = \bar{D(\theta)} - D(\bar{\theta}) \qquad (4.3)$$

where $D(\theta) = -2log\{p(y|\theta)\} + 2log\{f(y)\}$ and $\bar{\theta} = E[\theta|y]$, the posterior mean of the parameters. $D(\theta)$ should be called 'Bayesian Deviance' in general. Therefore, from Eqn (4.3) we can see that p_D can be considered as a 'mean deviance minus deviance of mean' ([38]).

After having the dimension of the model, DIC is defined as this:

$$\begin{aligned} DIC &= D(\bar{\theta}) + 2p_D \qquad (4.4) \\ &= \bar{D} + p_D \end{aligned}$$

It can be considered as a 'Bayesian measure of fit or adequacy,penalized by an additional complexity term p_D' ([38]).

DIC is very easy to compute especially when all the MCMC samples are available. Suppose θ_i, $i = 1, 2, \ldots, M$ are the posterior samples from MCMC, it gives us

$$\begin{aligned} DIC &= 2\bar{D} - D(\bar{\theta}) \\ &= 2[\frac{1}{M}\sum_{i=1}^{M} D(\theta_i)] - D(\frac{1}{M}\sum_{i=1}^{M}\theta_i) \end{aligned} \tag{4.5}$$

However, according to the definition, DIC, as well as p_D, could be negative under certain circumstances. Having negative values for p_D would make the interpretation as the dimension of the model very complicated. When the model has a complicated structure, for example, cure-rate model for survival data, DIC may not be able to give a good estimate for p_D, which will lead to a not very good model comparison result. Besides, DIC also have some problems with choosing the correct model as sample size increase. We conducted a simulation study to investigate the disadvantages of DIC and will show the results in a later section.

4.2 Proposed Model Comparison Method

We propose a new measure of model diagnostic for censored data. For censored data, since the censoring indicator is also part of the observed data, the closeness between the observed Y and the predicted Y_p is not enough. For example, when $\delta_{obs} = 1$ and observed $y_{obs} = t_{obs} = 10$, then prediction of censoring at $y_p = 11$ is not as good as failure/death at $y_p = 11$, although $(y_{obs} - y_p)^2$ are the same for both cases. In presence of censoring and cured fraction (i.e. $E(T)$ is not finite), it may not be appropriate to perform selection based on either $E(T_i - T_{ip})^2$ or $E(Y_i - Y_{ip})^2$. To avoid any of these problems, we need to give a suitable penalty that balances the two parts. We use the counting process of number of deaths over time to compare with the number of observed deaths over time to define our measure of model adequacy. Let $Y_i = \min(T_i, C_i)$ and $\delta_i = 1_{[T_i \leq C_i]}$. We define the counting process $N_+(t) = \sum_{i=1}^{n} N_i(t)$, where $N_i(t) = 1_{[Y_i \leq t, \delta_i = 1]}$. The $N_+(t)$ denotes the number of observed failures before time t. Let $N_o(t)$ be the observed value of the counting process $N_+(t)$ from data D. Let $N_p(t)$ denote the predicted value of the counting process $N_+(t)$.

Our new measure, which we call the *M-measure* for model m, is defined as

$$M_m = E\left[\int_0^\tau \|N_o(t) - N_p(t)\|dt|Data, F_c = \hat{F}_{CKM}\right], \tag{4.6}$$

where $\tau = \max_i\{Y_i\}$ and the distribution of censoring variable C_i is assumed to be known as $\hat{F}_{CKM}$, the Kaplan-Meier estimator of the cumulative distribution function of C from D. Different forms of norm can be used in the formula, such as the absolute value and the square.

Smaller M-measures mean better adequacy of fit and more precise predictive fit. The advantage of this criterion, compared to *deviance information criterion* (DIC), is that only very weak assumption is made for the computation of M-measure. The Kaplan-Meier estimator of the cumulative distribution function of C is not necessary. Other assumptions of the censoring time can also be used, such as the exponential distribution with the rate that has a Gamma distribution. So a wide range of models can be compared according to the M-measure.

4.3 Proportional odds model with cure-rate for survival data

With rapid developments in medical and health sciences, we now encounter more survival studies where some patients are expected to be *cured*. Survival models that account for cure are important for understanding prognosis in potentially terminal diseases. Traditional parametric survival models such as Weibull or Gamma ([11]) do not account for the probability of cure. Although subtle, one needs to distinguish between the concepts of censoring and cure: censoring refers to a subject who does not fail within the monitoring time window of a particular subject , while cure refers to one who will not fail within any reasonable monitoring time window. Indeed the latter is an abstraction as we never "observe" a cure (due to a finite monitoring time). Still estimating the probability of such an outcome, especially in various cancer-relapse settings, can help expose unknown health issues concerning that population.

Recently much attention has been devoted to formulating parametric survival models incorporating a *cured fraction* – a non-zero tail probability of the survival function. These have focused upon cancer-relapse trials including breast cancer, non-Hodgkins lymphoma, leukemia, prostate cancer, melanoma, and head and neck cancer, where due to recent advances in therapy and treatment, a significant proportion of patients are expected to be "cured", that is to remain disease-free even after really long follow-ups. *Cure rate models* incorporating a cured fraction, defined as a non-zero tail-probability of the survival function, adjust for this feature of the data and date back to the mixture model by [5] (mixture-cure model, in short) and has been extensively discussed by several authors, including [13], [14], [18], [27], [12], and [39]. In this model, the survival function for the entire population is given by

$$P[T > t] = S_p(t) = \pi + (1 - \pi)S(t), \tag{4.7}$$

where $\pi = S_p(+\infty)$ is the "cured fraction", and $S(t)$ with $S(+\infty) = 0$ is the *proper* survivor function for the non-cured group. In the presence of the $p \times 1$ vector of covariates $\boldsymbol{x}_i' = (x_{i1}, \ldots, x_{ip})$ for the i^{th} subject, assuming an accelerated failure time model $S(t|x_i) = S_0(t\theta(x_i))$ for non-cured subject and the cured fraction π to be free of x_i, we get obtain an accelerated failure time model

$$S_p(t|x_i) = \pi + (1 - \pi)S_0(t\theta(x_i)) = S_{p0}(t\theta(x_i)) \tag{4.8}$$

for the population survival function.

Another class of models, the Bounded Cumulative Hazard model, formulated by [44], [43], [45] and [6] (BCH model, in short) in cancer relapse settings, assume that a latent biological process of propagation of latent clonogenic tumor cells (latent factors) is generating the observed failure (relapse). [8] generalized this framework to a flexible class of cure models under latent activation schemes. Consider a typical cancer setting where for each individual in the population under study, we posit a certain unknown number, N, of latent factors. Let Z_i be the time (promotion time) for the i^{th} latent factor. Given $N > 0$, $Y_1, \cdots, Y_N$ are assumed to be independent and identically distributed with a common distribution function $F(y) = 1 - S(y)$ that does not depend upon N. The time to final event of interest can be defined by $T = \min\{Y_i : 1 \leq i \leq N\}$, when $N > 1$. This model can be used in cancer relapse or other disease models whenever we can envisage one or several *latent factors* or *events* corresponding to each patient. For an individual to be at *risk* of failure, he/she must be exposed to at least one of these latent factors. If $N = 0$, then the individual is not at risk of final event and is considered *cured.* Failure is observed when one (or some) of these latent factors get *activated/prommoted.*

Notice that N must be modeled using a stochastic mechanism. The number of possible latent events N can have any finite-mean integer-valued distribution (e.g., Binary, Geometric, etc.) with the moment generating function $m(t) = E[\exp(tN)]$ and a *cure fraction* defined as $P(N = 0) = m(-\infty)$. In this setting, the marginal distribution of T is given in terms of $m(t)$ as ([8]):

$$S_p(t) = E_N[P(T \geq t|N)] = m[\log S(t)]. \tag{4.9}$$

For example, in the traditional mixture-cure model, N is binary $N \sim Ber(\theta)$ $(0 \leq \theta \leq 1)$ with $m(t) = 1 - \theta(1 - e^t)$ to give

$$S_p(t) = (1 - \theta) + \theta S(t), \tag{4.10}$$

with cure fraction $1 - \theta \leq 1$. The BCH model assumes that N has a Poisson distribution with $m(t) = \exp[-\theta(1 - e^t)]$ for $\theta > 0$ and the corresponding marginal cure rate model is

$$S_p(t) = e^{-\theta(1-S(t))}, \tag{4.11}$$

with cure fraction $\exp(-\theta)$. The biological arguments for using this assumption for a cancer relapse study are put forward by [19] among others. However, [41], among others, find the Poisson assumption at best debatable irrespective of any situation involving cure in cancer. The class of models in (4.9) is far more general than these two competing models in existing literature of cure-rate survival data.

The hazard function of the BCH model in (4.11) is given by

$$h_p(y) = \theta f(y), \tag{4.12}$$

when the covariate vector $\boldsymbol{x}_i$ for the i^{th} subject is incorporated through the cure rate parameter θ_i as $\theta_i \equiv \theta(\boldsymbol{x}_i'\boldsymbol{\beta}) = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ denote the corresponding vector of regression coefficients, and $F(t)$ is assumed to be free of $\boldsymbol{x}_i$ to get a proportional hazards structure for the population hazard in (4.12).

The cure models envisaged by [8] and others deal with failure (relapse) times at two different levels: an *observed* failure time, say T, corresponding to the time when the individual *fails*, and the *latent* event times, $Y_k, k = 1, \ldots, N$, the activation times for the N latent factors that generate the observed failure at time T. Note that if $N = 0$ then the individual is not exposed to any of the latent factors and is considered immune from failure. Conditional upon N, the Y_k's are assumed to be independently and identically distributed with a *latent survival function* $P(Y > t) = S(t) = 1-F(t)$. When N is distributed as $Geo(\theta)$ with p.d.f. $P[N = k] = \theta^k/(1+\theta)^{k+1}$, then we get the population survival function as

$$S_p(t) = [1 + \theta F(t)]^{-1}. \tag{4.13}$$

This survival function in (4.13) has proportional odds structure when covariate $\boldsymbol{x}$ is modeled via $\theta(\boldsymbol{x})$ and the latent survival $S(t) = 1 - F(t)$ is free of $\boldsymbol{x}$, because

$$\{1 - S_p(t|\boldsymbol{x})\}/S_p(t|\boldsymbol{x}) = \theta(\boldsymbol{x})F(t)\ . \tag{4.14}$$

There is an alternative derivation of the proportional odds model with cure as an extension of the transformation-cure model ([47]).

The corresponding hazard for (4.14)is

$$h_p(t|\boldsymbol{x}) = -\frac{d}{dt} log\{S_p(t|\boldsymbol{x})\} = \frac{\theta(\boldsymbol{x})f(t)}{1+\theta(\boldsymbol{x})F(t)}\ , \tag{4.15}$$

where the density $f(t)$ of $F(t)$ is assumed to be continuous except at finite time points. For proportional odds model with no cured fraction, the ratio of hazards $h(t|\boldsymbol{x}_1)/h(t|\boldsymbol{x}_2)$ goes to one as $t \to \infty$ and it goes to $\theta(\boldsymbol{x}_1)/\theta(\boldsymbol{x}_2)$ as $t \to 0$. However, for the proportional odds-model with cured fraction in (4.13), the ratio $h_p(t|\boldsymbol{x}_1)/h_p(t|\boldsymbol{x}_2)$ goes to $[1+\theta(\boldsymbol{x}_2)]\theta(\boldsymbol{x}_1)/\{[1+\theta(\boldsymbol{x}_1)]\theta(\boldsymbol{x}_2)\}$ as $t \to \infty$, because $F(+\infty) = 1$.

We specify the latent survival function $S(t)$ using a two-parameter Weibull distribution $Weib(\rho, \eta)$ with survival function $S(t) = \exp(-\eta t^\rho)$. This implicitly assumes that hazard $h(t) = \eta\rho t^{\rho-1}$ is either increasing (for $\rho \geq 1$) or decreasing (for $\rho \leq 1$). However, the corresponding $h_p(t|\boldsymbol{x})$ may not have the same monotonic trend. When $h(t) = \eta$ (constant), the corresponding $h_p(t|\boldsymbol{x}) = \{\theta(\boldsymbol{x})\eta\}/\{(\theta(\boldsymbol{x}) + 1)e^{\eta t} - \theta\}$ is strictly decreasing.

For the i^{th} individual, our observed data $\mathbf{D}_i = \{y_i, \delta_i, \boldsymbol{x}_i\}$ consists of covariate vector $\boldsymbol{x}_i$, $y_i = \min(T_i, C_i)$ as the observed failure time, $\delta_i = I[T_i \leq C_i]$ as the failure indicator, where C_i is the non-informative random censoring time. We denote the model parameters (and hyper-parameters) into Ω, which actually depends on the specific model. The contribution of subject i to the data likelihood (in a right-censored setting) is

$$L(\Omega \,|\, \mathbf{D}_i) = S_p(y_i \,|\, \Omega; \boldsymbol{x}_i) \times \{h_p(y_i \,|\, \Omega; \boldsymbol{x}_i)\}^{\delta_i}$$

where for the proportional odds model with cure rate, $S_p(t \,|\, \Omega; \boldsymbol{x}_i)$ and $h_p(t \,|\, \Omega; \boldsymbol{x}_i)$ are given in (4.13) and (4.15) respectively. For other models, such as the mixture-cure and the BCH model, the S_p and h_p will be corresponding to the chosen model.

The posterior distribution of Ω is

$$p(\Omega \,|\, \mathbf{D}) \propto \left[\prod_{i=1}^{n} L(\Omega \,|\, \mathbf{D}_i)\right] \times \pi(\Omega), \tag{4.16}$$

where $\mathbf{D} = \{\mathbf{D}_i\}_{i=1}^{n}$ denotes the observed data and $\pi(\Omega)$ is the joint prior of Ω. For the model in (4.15), it is assumed to be $\pi(\Omega) = \pi_1(\rho, \eta)\pi_2(\boldsymbol{\beta}|\rho, \eta)$. A more precise notation would acknowledge L and Ω to depend on the model m, but we supress the dependence of L and Ω on m in the notation for ease of presentation. In general the marginalization of $p(\Omega \,|\, \mathbf{D})$ is analytically intractable and is performed using Markov Chain Monte Carlo tool, which iteratively samples from the joint posterior using possibly Metropolis updates for the full conditionals. In general, we adopt normal proposals for $\boldsymbol{\beta}$, log-normal for η and Gamma for ρ (as the case may be).

For proportional odds model with cure-rate, the *M-measure* can be computed as a two-step procedure. First $\{Z_{ipj}, \delta_{ipj}\}$ are simulated from the posterior predictive distribution of (Y_i, δ_i), i.e., $p(Z_{ip}, \delta_{ip}|\Omega_j, F_c = \hat{F}_{CKM})$, where Ω_j is the j-th MCMC sample ($j = 1, ..., N$) from the posterior density $p(\Omega|D)$ for model M. For each C^*_{ij} sampled from the Kaplan-Meier estimated cdf $\hat{F}_{CKM}$, we sample $\delta^*_{ij} \sim Ber(F_p(C_{ij}))$ where $F_p(C_{ij}) = \theta_i F(C^*_{ij})/(1 + \theta_i F(C^*_{ij}))$ and set $Z^*_{ij} = C^*_{ij}$ if $\delta^*_{ij} = 0$. When $\delta^*_{ij} = 1$, we sample $U^*_{ij} \sim U(0, 1)$ and set $Z^*_{ij} = F^{-1}_{cure}(U^*_i)$, where

$$F_{cure}(y) = \frac{F(y)(1 + \theta_i F(C^*_{ij}))}{\{1 + \theta_i F(y)\}F(C^*_{ij})}$$

for $0 < y < C^*_{ij}$. Once the posterior predictive samples are obtained, for each j we compute

$$M_j = \int_0^{\tau} |N_o(t) - N_{pj}(t)|dt = \sum_{j'=1}^{K} |N_o(a_{j'}) - N_{pj}(a_{j'})|\Delta_{j'},$$

where $0 = a_0 < a_1 < a_2 < \ldots < a_K < a_{K+1} = \tau$ are distinct points where $N_o(t)$ and $N_{pj}(t)$ have jumps, and $\Delta_{j'} = a_{j'+1} - a_{j'}$. Finally, M_m is obtained as an average over the M_j's for $j = 1, ..., N$, where N is a large number (chosen based on achieving certain level of Monte Carlo error).

4.4 Simulation

To compare our proposed model comparison method with DIC, a couple of simulation studies were conducted under different situations. First we want to see the difference of performance of the two methods on the proportional model with cure-rate and YCIS (BCH) model ([44], [43], [45] and [6]). m sets of data are generated from the YCIS model with n subjects in each data set.

Here is the algorithm we use to simulate the data from YCIS model.

1. Generate covariates $X_{k,i} \sim Binomial(1, p)$ and censoring time $C_{k,i} \sim Uniform(1.5, 3)$, for $k = 1, 2, \ldots, m$ and $i = 1, 2, \ldots, n$.

2. Generate $N_{k,i} \sim Poisson(e^{\theta})$ where $\theta = \beta_0 + \beta_1 x$, for $k = 1, 2, \ldots, m$.

3. If $N_{k,i} = 0$, which means this subject is cured, then $Y_{k,i} = C_{k,i}$, for $k = 1, 2, \ldots, m$ and $i = 1, 2, \ldots, n$.

4. If $N_{k,i} \neq 0$, then latent times are generated, $Z_j \sim Exp(\lambda)$, for $j = 1, 2, \ldots, N_{k,i}$. It gives us the survival time $T_{k,i} = \min_j(Z_j)$ and the observed time $Y_{k,i} = \min(T_{k,i}, C_{k,i})$.

To generate proportional odds model with cure-rate, we follow similar steps except in step 2, we generate $N_{k,i} \sim Geo(\theta)$ where $\theta = \beta_0 + \beta_1 x$, for $k = 1, 2, \ldots, m$.

we use two models, the true model, i.e. YCIS model, and the proportional odds model with cure-rate, to fit the generated data sets. DIC and M-measure are computed for both models based on the formulas above. Here is some of the results.

In Table 4.1, the percentage of each method choosing the right model are listed.

Table 4.1: Results of simulation study for the model comparison method on survival data.

n	50	100	200	300
DIC	92	79	68	70
Mmeasure1	95	96	97	100
Mmeasure2	96	97	97	100

We used two different norms in our M-measure, the absolute value and the square function, which are called as Mmeasure1 and Mmeasure2, respectively. Four different sample sizes are tried here, $n = 50, 100, 200$, and 300. Just as it is mentioned in

Spiegelhalter, "As with AIC, DIC will not consistently select the true model from a fixed set with increasing sample sizes". As the sample size increases, the percentage of DIC choosing the right model gets lower. While for M-measure, the percentage of choosing the right model increases. The accuracy of M-measure is already very promising when the sample size is only 50, and it keeps improving when n increases. Also M-measure has a better performance in detecting the true model than DIC with all sample sizes. The percentages of M-measure are much higher than those of DIC. But different norms used in M-measure don't affect the result too much, which shows us the consistency between different norms.

4.5 Recurrent Events Data

The model comparison method we proposed is described using the notation of the counting process. Therefore we can extend it to the recurrent events data very easily.

We now describe a goodness-of-fit statistic which can be used to compare the two different link functions and different inference methods for our model with the same link function for the intensity and mean functions. Similar as the case of survival analysis, for subject i, $i = 1, \ldots, n$, our goodness-if-fit statistic is computed as the difference between the observed number of events $N_i(t)$ and the estimated mean function $\hat{\mu}(t)$, integrated over time, which has the expression

$$M = \sum_{i=1}^{n} \int_{0}^{C_i} |N_i(t) - \hat{\mu}(t)| dt, \tag{4.17}$$

where C_i is the censoring time for subject i. A smaller value for the measure indicated a better fit of the data from the model. We have the results from different link functions and different methods in Table 4.2. From the table we can see that the measure increases as the value of α increases for both link functions, which means smaller α gives a better fit when using integrated likelihood method. Both the partial likelihood method and the integrated likelihood method for small α (0.001) give a similar fit for the data since the difference of the measure is only 2%. The additive link is preferred based on the smaller measures across different α values.

Table 4.2: Goodness-of-fit results for partial likelihood estimate (PL) and integrated likelihood estimates with different α values using different link functions.

α	additive	exponentiated
PL	1033.9	1044.7
0.001	1055.3	1062.5
0.01	1061.5	1067.8
0.1	1072.3	1077.4
1	1093.4	1096.5

CHAPTER 5

DISCUSSION AND FUTURE WORK

We proposed semiparametric Bayesian model for analysing recurrent events data with the same link for intensity and mean functions. Our nonmultiplicative intensity model broads the possible models for recurrent events data. Comparison of risks for change in different covariates x can be made to provide information for the clinicians to schedule EMB after transplant. Meaningful prediction about the cumulative mean function can be given based on the intensity functions of both our models.

Empirical Bayes (integrated likelihood) method is used for solving inference problem of our heart transplant data set. In the future, we want to investigate empirical Bayesian method for our intensity models with dependent termination. Not only we want to incorporate the popular proportional hazard model for dependent termination into our model, we also want to investigate the possibility of expanding the model class for dependent termination time.

When considering the dependent termination in the model, we also need to introduce it into the goodness-of-fit statistic. A good model comparison method should be able to reveal which part of the model does not fit the data very well, the recurrent event part or the termination time part.

APPENDIX A

SECOND PARTIAL DERIVATIVES FOR THE NONMULTIPLICATIVE INTENSITY MODELS

All the second partial derivatives of the integrated likelihood function for the non-multiplicative intensity models are given here

$$\begin{aligned}
\frac{\partial^2 l}{\partial \beta^2} &= \sum_{j=1}^{M}\sum_{i=1}^{n}\left[-N_{ij}\frac{x_i^2(\alpha+r_j+B_{2j}\eta)^2}{\{\alpha\lambda(1+n_{ij}\eta)+x_i\beta(\alpha+r_j+B_{2j}\eta)\}^2}\right];\\
\frac{\partial^2 l}{\partial \eta^2} &= \sum_{j=1}^{M}\Big[\frac{\alpha\Lambda_j B_{2j}^2}{(\alpha+r_j+B_{2j}\eta)^2}-\\
&\quad \sum_{i=1}^{n} N_{ij}\frac{\alpha\lambda\{n_{ij}(\alpha+r_j)-B_{2j}\}}{\{\alpha\lambda(1+n_{ij}\eta)+x_i\beta(\alpha+r_j+B_{2j}\eta)\}^2(\alpha+r_j+B_{2j}\eta)^2}\times\\
&\quad \{(\alpha\lambda n_{ij}+2x_i\beta B_{2j})(\alpha+r_j+B_{2j}\eta)+\alpha\lambda B_{2j}(1+n_{ij}\eta)\}\Big];\\
\frac{\partial^2 l}{\partial \lambda^2} &= \sum_{j=1}^{M}\sum_{i=1}^{n}\left[-N_{ij}\frac{\alpha^2(1+n_{ij}\eta)^2}{\{\alpha\lambda(1+n_{ij}\eta)+x_i\beta(\alpha+r_j+B_{2j}\eta)\}^2}\right];\\
\frac{\partial^2 l}{\partial \beta\partial\lambda} &= \sum_{j=1}^{M}\sum_{i=1}^{n}\left[-N_{ij}\frac{\alpha(1+n_{ij}\eta)(\alpha+r_j+B_{2j}\eta)x_i}{\{\alpha\lambda(1+n_{ij}\eta)+x_i\beta(\alpha+r_j+B_{2j}\eta)\}^2}\right];\\
\frac{\partial^2 l}{\partial \beta\partial\eta} &= \sum_{j=1}^{M}\sum_{i=1}^{n}\left[-N_{ij}\frac{\alpha\lambda\{n_{ij}(\alpha+r_j)-B_{2j}\}x_i}{\{\alpha\lambda(1+n_{ij}\eta)+x_i\beta(\alpha+r_j+B_{2j}\eta)\}^2}\right];\\
\frac{\partial^2 l}{\partial \lambda\partial\eta} &= \sum_{j=1}^{M}\left[-\frac{\alpha\Delta_j B_{2j}}{\alpha+r_j+B_{2j}\eta}+\sum_{i=1}^{n}N_{ij}\frac{\alpha\{n_{ij}(\alpha+r_j)-B_{2j}\}\beta x_i}{\{\alpha\lambda(1+n_{ij}\eta)+x_i\beta(\alpha+r_j+B_{2j}\eta)\}^2}\right].
\end{aligned}$$

BIBLIOGRAPHY

[1] O. O. Aalen, J. Fosen, H. Weedon-Fekj ær, Ø Borgan, and E Husebye. Dynamic analysis of multivariate failure time data. *Biometrics*, 60:764–773, 2004.

[2] O.O. Aalen. A model for non-parametric regression analysis of counting processes. In *Lecture Notes in Statistics*, volume 2, pages 1–25. New York: Springer, 1980.

[3] H Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.

[4] P.K. Andersen and R.D. Gill. Cox's regression model for counting processes: a large-sample study. *The Annals of Statistics*, 10:1100–1200, 1982.

[5] J. Berkson and R.P. Gage. Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47:501–515, 1952.

[6] M.-H. Chen, J.G. Ibrahim, and D. Sinha. A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94:909–919, 1999.

[7] R.J. Cook and J.F. Lawless. *The statistical analysis of recurrent events*. Springer, 2007.

[8] F. Cooner, S. Banerjee, B.P. Carlin, and D. Sinha. Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*, 102:560–572, 2007.

[9] D. R. Cox. Regression models and life-tables (with discussion). *J. R. Statist. Soc.*, 34:187–220, 1972.

[10] D. R. Cox. Partial likelihood. *Biometrika*, 62:269–276, 1975.

[11] D.R. Cox and D. Oakes. *Analysis of Survival Data*. London: Chapman and Hall, 1984.

[12] M. Ewell and J.G. Ibrahim. The large sample distribution of the weighted log-rank statistic under general local alternatives. *Lifetime Data Analysis*, 3:5–12, 1997.

[13] V.T. Farewell. The use of mixture models for the analysis of survival data with long term survivors. *Biometrics*, 38:1041–1046, 1982.

[14] V.T. Farewell. Mixture models in survival analysis: Are they worth the risk? *Canadian Journal of Statistics*, 14:257–262, 1986.

[15] I. J. Good. The bayesian influence, or how to sweep subjectivism under the carpet. In C. A. Hooker and W. Harper, editors, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, volume 2, pages 125–174, 1973.

[16] I. J. Good. *the Estimation of Probabilities.* M.I.T. Press, Cambridge, Massachusetts, 2001.

[17] W.Q. Gradek, C. D'Amico, A.L. Smith, D Vega, and W.M. Book. Routine surveillance endomyocardial biopsy continues to detect significant rejection late after heart transplantation. *The Journal of Heart and Lung Transplant*, 20:803, 2001.

[18] R.J. Gray and A.A. Tsiatis. A linear rank test for use when the main interest is in differences in cure rate. *Biometrics*, 45:899–904, 1989.

[19] L.G. Hanin, M. Zaider, and A.Y. Yakovlev. Distribution of the number of clonogens surviving fractionated radiotherapy: A long-standing problem revisited. *International Journal of Radiation Biology*, 77:205–213, 2001.

[20] J.G. Ibrahim and M Chen. Prior distributions and bayesian computation for proportional hazards models. *the Indian Journal of Statistics*, 60:48–64, 1998.

[21] J.G. Ibrahim, M.-H. Chen, and D. Sinha. *Bayesian Survival Analysis.* New York: Springer-Verlag, 2001.

[22] J.G. Ibrahim, M.-H. Chen, and D. Sinha. Criterion based methods for bayesian model assessment. *Statistica Sinica*, 2001.

[23] J.D. Kalbfleisch. Nonparametric bayesian analysis of survival time data. *J.R. Statist. Soc.*, 40:214–221, 1978.

[24] J. F. Lawless, C. Nadeau, and R. J Cook. Analysis of mean and rate functions for recurrent event. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, pages 37–49. New York: Springer-Verlag, 1997.

[25] J.F. Lawless and C. Nadeau. Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37:158–168, 1995.

[26] D. Y. Lin, L. J. Wei, I. Yang, and Z. Ying. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society*, 62:711–730, 2000.

[27] R.A. Maller and S. Zhou. *Survival analysis with long-term survivors*. New York: Wiley, 1996.

[28] M. Miloslavsky, M.J. Keles, S.and Laan, and S. Butler. Recurrent events analysis in the presence of time dependent covariates and dependent censoring. *Journal of the Royal Statistical Society*, 66:239–257, 2004.

[29] D. Oakes. Frailty models for multiple event times. pages 371–379, 1992.

[30] M.S. Pepe and J. Cai. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association*, 88:811–820, 1993.

[31] D.E. Schaubel, D. Zeng, and J Cai. A semiparametric additive rates model for recurrent event data. *Lifetime Data Analysis*, 12:389–406, 2006.

[32] T.H. Scheike. The additive nonparametric and semiparametric aalen model as the rate function for a counting process. *Lifetime Data Analysis*, 10:247–262, 2002.

[33] G. E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

[34] D. Sinha. Semiparametric bayesian analysis of multiple event time data. *Journal of the American Statistical Association*, 88:979–983, 1993.

[35] D. Sinha, J. G. Ibrahim, and M.-H. Chen. A bayesian justification of cox's partial likelihood. *Biometrika*, 90:629–641, 2003.

[36] D. Sinha, T. Maiti, J.G. Ibrahim, and Ouyang B. Current methods for recurrent events data with dependent termination: a bayesian perspective. *Journal of the American Statistical Association*, 103:866–878, 2008.

[37] D. Sinha, M.B. McHenry, S.R. Lipsitz, and M. Ghosh. Empirical bayes estimation for additive hazards regression models. *Biometrika*, 96:545–558, 2009.

[38] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, 64:583–639, 2002.

[39] D.K. Stangl and J.B. Greenhouse. Assessing placebo response using bayesian hierarchical survival models. *Lifetime Data Analysis*, 4:5–28, 1998.

[40] J. Stehlik, R. Starling, M. Movsesian, J. Fang, R. Brown, M. Hess, N. Lewis, and J. Kirklin. Utility of long-term surveillance endomyocardial biopsy: A multi-institutional analysis. *The Journal of Heart and Lung Transplant*, 25:1402–1409, 2006.

[41] S.L. Tucker and J.M.G Taylor. Improved models of tumour cure. *International Journal of Radiation Biology*, 70:539–553, 1996.

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

[43] A.Y. Yakovlev. Letter to the editore. *Statistics in Medicine*, 13:983–986, 1994.

[44] A.Y. Yakovlev, B. Asselain, V.J. Bardou, A. Fourquet, T. Hoang, A. Rochefediere, and A.D. Tsodikov. A simple stochastic model of tumor recurrence and its applications to data on pre-menopausal breast cancer. *Biometrie et Analyse de Dormees Spatio-Temporelles*, 12:66–82, 1993.

[45] A.Y. Yakovlev and A.D. Tsodikov. *Stochastic models of tumor latency and their biostatistical applications*. New Jersey: World Scientific, 1996.

[46] D. Zeng and D. Y. Lin. Semiparametric transformation models with random effects for recurrent events. *Journal of the American Statistical Association*, 102:167–180, 2007.

[47] D. Zeng, G. Yin, and J.G. Ibrihim. Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*, 101:670–684, 2006.

BIOGRAPHICAL SKETCH

Yu Gu was born on August 8th, 1984, Zibo, China. She entered Peking University in China in the fall of 2002 and obtained her bachelor's degree in Applied Mathematics in the summer of 2006. She studied Statistics in Department of Statistics at Florida State University. After obtaining her Master's degree, she continued into the PhD program and defended her dissertation in the spring of 2011.

Yu's current research interest includes semiparametric methods for recurrent events data, empirical Bayesian methods for recurrent events data, Bayesian survival analysis, cure-rate model, model comparison methods for survival data and recurrent events data and clinical trials.