Biometrika (2009), **xx**, x, *pp.* 1–30

Printed in Great Britain

Semiparametric Additive Rate Model for Recurrent Events with Informative Terminal Event

BY DONGLIN ZENG AND JIANWEN CAI

Department of Biostatistics, CB# 7420, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7420, U.S.A.

dzeng@bios.unc.edu, cai@bios.unc.edu

SUMMARY

We propose a semiparametric additive rate model for modelling recurrent events in the presence of the terminal event. The dependence between recurrent events and terminal event is fully nonparametric and is due to some latent process in the baseline rate function. Additionally, a general transformation model is used to model the terminal event given covariates. We construct an estimating equation for parameter estimation. The asymptotic distributions of the proposed estimators are derived. Simulation studies demonstrate that the proposed inference procedure performs well in realistic settings. Application to a medical study is presented.

Some key words: Additive rate model; Estimating equation; Recurrent event; Terminal event; Transformation models.

1. INTRODUCTION

Recurrent events are common in medical practice or epidemiologic studies when each subject experiences a particular event repeatedly over time. Examples of recurrent events include multiple infection episodes, tumor recurrences, and repeated drug use. Interest of recurrent event analysis usually focuses on identifying risk factors which may elevate or decrease the frequencies of recurrent events.

In most practices, recurrent event times are subject to censoring. One typical censoring is caused by the termination of the follow-up due to the subject's death. Such terminating censorship is very likely informative about the recurrent events so it should be accounted for in the analysis. In the literature, most of the existing methods on recurrent event analysis (e.g., Andersen and Gill, 1982; Prentice, Williams and Peterson, 1981; Wei, Lin and Weissfeld, 1989) require non-informative censorship and may yield misleading results when recurrent event times are actually informatively censored. Recently, jointly modelling both recurrent events and terminal event through shared frailty or random-effects have been developed. Such joint models attribute the association between the two types of events to some latent effects, which are included in the regression models either as frailty or random effects. For example, Wang, Qin and Chiang (2001) and Huang and Wang (2004) studied a shared frailty model with proportional intensity and proportional hazards assumptions for recurrent events and the terminal event, respectively. The model allows an unknown distribution for the shared frailty. Liu, Wolfe and Huang (2004) considered the same model but assumed a gamma frailty distribution. In a recent paper, Zeng and Lin (2009) studied the general transformation models in this joint modelling approach. For all these joint modelling approaches, one strong assumption is that the dependence between the recurrent events and the terminal event is modelled via an explicit and parametric latent effect, which may not be true in practice. The computation involved in the joint modelling approach is usually intensive.

Compared to the intensity models used in the joint modelling approaches mentioned above, rate models have also been popular in analyzing recurrent events because the regression coefficients reflect the covariate effects on the frequency of the recurrent events which is practically more intuitive. Examples include the proportional rate model or its transformed form as proposed by Pepe and Cai (1993), Lawless and Nadeau (1995) and Lin, Wei, Yang and Ying (2000). All these models assume the effect of the covariates to be multiplicative and the non-informative censoring. Work on extension to incorporating the informative terminal event is limited: Cook and Lawless (1997) studied the mean and rate of the recurrent events among survivors at certain time points. Ghosh and Lin (2000) proposed an nonparametric estimator for the rate function of the recurrent event by incorporating the survival probabilities of the terminal event. They further considered the proportional rate model with covariates in Ghosh and Lin (2002), where the inverse probability weighted estimating equation was used to obtain the consistent estimators for the regression coefficients. An expanded version of the same type of the inverse weighted estimating equation was adopted to improve the efficiency in Miloslavsky et al (2004) for the proportional rate model.

A useful and important alternative to the proportional rate model is the additive rate model, where the true underlying covariate effects may add to, rather than multiply, the baseline event rate. As pointed out in Schaubel et al (2006), in many practical applications, an additive model may indeed be more appropriate, particularly with respect to continuous covariates. In situations where the additive and multiplicative models fit the data equally well, the additive model may be preferred due to the interpretation of the regression parameter. For the additive rate model as given in Lin and Ying (1994), no work has been done to incorporate the informative terminal event.

In this paper, we focus on the additive rate model for recurrent events. Only covariates of interest are parametrically modelled as an additive component in this model. In our additive model, the baseline rate function is nonparametric and depends on some latent random variables which are associated with the terminal event. However, such an association is fully nonparametric. A general transformation model (Zeng and Lin, 2006) is used for modelling terminal event.

2. Models and Inference

2·1. *Models*

Let $N(t)$ denote the counting process associated with recurrent event and let T denote the terminal event time. The covariates of interest are denoted by X. For the terminal event time T, we assume the following linear transformation model

$$\Lambda(t|X) = G(e^{-X^T\beta}\Lambda(t)), \tag{1}$$

where $\Lambda(t|X)$ is the conditional hazard function of T given X, $\Lambda(\cdot)$ is an unknown and monotone transformation with $\Lambda(0) = 0$ and G is a given transformation function. The usual proportional hazards model and the proportional odds model are both special cases of the linear transformation model with $G(x) = x$ and $G(x) = \log(1 + x)$. Note that model (1) is equivalent to

$$\log \Lambda(T) = X^T\beta + \epsilon,$$

where ϵ is an independent error following a distribution with cumulative density function $1 - e^{-G(e^\epsilon)}$. For the recurrent event process, we let ν be subject-specific latent effect which is independent of X and may be associated with the terminal event residual ϵ. For any time t, given ν and $T > t$, we assume that the rate of the recurrent event at time t is independent of T. Furthermore, we model this rate function of the recurrent event process via an additive model by assuming

$$E[dN(t)|X, T > t, \nu] = I(T > t)\left\{dR(t, \nu) + X^T\gamma dt\right\}, \tag{2}$$

where $R(t,\nu)$ is the subject-specific baseline cumulative rate function and assumed to be unknown. Moreover, $R(0,\nu) = 0$ and $R(t,\nu)$ is an increasing function of t for $t \leq T$. Particularly, the parameter γ represents the rate difference for one unit change in X for a given subject-specific latent effect ν. The latent effect ν explains the dependence between the recurrent event process and the terminal event.

2·2. *Inference Procedure*

Suppose that we observed data from n i.i.d subjects subject to right censoring. We denote them as

$$Y_i = T_i \wedge C_i, \;\; \Delta_i = I(T_i \leq C_i)$$

and $(N_i(t), t \leq Y_i)$ for $i = 1, ..., n$, where C_i is censoring time for subject i, $T_i \wedge C_i$ is the minimum of T_i and C_i, and $I(T_i \leq C_i)$ is the failure indicator. We assume that the right-censoring is noninformative satisfying that C_i is independent of ν, $N_i(t)$ and T_i given X_i.

Our goal is to estimate β and γ. First, we use the survival data $(Y_i, \Delta_i, X_i), i = 1, ..., n,$ to estimate the parameters in model (1). Particularly, the nonparametric maximum likelihood estimation approach (Zeng and Lin, 2006) is used to derive the estimates for β and Λ and we denote the estimates as $\widehat{\beta}$ and $\widehat{\Lambda}$ respectively. That is, $\widehat{\beta}$ and $\widehat{\Lambda}$ maximize

$$\prod_{i=1}^{n} \left[\left\{ \Lambda\{Y_i\} e^{-X_i^T\beta} G'(\Lambda(Y_i) e^{-X_i^T\beta}) \right\}^{\Delta_i} \exp\left\{ -G(\Lambda(Y_i) e^{-X_i^T\beta}) \right\} \right],$$

where $\Lambda\{t\}$ denotes the jump size of Λ at t. The details of computing $\widehat{\beta}$ and $\widehat{\Lambda}$ can be found in Zeng and Lin (2006).

To estimate γ, since T can be censored, we may not be able to estimate the rate function given T directly; instead, we need to consider the observed rate function given the observed end point

Y. From model (2), we have

$$E[dN(t)|X, Y > t] = I(Y > t)\left\{dE[R(t,\nu)|X, Y > t] + X^T\gamma dt\right\}.$$

Since C is independent of ν and T given X,

$$E[R(t,\nu)|X, Y > t] = E[R(t,\nu)|X, T > t] = E[R(t,\nu)|X, \epsilon > \log\Lambda(t) - X^T\beta].$$

Following the assumption that (ϵ, ν) are independent of X, we obtain

$$E[dN(t)|X, Y > t] = I(Y > t)\left\{dE[R(t,\nu)|\epsilon > s]\Big|_{s=\log\Lambda(t)-X^T\beta} + X^T\gamma dt\right\}. \qquad (3)$$

Thus, if define $dH(t,s)$ as $E[dR(t,\nu)|\epsilon > s]$, then it is necessary to be able to estimate $dH(t,s)$ using the observed data. Note that from the fact (ν,ϵ) is independent of X and C, we have

$$E[dR(t,\nu)|\epsilon > s] = \frac{E[dR(t,\nu)I(\epsilon > s)]}{E[I(\epsilon > s)]} = \frac{E[dR(t,\nu)I(\Lambda(Y)e^{-X^T\beta} > e^s)g(X)]}{E[I(\Lambda(Y)e^{-X^T\beta} > e^s)g(X)]}$$

for any integrable function $g(X)$. Particularly, we choose $g(X)$ to be of the form $I(X^T\beta \geq \log\Lambda(t) - s)$ so that both $\Lambda(Y)e^{-X^T\beta} > e^s$ and $X^T\beta \geq \log\Lambda(t) - s$ implies $Y > t$. Then,

$$E[dR(t,\nu)|\epsilon > s] = \frac{E[(dN(t) - X^T\gamma dt)I(\Lambda(Y)e^{-X^T\beta} > e^s, X^T\beta \geq \log\Lambda(t) - s)]}{E[I(\Lambda(Y)e^{-X^T\beta} > e^s, X^T\beta \geq \log\Lambda(t) - s)]}.$$

Hence, we can estimate $dH(t,s)$ using the empirical observations as

$$d\widehat{H}(t,s) \equiv \frac{\sum_{j=1}^n (dN_j(t) - X_j^T\gamma dt)I(\widehat{\Lambda}(Y_j)e^{-X_j^T\widehat{\beta}} > e^s, X_j^T\widehat{\beta} \geq \log\widehat{\Lambda}(t) - s)}{\sum_{j=1}^n I(\widehat{\Lambda}(Y_j)e^{-X_j^T\widehat{\beta}} > e^s, X_j^T\widehat{\beta} \geq \log\widehat{\Lambda}(t) - s)}.$$

From (3), this implies that the following term

$$I(Y_i > t)\left\{dN_i(t) - d\widehat{H}(t, \log\widehat{\Lambda}(t) - X_i^T\widehat{\beta}) - X_i^T\gamma dt\right\}$$

has mean approximating zero given X_i; equivalently, if define

$$d\overline{N}_i(t) = \frac{\sum_{j=1}^n dN_j(t) I(\widehat{\Lambda}(Y_j) e^{-X_j^T \widehat{\beta}} > \widehat{\Lambda}(t) e^{-X_i^T \widehat{\beta}}, X_j^T \widehat{\beta} \geq X_i^T \widehat{\beta})}{\sum_{j=1}^n I(\widehat{\Lambda}(Y_j) e^{-X_j^T \widehat{\beta}} > \widehat{\Lambda}(t) e^{-X_i^T \widehat{\beta}}, X_j^T \widehat{\beta} \geq X_i^T \widehat{\beta})}$$

and

$$\overline{X}_i(t) = \frac{\sum_{j=1}^n X_j I(\widehat{\Lambda}(Y_j) e^{-X_j^T \widehat{\beta}} > \widehat{\Lambda}(t) e^{-X_i^T \widehat{\beta}}, X_j^T \widehat{\beta} \geq X_i^T \widehat{\beta})}{\sum_{j=1}^n I(\widehat{\Lambda}(Y_j) e^{-X_j^T \widehat{\beta}} > \widehat{\Lambda}(t) e^{-X_i^T \widehat{\beta}}, X_j^T \widehat{\beta} \geq X_i^T \widehat{\beta})},$$

then

$$I(Y_i > t) \left\{ dN_i(t) - d\overline{N}_i(t) - (X_i - \overline{X}_i(t))^T \gamma dt \right\}$$

is approximately zero for given X_i.

Hence, to estimate γ, we propose the following estimating equation for inference:

$$\sum_{i=1}^n \int \omega(t) I(Y_i > t)(X_i - \overline{X}_i(t)) \left\{ dN_i(t) - d\overline{N}_i(t) - (X_i - \overline{X}_i(t))^T \gamma dt \right\} = 0,$$

where $\omega(t)$ is any deterministic weight function. Equivalently, the estimator for γ, denoted as $\widehat{\gamma}$, is given as

$$\left[\sum_{i=1}^n \int I(Y_i > t) \omega(t) (X_i - \overline{X}_i(t))^{\otimes 2} dt \right]^{-1} \left[\sum_{i=1}^n \int I(Y_i \geq t) \omega(t) (X_i - \overline{X}_i(t)) \left\{ dN_i(t) - d\overline{N}_i(t) \right\} \right]. \tag{4}$$

Note that there is some possibility that the denominator in the calculation of $d\overline{N}_i(t)$ and $\overline{X}_i(t)$, i.e.,

$$\sum_{j=1}^n I(\widehat{\Lambda}(Y_j) e^{-X_j^T \widehat{\beta}} > \widehat{\Lambda}(t) e^{-X_i^T \widehat{\beta}}, X_j^T \widehat{\beta} \geq X_i^T \widehat{\beta}),$$

could be zero. In this case, we define $0/0$ as zero so that the corresponding $d\overline{N}_i(t)$ and $\overline{X}_i(t)$ are zeros.

2·3. *Extension to time-dependent covariates*

Our model and inference method can be extended to incorporate external time-dependent covariates $X(t)$ in the above formulation. Particularly, when $X(t)$ is time-dependent, the transformation model (1) for the terminal event becomes

$$\Lambda(t|X) = G\left(\int_0^t e^{-X(s)^T\beta} d\Lambda(s)\right),$$

where $\Lambda(t|X)$ is the conditional hazard function of T given X. The above model is also equivalent to

$$\log \int_0^T e^{-X(s)^T\beta} d\Lambda(s) = \epsilon,$$

where ϵ is independent of X with cumulative density function $1 - \exp\{-G(e^\epsilon)\}$. Thus, if we re-define $d\overline{N}_i(t)$ as

$$\frac{\sum_{j=1}^n dN_j(t) I(\int_0^{Y_j} e^{-X_j(s)^T\widehat{\beta}} d\widehat{\Lambda}(s) > \int_0^t e^{-X_i(s)^T\widehat{\beta}} d\widehat{\Lambda}(s), \int_0^t e^{-X_j(s)^T\widehat{\beta}} d\widehat{\Lambda}(s) \le \int_0^t e^{-X_i(s)^T\widehat{\beta}} d\widehat{\Lambda}(s))}{\sum_{j=1}^n I(\int_0^{Y_j} e^{-X_j(s)^T\widehat{\beta}} d\widehat{\Lambda}(s) > \int_0^t e^{-X_i(s)^T\widehat{\beta}} d\widehat{\Lambda}(s), \int_0^t e^{-X_j(s)^T\widehat{\beta}} d\widehat{\Lambda}(s) \le \int_0^t e^{-X_i(s)^T\widehat{\beta}} d\widehat{\Lambda}(s))}$$

and redefine $\overline{X}_i(t)$ as

$$\frac{\sum_{j=1}^n X_j(t) I(\int_0^{Y_j} e^{-X_j(s)^T\widehat{\beta}} d\widehat{\Lambda}(s) > \int_0^t e^{-X_i(s)^T\widehat{\beta}} d\widehat{\Lambda}(s), \int_0^t e^{-X_j(s)^T\widehat{\beta}} d\widehat{\Lambda}(s) \le \int_0^t e^{-X_i(s)^T\widehat{\beta}} d\widehat{\Lambda}(s))}{\sum_{j=1}^n I(\int_0^{Y_j} e^{-X_j(s)^T\widehat{\beta}} d\widehat{\Lambda}(s) > \int_0^t e^{-X_i(s)^T\widehat{\beta}} d\widehat{\Lambda}(s), \int_0^t e^{-X_j(s)^T\widehat{\beta}} d\widehat{\Lambda}(s) \le \int_0^t e^{-X_i(s)^T\widehat{\beta}} d\widehat{\Lambda}(s))},$$

then an estimator for γ is given similar to (4) as

$$\left[\sum_{i=1}^n \int I(Y_i > t)\omega(t)(X_i(t) - \overline{X}_i(t))^{\otimes 2} dt\right]^{-1} \left[\sum_{i=1}^n \int I(Y_i \ge t)\omega(t)(X_i(t) - \overline{X}_i(t)) \left\{dN_i(t) - d\overline{N}_i(t)\right\}\right].$$

3. ASYMPTOTIC RESULTS

We provide the asymptotic results for the estimators $(\widehat{\beta}, \widehat{\Lambda})$ and $\widehat{\gamma}$, assuming X and its effect to be time-independent. The same results apply to the case when X contains time-dependent components. We need the following assumptions.

(C.1) The true parameter β_0 belongs to a known compact set and the hazards function $\Lambda_0(t)$ is continuously differentiable and strictly increasing in $[0, \tau]$, where τ is the study duration and assumed to be finite.

(C.2) Covariates X are bounded and satisfy the following condition: if $\alpha_0 + \alpha_1^T X = 0$ with probability one, then $\alpha_0 = 0$ and $\alpha_1 = 0$.

(C.3) Transformation function $G(x)$ is three-times continuously differentiable and strictly increasing. Moreover, there exists a positive constant ρ_0 such that

$$\lim \sup_{x\to\infty} (1 + x)^{\rho_0} e^{-G(x)} < \infty, \ \ \lim \sup_{x\to\infty} (1 + x)^{1+\rho_0} G'(x) e^{-G(x)} < \infty.$$

(C.4) There exists some positive constant δ_0 such that $P(C \geq \tau | X) > \delta_0$.

The conditions in both (C.1) and (C.4) are standard in the practice of survival analysis context. Condition (C.2) is equivalent to saying that the design matrix $[1, X]$ is full rank with some positive probability. Condition (C.3) stipulates the tail behavior of the transformation function $G(x)$. It is easy to check that transformations $G(x) = \rho^{-1}\{(1 + x)^\rho - 1\}$ for $\rho \geq 0$ and $G(x) = r^{-1} \log(1 + rx)$ for $r \geq 0$ satisfy this condition. The same condition is used in Zeng and Lin (2006) for transformation models.

The first result concerns the asymptotic distribution of $(\widehat{\beta}, \widehat{\Lambda})$, which has been given in Zeng and Lin (2006). We quote this result in the following theorem.

Theorem 1 (from Zeng and Lin, 2006). Under conditions (C.1)-(C.4), $(\widehat{\beta}, \widehat{\Lambda})$ are strongly consistent in the sense

$$|\widehat{\beta} - \beta_0| + \sup_{t\in[0,\tau]} |\widehat{\Lambda}(t) - \Lambda_0(t)| \to_{a.s.} 0;$$

moreover, $n^{1/2}(\widehat{\beta}-\beta_0,\widehat{\Lambda}-\Lambda_0)$ converges in distribution to a tight Gaussian process in the metric space $R^d \times l^\infty[0,\tau]$, where d is the dimension of β_0 and $l^\infty[0,\tau]$ consists all the bounded function in $[0,\tau]$ equipped with the supreme norm.

Furthermore, according to Zeng and Lin (2006), we have the following asymptotic linear expansion for $\widehat{\beta}$ and $\widehat{\Lambda}$:

$$n^{1/2}(\widehat{\beta}-\beta_0) = \mathcal{G}_n S_\beta(Y,\Delta,X;\beta_0,\Lambda_0) + o_p(1),$$

$$n^{1/2}(\widehat{\Lambda}(t)-\Lambda_0(t)) = \mathcal{G}_n S_\Lambda(Y,\Delta,X,t;\beta_0,\Lambda_0) + o_p(1), \tag{5}$$

where S_β and S_Λ are the respective influence function for $\widehat{\beta}$ and $\widehat{\Lambda}$, $\mathcal{G}_n$ is the empirical process defined as $n^{1/2}(\mathcal{P}_n-\mathcal{P})$ with $\mathcal{P}_n$ being the empirical measure and $\mathcal{P}$ being its expectation, and $o_p(1)$ denotes the random element converging to zero in probability in the metric space of Theorem 1. Moreover, using the consistent estimator of the information matrix for $\widehat{\beta}$ and $\widehat{\Lambda}$ as given in Zeng and Lin (2006), we can estimate S_β and S_Λ consistently in the uniform sense of (Y,Δ,X) and $t\in[0,\tau]$; so we denote such estimators as $\widehat{S}_\beta$ and $\widehat{S}_\Lambda$ respectively.

The following theorem gives the asymptotic distribution for $\widehat{\gamma}$.

Theorem 2. Under conditions (C.1)-(C.4),

$$n^{1/2}(\widehat{\gamma}-\gamma_0) = \mathcal{G}_n S_\gamma(N,Y,\Delta,X;\beta_0,\gamma_0,\Lambda_0) + o_p(1),$$

where S_γ is the mean-zero influence function for $\widehat{\gamma}$ and is given in the appendix. As the result, $n^{1/2}(\widehat{\gamma}-\gamma_0)$ converges in distribution to a mean-zero Gaussian distribution with variance $\Sigma_\gamma = Var(S_\gamma)$.

We need to estimate the asymptotic covariance of $\widehat{\gamma}$. However, since S_γ is complicated and involves the Hadamard derivatives in the metric space of Theorem 1, direct estimation of S_γ is not feasible. Therefore, we propose the following Monte-Carlo method: from the proof of Theorem 2, we note that in the expression (4), $\widehat{\gamma}$'s variation only comes from the term $N_i(t) - \overline{N}_i(t)$ and the variation in the empirical summations in the numerator and denominator of $\overline{N}_i(t)$, as well as the plug-in estimator $(\widehat{\beta}, \widehat{\Lambda})$. Therefore, we wish to use the Monte-Carlo method to capture all these variations.

Specifically, we generate n i.i.d random variables $\mathcal{Z}_1, ..., \mathcal{Z}_n$ from the standard normal distribution. Then the contribution to $\widehat{\gamma}$'s variation due to $N_i(t) - \overline{N}_i(t)$ in expression (4) is equivalent to the variation of the following function of $(\mathcal{Z}_1, ..., \mathcal{Z}_n)$,

$$\Omega_1 = \left[\sum_{i=1}^n \int I(Y_i > t)\omega(t)(X_i - \overline{X}_i(t))^{\otimes 2} dt\right]^{-1} \times$$

$$\left[\sum_{i=1}^n \mathcal{Z}_i \int I(Y_i \geq t)\omega(t)(X_i - \overline{X}_i(t)) \left\{dN_i(t) - d\overline{N}_i(t)\right\}\right],$$

given the observed data. The contribution due to the numerator and denominator of $\overline{N}_i(t)$ is equivalent to

$$\Omega_2 = \left[\sum_{i=1}^n \int I(Y_i > t)\omega(t)(X_i - \overline{X}_i(t))^{\otimes 2} dt\right]^{-1} \left[\sum_{i=1}^n \int I(Y_i \geq t)\omega(t)(X_i - \overline{X}_i(t)) \times\right.$$

$$\left\{ - \frac{\sum_{j=1}^n \mathcal{Z}_j (dN_j(t) - X_j^T \widehat{\gamma} dt) I(\widehat{\Lambda}(Y_j) e^{-X_j^T \widehat{\beta}} > \widehat{\Lambda}(t) e^{-X_i^T \widehat{\beta}}, X_j^T \widehat{\beta} \geq X_i^T \widehat{\beta})}{\sum_{j=1}^n I(\widehat{\Lambda}(Y_j) e^{-X_j^T \widehat{\beta}} > \widehat{\Lambda}(t) e^{-X_i^T \widehat{\beta}}, X_j^T \widehat{\beta} \geq X_i^T \widehat{\beta})}\right.$$

$$+ \frac{\sum_{j=1}^n (dN_j(t) - X_j^T \widehat{\gamma} dt) I(\widehat{\Lambda}(Y_j) e^{-X_j^T \widehat{\beta}} > \widehat{\Lambda}(t) e^{-X_i^T \widehat{\beta}}, X_j^T \widehat{\beta} \geq X_i^T \widehat{\beta})}{\left(\sum_{j=1}^n I(\widehat{\Lambda}(Y_j) e^{-X_j^T \widehat{\beta}} > \widehat{\Lambda}(t) e^{-X_i^T \widehat{\beta}}, X_j^T \widehat{\beta} \geq X_i^T \widehat{\beta})\right)^2}$$

$$\left.\left.\left(\sum_{j=1}^n \mathcal{Z}_j I(\widehat{\Lambda}(Y_j) e^{-X_j^T \widehat{\beta}} > \widehat{\Lambda}(t) e^{-X_i^T \widehat{\beta}}, X_j^T \widehat{\beta} \geq X_i^T \widehat{\beta})\right)\right\}\right].$$

Finally, to account for the variation in estimating β and Λ, we generate

$$\widetilde{\beta} = \widehat{\beta} + \frac{1}{n}\sum_{i=1}^{n} \mathcal{Z}_i \widehat{S}_\beta(Y_i, \Delta_i, X_i), \quad \widetilde{\Lambda}(t) = \widehat{\Lambda}(t) + \frac{1}{n}\sum_{i=1}^{n} \mathcal{Z}_i \widehat{S}_\Lambda(Y_i, \Delta_i, X_i, t).$$

We then obtain

$$\Omega_3 = \left[\sum_{i=1}^{n} \int I(Y_i > t)\omega(t)(X_i - \overline{X}_i(t))^{\otimes 2} dt\right]^{-1}$$

$$\times \left[\sum_{i=1}^{n} \int I(Y_i \geq t)\omega(t)(X_i - \overline{X}_i(t)) \left\{dN_i(t) - d\widetilde{N}_i(t)\right\}\right],$$

where $\widetilde{N}_i(t)$ is defined the same way as $\overline{N}_i(t)$ except that $(\widehat{\beta}, \widehat{\Lambda})$ is replaced with $(\widetilde{\beta}, \widetilde{\Lambda})$. Thus, intuitively, the pure variation due to $(\widehat{\beta}, \widehat{\Lambda})$ is reflected in $\Omega_3 - \widehat{\gamma}$.

We combine all these together and obtain one statistic

$$\widetilde{\gamma} = \Omega_1 + \Omega_2 + \Omega_3.$$

We repeat such Monte-Carlo method a number of times. The sample variation of these generated statistics $\{\widetilde{\gamma}\}$ is considered as an estimator for the asymptotic covariance of $\widehat{\gamma}$.

The following theorem justifies the validity of the above Monte-Carlo method, whose proof is given in the appendix.

Theorem 3. Let $E_{\mathcal{Z}}$ denote the conditional expectation with respect to $\mathcal{Z}_1, ..., \mathcal{Z}_n$ given the observed data. Then

$$E_{\mathcal{Z}}\left[(\widetilde{\gamma} - \widehat{\gamma})^{\otimes 2}\right] \to_p \Sigma_\gamma.$$

The proof of Theorem 2 utilizes the theory of empirical process and Theorem 1. Particularly, we expand $n^{1/2}(\widehat{\gamma} - \gamma_0)$ linearly as the summation of independent components. The proof of Theorem 3 is in the same spirit as of Theorem 2. All the details are given in the appendix.

4. Partly Linear Additive Risk Model

In this section, we consider an even more general model for the recurrent events called partly parametric additive risk model. In this model, we allow some covariates to have time-dependent effects but other covariates to have linear effects. Specifically, let W and Z denote those covariates whose effects are time-dependent and linear respectively and $X = (W, Z)$. Then a partly linear additive risk model for the recurrent events assumes

$$E[dN(t)|X, T > t, \nu] = I(T > t)\left\{dR(t, \nu) + W^T\alpha(t)dt + Z^T\theta dt\right\}$$

where the parameter $\alpha(t)$ is an unknown function of t. Such a model is similar to the partly parametric additive model proposed in McKeague and Sasieni (1994) but we allow the baseline function to depend on an unknown latent effect which is also associated with the terminal event T.

We can apply the same idea as in Section 2 to estimate $\alpha(t)$ and θ. Particularly, a similar equation to (3) holds:

$$E[dN(t)|X, Y > t] = I(Y > t)\left\{dH(t, \log\Lambda(t) - X^T\beta) + W^T\alpha(t)dt + Z^T\theta dt\right\}.$$

Again, $dH(t, s)$ can be estimated using the empirical observations as

$$d\widehat{H}(t, s) \equiv \frac{\sum_{j=1}^n (dN_j(t) - W_j^T\alpha(t)dt - Z_j^T\theta dt) I(\widehat{\Lambda}(Y_j)e^{-X_j^T\widehat{\beta}} > e^s, X_j^T\widehat{\beta} \geq \log\widehat{\Lambda}(t) - s)}{\sum_{j=1}^n I(\widehat{\Lambda}(Y_j)e^{-X_j^T\widehat{\beta}} > e^s, X_j^T\widehat{\beta} \geq \log\widehat{\Lambda}(t) - s)}.$$

Therefore, this implies that

$$I(Y_i > t)\left\{dN_i(t) - d\widehat{H}(t, \log\widehat{\Lambda}(t) - X_i^T\widehat{\beta}) - W_i^T\alpha(t)dt - Z_i^T\theta dt\right\}$$

has mean approximating zero given X_i. If define $\overline{N}_i(t)$, $\overline{W}_i(t)$ and $\overline{Z}_i(t)$ similarly as before, we conclude that

$$I(Y_i > t)\left\{dN_i(t) - d\overline{N}_i(t) - (W_i - \overline{W}_i(t))^T\alpha(t)dt - (Z_i - \overline{Z}_i(t))^T\theta dt\right\}$$

is approximately zero for given X_i.

Hence, we propose the following estimating equations to estimate $\alpha(t_0)$ for any t_0 and θ:

$$\sum_{i=1}^n \int K_{a_n}(t-t_0)I(Y_i > t)(W_i - \overline{W}_i(t))\left\{dN_i(t) - d\overline{N}_i(t) - (W_i - \overline{W}_i(t))^T\alpha(t_0)dt\right.$$
$$\left.-(Z_i - \overline{Z}_i(t))^T\theta dt\right\} = 0, \quad (6)$$

and

$$\sum_{i=1}^n \int I(Y_i > t)(Z_i - \overline{Z}_i(t))\left\{dN_i(t) - d\overline{N}_i(t) - (W_i - \overline{W}_i(t))^T\alpha(t)dt - (Z_i - \overline{Z}_i(t))^T\theta dt\right\} = 0, \quad (7)$$

where $K_{a_n}(t) = a_n^{-1}K(t/a_n)$ with $K(\cdot)$ being a symmetric kernel function and a_n being a bandwidth. Solving (6) yields

$$\hat{\alpha}(t_0;\theta) = \Sigma_{WW}(t_0)^{-1}\left\{\Sigma_{WN}(t_0) - \Sigma_{WZ}(t_0)\theta\right\},$$

where

$$\Sigma_{WW}(t_0) = \sum_{i=1}^n \int K_{a_n}(t-t_0)I(Y_i > t)(W_i - \overline{W}_i(t))^{\otimes 2}dt,$$

$$\Sigma_{WN}(t_0) = \sum_{i=1}^n \int K_{a_n}(t-t_0)I(Y_i \geq t)(W_i - \overline{W}_i(t))\left\{dN_i(t) - d\overline{N}_i(t)\right\},$$

and

$$\Sigma_{WZ}(t_0) = \sum_{i=1}^{n} \int K_{a_n}(t-t_0) I(Y_i \geq t)(W_i - \overline{W}_i(t))(Z_i - \overline{Z}_i(t))^T dt.$$

After substituting this into equation (7), we obtain that the estimator for θ is given as

$$\hat{\theta} = \left[\sum_{i=1}^{n} \int I(Y_i \geq t) \left\{(Z_i - \overline{Z}_i(t))^{\otimes 2} - (Z_i - \overline{Z}_i(t))(W_i - \overline{W}_i(t))^T \Sigma_{WW}(t)^{-1} \Sigma_{WZ}(t)\right\} dt\right]^{-1}$$

$$\times \left[\sum_{i=1}^{n} \int I(Y_i \geq t)(Z_i - \overline{Z}_i(t)) \left\{dN_i(t) - d\overline{N}_i(t) - (W_i - \overline{W}_i(t))^T \Sigma_{WW}(t)^{-1} \Sigma_{WN}(t) dt\right\}\right].$$

The estimator for $\alpha(t)$ is then given as $\hat{\alpha}(t; \hat{\theta})$.

Notice that the expression of $\hat{\theta}$ takes a similar expression as $\hat{\gamma}$ in (4), except that additional projections on the covariate W-space are subtracted from both Z and $dN(t)$. Therefore, under some regularity conditions and assuming $na_n \to \infty$ and $na_n^4 \to 0$, following the similar arguments as proving Theorem 2, we can show that $\hat{\theta}$ is consistent and $n^{1/2}(\hat{\theta} - \theta_0)$ converges in distribution to a mean-zero normal distribution. Moreover, the estimator for $\alpha(t)$ can be shown to be point-wise consistent and asymptotically normal.

5. SIMULATION STUDIES

We conduct simulation studies to examine the performance of the proposed method. In the simulation studies, for each subject i, we generate two covariates with X_{1i} from a Bernoulli distribution with success probability 0.5 and X_{2i} from the uniform distribution in $[0, 1]$. To generate the terminal event, we use the transformation model

$$\log \frac{T_i}{2} = X_{1i} - 0.5 X_{2i} + \epsilon_i.$$

Thus, the true cumulative hazards function $\Lambda_0(t) = t/2$ and the corresponding $\beta_0 = (1, -0.5)^T$. Furthermore, we generate ϵ from the extreme-value distribution so the model for the terminal event is the proportional hazards model.

To generate the recurrent events, we use the following intensity model:

$$\lambda_i(t) = \xi_i I(T_i > t) \left\{ 0.5 - \psi_0 \exp(\nu_i)/\log \epsilon_i + 0.5X_{1i} + 0.8X_{2i} \right\},$$

where $\lambda_i(t)$ denotes the intensity function at time t for subject i, ξ_i is generated independently from a Gamma-distribution with mean 1 and variance 0.5, and ν_i is independently generated from the uniform distribution in $[0, 1]$. Additionally, the coefficient ψ_0 is a given constant. Clearly, this intensity model implies the following rate model

$$E[dN_i(t)|X_{1i}, X_{2i}, \nu_i, \epsilon_i] = I(T_i > t) \left\{ 0.5 - \psi_0 \exp(\nu_i)/\log \epsilon_i + 0.5X_{1i} + 0.8X_{2i} \right\} dt.$$

Thus, the corresponding coefficient $\gamma_0 = (0.5, 0.8)^T$. The first component $-\psi_0 \exp(\nu_i)/\log \epsilon_i$ reflects the dependence between the rate of the recurrent events and the terminal event. Particularly, when $\psi_0 = 0$, we obtain the situation when the terminal event is non-informative of the recurrent events; when ψ_0 is non-zero, this implies the informativeness of the terminal event. For the latter, we choose $\psi_0 = 1$ in the simulations. Finally, the right-censoring time is generated from the minimum of the uniform distribution in $[1.5, 8]$ and 3, which yields 35% censoring. The average number of the recurrent events per subjects is around 3 to 3.5.

For each simulated data, we first implement the algorithm in Zeng and Lin (2006) to estimate β and Λ as well as their influence functions. The estimator for γ is obtained using the formula (4). The procedure based on the Monte-Carlo resampling method, which was given in the previous section, is used to estimate the asymptotic covariance. Particularly, we use 100 Monte-Carlo samples and find the variance estimation to be fairly accurate. The following two tables summarize the results from sample sizes $n = 100, 200$ and 400, with Table 1 from the simulations corresponding to $\psi_0 = 1$ and Table 2 from the simulations corresponding to $\psi_0 = 0$. In the tables, column "Bias" is the average bias from 1000 repetitions; "SE" is the sample standard deviation of the empirical estimates; "ESE" is the average value of the estimated standard errors obtained from the resampling approach; "CP" is the coverage probability of the 95% confidence interval based on the normal approximation. The results indicate that the biases of the estimators are small and decrease quickly with the increasing sample sizes; the estimated standard errors are reasonably close to the empirical standard errors; the confidence intervals all have reasonable nominal levels.

For comparison, we also report the results by treating the terminal event as non-informative; that is, we estimate the effects of the covariates on the recurrent event rate by fitting a simple additive rate model as follows:

$$E[dN(t)|T > t, X] = I(T > t)(dR(t) + X^T \gamma dt).$$

Such naive estimators can be obtained using the same expression (4) except that we set $\hat{\beta} = 0$ and $\hat{\Lambda}(Y) = Y$. Note that our model (2) does not reduce to this model. As expected, the naive estimators treating the terminal event as non-informative can have very large bias when the recurrent events and the terminal event are actually dependent due to some latent process (i.e., $\psi_0 = 1$) while its bias is small when there are no such dependence (i.e., $\psi_0 = 0$). From the simulation studies, when the recurrent event is independent of the terminal event, our estimators generally have larger variance than the naive estimators, mainly because the latter utilizes the independence information in estimation. However, under the situation when the two types of events are actually dependent ($\psi_0 = 1$), the naive estimator produce large bias while our estimator is still approximately unbiased. The ratios between the mean square errors from our method and the

Table 1. *Simulation Results from 1000 Repetitions with Non-informative Terminal Events*

			Our approach				Naive	
n	Par.	True	Bias ($\times 10^{-2}$)	SE ($\times 10^{-2}$)	ESE ($\times 10^{-2}$)	CP ($\times 10^{-2}$)	Bias ($\times 10^{-2}$)	SE ($\times 10^{-2}$)
100	β_1	1.0	2.6	26.2	26.3	94.6	-	-
	β_2	-0.5	-0.7	45.4	43.6	94.8	-	-
	γ_1	0.5	2.5	24.5	26.8	96.2	0.5	20.4
	γ_2	0.8	2.9	40.2	41.1	95.0	0.9	40.5
200	β_1	1.0	0.1	18.2	18.4	94.7	-	-
	β_2	-0.5	-0.8	31.9	30.4	94.0	-	-
	γ_1	0.5	1.4	17.0	19.3	98.3	0.3	14.3
	γ_2	0.8	0.7	28.1	29.2	95.6	0.1	27.8
400	β_1	1.0	-1.2	13.5	13.0	93.4	-	-
	β_2	-0.5	-0.0	20.9	21.4	95.1	-	-
	γ_1	0.5	0.5	12.3	13.8	96.7	-0.3	10.4
	γ_2	0.8	0.3	19.5	20.7	95.8	-0.4	19.2

native estimators decrease from 90% to 40% in estimating γ_1 when the sample size increases from 100 to 400. These ratios are close to 1 in estimating γ_2 but also decrease significantly when the sample size increases.

We repeat the same simulation study using the same setting except that ϵ is generated from the logistic distribution, that is, the terminal event follows the proportional odds model. The results and conclusions are similar (results not shown).

6. Real Example

We apply our method to analyze the data from a subgroup in the AIDS Links to Intravenous Experiences (ALIVE) cohort study (Vlahov et al., 1991). In this study, a group of intravenous drug users with HIV infections were followed between August 1, 1993 and December 31, 1997, where the collected data included their in-patient admissions and other variables. The terminal

Table 2. *Simulation Results from 1000 Repetitions with Informative Terminal Events*

			Our approach				Naive	
n	Par.	True	Bias ($\times 10^{-2}$)	SE ($\times 10^{-2}$)	ESE ($\times 10^{-2}$)	CP ($\times 10^{-2}$)	Bias ($\times 10^{-2}$)	SE ($\times 10^{-2}$)
100	β_1	1.0	2.6	26.2	26.3	94.6	-	-
	β_2	-0.5	-6.8	45.4	43.6	94.8	-	-
	γ_1	0.5	13.3	47.1	49.5	96.5	42.3	37.7
	γ_2	0.8	1.5	80.4	77.1	95.5	-23.8	73.2
200	β_1	1.0	0.1	18.2	18.4	94.7	-	-
	β_2	-0.5	-0.8	31.9	30.4	94.0	-	-
	γ_1	0.5	7.8	32.5	35.5	96.5	43.6	26.0
	γ_2	0.8	0.2	54.2	54.2	95.2	-21.6	49.0
400	β_1	1.0	-1.2	13.5	13.0	93.4	-	-
	β_2	-0.5	-0.0	20.9	21.4	95.1	-	-
	γ_1	0.5	3.1	23.5	25.3	96.4	42.2	19.1
	γ_2	0.8	0.4	37.9	38.6	93.9	-21.3	33.8

event was death. For illustration, we only consider the female patients of 471 subjects. On average, each patient had 1.3 hospital admissions and there were 83 deaths. The interest focuses on the effects of the baseline HIV status (positive vs negative) and age on both recurrent hospital admissions and death.

First, to determine the survival model for the death, we consider the class of logarithmic transformations $r^{-1}\log(1+rx)$ for $G(x)$ by varying r from 0 to 1. The AIC criterion chooses the best transformation to be the proportional odds model ($r=1$). We then proceed to fit the additive rate model for the recurrent hospital admissions using our approach. The result is given in the first half of Table 3, which shows that the HIV positive patients tended to die earlier and experience more hospital admission, as compared to the HIV negative patients; the patient's age was significantly associated with the death but not the hospital admission.

To assess the goodness of fit using our model, we examine the following total summation of the residuals for each subject

$$\int_0^{Y_i} \left\{ dN_i(t) - d\widehat{H}(t, \log \widehat{\Lambda}(t) - X_i^T \widehat{\beta}) - X_i^T \widehat{\gamma} dt \right\},$$

Table 3. *Analysis of HIV Data*

	Death Model				Recurrent Event Model			
Covariates	Est	SE	Z-stat	p-value	Est	SE	Z-stat	p-value
				Data contain all 471 subjects				
HIV+ vs HIV-	1.570	0.278	5.641	< 0.001	0.135	0.057	2.359	0.018
Age	0.057	0.018	3.179	0.001	0.004	0.003	1.431	0.152
				Data exclude 11 extreme subjects				
HIV+ vs HIV-	1.651	0.356	4.640	< 0.001	0.105	0.044	2.408	0.016
Age	0.056	0.021	2.718	0.007	0.006	0.003	2.178	0.029

11 subjects are those who had at least 9 admissions.

equivalently,

$$\int_0^{Y_i} \left\{ dN_i(t) - d\bar{N}_i(t) - (X_i - \bar{X}_i(t))^T \widehat{\gamma} dt \right\}.$$

As shown in Section 2, when our model is correct, the above statistics should have an approximate mean zero and be independent of X_i. Therefore, a graphical way to assess the model fit is to plot the above residual quantity against covariate X_i. We plot in Figure 1 the summed residuals for each subject versus the patient's age within the HIV positive and negative groups respectively. Overall, we find that the residuals fluctuate around zero and appear to be random. The residuals for the subjects in HIV+ group appear to be slightly more spread-out than the ones for the subjects in HIV- group. In addition, we notice that there are 11 subjects who have residuals larger than 5. Interestingly, these subjects are all extreme cases who experienced at least 9 admissions; thus, their observations can be very influential in the model fitting. For instance, after removing these subjects, the average number of the admission reduces to 1.11; moreover, the result from the model fit, as given in the second half of Table 3, shows that the age's effect becomes much more significant for the recurrent event model.

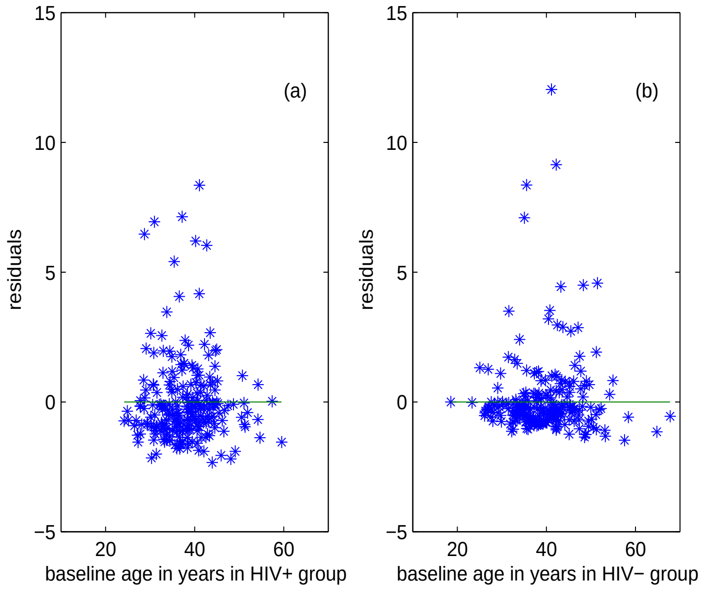

Fig. 1. The plot of the residuals vs the baseline ages. Plot (a) plots the residuals for the subjects in HIV+ group (b) plots the residuals for the subjects in HIV- group.

7. Discussion

In this paper, the general transformation models were used to model the terminal event given covariates. However, such models are not essential in our approach. Other models such as the accelerated failure time model or the additive hazards model can also be used. The choice of the model for the terminal event depends on data fitting.

In obtaining the estimating equation for γ, we constructed the risk set at time t based on the ranks of both the terminal event residual ϵ and $X^T\beta$ and gave each subject in the risk set equal weights. One possibility is to assign different weights based on each subject's covariate information. It is unclear what weight functions can lead to a more efficient estimator for γ. Another possibility to construct the risk set is to adapt the artificial censoring idea which was used in Lin, Robins and Wei (1996) and Ghosh and Lin (2003) under different contexts and models. This idea will further trim the risk set we constructed here. It remains unknown how much efficiency gain/loss the artificial censoring will have. A better alternative approach is to combine the estimators from our method and the artificial censoring approach in an optimal way, which will guarantee the efficiency improvement. We will explore this approach in the future.

Although we focused on the additive rate model for the recurrent event, our inference method also applies to the proportional rate model, where the rate function is given as

$$E[dN(t)|T > t, \nu, X] = I(T > t)e^{X^T\gamma}dR(t, \nu).$$

The same estimating equation can be constructed as in Section 2. However, the interpretation of the coefficient γ is different between the additive rate model and the proportional rate model. Finally, we can model the mean function of the recurrent event instead of the rate function by assuming

$$E[N(t)|X, T > t, \nu] = I(T > t)\left\{R(t, \nu) + X^T\gamma t\right\}.$$

Note that this model may only imply the rate model if X is time-independent.

Acknowledgement

This work was partially supported by the National Institutes of Health grant R01-HL57444.

Appendix

Proof of Theorem 2

To prove Theorem 2, we define $d\mathcal{R}(t) = dN(t) - X^T\gamma_0 dt$ and

$$d\overline{\mathcal{R}}(t, X; \beta, \Lambda) = \frac{\sum_{j=1}^n d\mathcal{R}_j(t) I(\Lambda(Y_j)e^{-X_j^T\beta} > \Lambda(t)e^{-X^T\beta}, X_j^T\beta \geq X^T\beta)}{\sum_{j=1}^n I(\Lambda(Y_j)e^{-X_j^T\beta} > \Lambda(t)e^{-X^T\beta}, X_j^T\beta \geq X^T\beta)}.$$

Moreover, based on 2.10.4 of van der Vaart and Wellner, the class

$$\{\Lambda(Y) : \Lambda \text{ is non-decreasing and right-contiuous and bounded by } c_0\}$$

is a VC-hull class; the same holds for the finite dimensional space $\left\{X^T\beta : \beta \in R^d\right\}$. Thus,

$$\left\{\Lambda(Y)e^{-X^T\beta} : \|\Lambda - \Lambda_0\| + |\beta - \beta_0| < \delta_0\right\}$$

is a universally Donsker class. Therefore, from the Glivenko-Cantelli theorem, it is clear that the asymptotic limit of $d\overline{\mathcal{R}}(t, X; \beta, \Lambda)$ is equal to

$$\frac{E\left[d\mathcal{R}_j(t)I(\Lambda(Y_j)e^{-X_j^T\beta} > \Lambda(t)e^{-X^T\beta}, X_j^T\beta \geq X^T\beta)\right]}{E\left[I(\Lambda(Y_j)e^{-X_j^T\beta} > \Lambda(t)e^{-X^T\beta}, X_j^T\beta \geq X^T\beta)\right]},$$

which is denoted as $d\mathcal{R}_0(t, X; \beta, \Lambda)$. Moreover, such convergence is uniformly in $t \in [0, \tau]$, X, and (β, Λ) is the neighborhood of (β_0, Λ_0). Similarly, we define the limit of $\overline{X}_i(t)$ as

$$E_0(X, t; \beta, \Lambda) = \frac{E\left[X_j I(\Lambda(Y_j)e^{-X_j^T\beta} > \Lambda(t)e^{-X^T\beta}, X_j^T\beta \geq X^T\beta)\right]}{E\left[I(\Lambda(Y_j)e^{-X_j^T\beta} > \Lambda(t)e^{-X^T\beta}, X_j^T\beta \geq X^T\beta)\right]}$$

evaluated at $X = X_i, \beta = \widehat{\beta}, \Lambda = \widehat{\Lambda}$.

From expression (4), we have

$$\widehat{\gamma} - \gamma_0 = \left[\sum_{i=1}^{n} \int \omega(t)I(Y_i > t)(X_i - \overline{X}_i(t))^{\otimes 2}dt\right]^{-1}$$

$$\times \left[\sum_{i=1}^{n} \int \omega(t)I(Y_i > t)(X_i - \overline{X}_i(t))d\left\{\mathcal{R}_i(t) - \overline{\mathcal{R}}_i(t; \widehat{\beta}, \widehat{\Lambda})\right\}\right].$$

Note that with probability one,

$$\frac{1}{n}\sum_{i=1}^{n} \int \omega(t)I(Y_i > t)(X_i - \overline{X}_i(t))^{\otimes 2}dt \to \Sigma_X \equiv E\left[\int \omega(t)I(Y > t)\left(X - E_0(X, t; \beta_0, \Lambda_0)\right)^{\otimes 2}\right].$$

Since $E_0(X, t; \beta_0, \Lambda_0)$ is a function of ϵ and X and ϵ are independent, from condition (C.2), the above limit must be positive definite. Thus, it holds

$$n^{1/2}(\widehat{\gamma} - \gamma_0) = n^{1/2}\left(\Sigma_X + o(1)\right)^{-1}\left[n^{-1}\sum_{i=1}^{n} \int \omega(t)I(Y_i > t)(X_i - \overline{X}_i(t))d\left\{\mathcal{R}_i(t) - \overline{\mathcal{R}}_i(t; \widehat{\beta}, \widehat{\Lambda})\right\}\right]$$

$$= n^{1/2}\left(\Sigma_X + o(1)\right)^{-1}\left[n^{-1}\sum_{i=1}^{n}\int \omega(t) I(Y_i > t)(X_i - E_0(X_i, t; \widehat{\beta}, \widehat{\Lambda})) d\left\{\mathcal{R}_i(t) - \overline{\mathcal{R}}_i(t; \widehat{\beta}, \widehat{\Lambda})\right\}\right]$$

$$-n^{1/2}\left(\Sigma_X + o(1)\right)^{-1}\left[n^{-1}\sum_{i=1}^{n}\int \omega(t) I(Y_i > t)(\overline{X}_i(t) - E_0(X_i, t; \widehat{\beta}, \widehat{\Lambda})) d\left\{\mathcal{R}_i(t) - \overline{\mathcal{R}}_i(t; \widehat{\beta}, \widehat{\Lambda})\right\}\right]. \quad (A.1)$$

On the other hand, we note

$$\overline{\mathcal{R}}_i(t; \widehat{\beta}, \widehat{\Lambda}) - \mathcal{R}_0(X_i, t; \beta_0, \Lambda_0)$$

$$= [\overline{\mathcal{R}}_i(t; \widehat{\beta}, \widehat{\Lambda}) - \mathcal{R}_0(X_i, t; \widehat{\beta}, \widehat{\Lambda})] + [\mathcal{R}_0(X_i, t; \widehat{\beta}, \widehat{\Lambda}) - \mathcal{R}_0(X_i, t; \beta_0, \Lambda_0)]. \quad (A.2)$$

The first term of (A.2) can be rewritten

$$\frac{(\mathcal{P}_n - \mathcal{P})\left[\mathcal{R}(t) I(\Lambda_0(Y)e^{-X^T\beta_0} > \Lambda_0(t)e^{-X_i^T\beta_0}, X^T\beta_0 \geq X_i^T\beta_0)\right]}{E\left[I(\Lambda_0(Y)e^{-X^T\beta_0} > \Lambda_0(t)e^{-X_i^T\beta_0}, X^T\beta_0 \geq X_i^T\beta_0)\right]}$$

$$-\frac{E\left[\mathcal{R}(t) I(\Lambda_0(Y)e^{-X^T\beta_0} > \Lambda_0(t)e^{-X_i^T\beta_0}, X^T\beta_0 \geq X_i^T\beta_0)\right]}{E\left[I(\Lambda_0(Y)e^{-X^T\beta_0} > \Lambda_0(t)e^{-X_i^T\beta_0}, X^T\beta_0 \geq X_i^T\beta_0)\right]^2}$$

$$\times(\mathcal{P}_n - \mathcal{P})\left[I(\Lambda_0(Y)e^{-X^T\beta_0} > \Lambda_0(t)e^{-X_i^T\beta_0}, X^T\beta_0 \geq X_i^T\beta_0)\right] + o_p(n^{-1/2}). \quad (A.3)$$

Using the mean-value theorem, the second term of (A.2) becomes

$$\nabla_\beta \frac{E\left[\mathcal{R}(t) I(\Lambda_0(Y)e^{-X^T\beta_0} > \Lambda_0(t)e^{-X_i^T\beta_0}, X^T\beta_0 \geq X_i^T\beta_0)\right]}{E\left[I(\Lambda_0(Y)e^{-X^T\beta_0} > \Lambda_0(t)e^{-X_i^T\beta_0}, X^T\beta_0 \geq X_i^T\beta_0)\right]}(\widehat{\beta} - \beta_0)$$

$$+\nabla_\Lambda \frac{E\left[\mathcal{R}(t) I(\Lambda_0(Y)e^{-X^T\beta_0} > \Lambda_0(t)e^{-X_i^T\beta_0}, X^T\beta_0 \geq X_i^T\beta_0)\right]}{E\left[I(\Lambda_0(Y)e^{-X^T\beta_0} > \Lambda_0(t)e^{-X_i^T\beta_0}, X^T\beta_0 \geq X_i^T\beta_0)\right]^2}[\widehat{\Lambda} - \Lambda_0] + o_p(n^{-1/2}), \quad (A.4)$$

where ∇_β denotes the derivative with respect to β and ∇_Λ denotes the Hadmard derivative with respect to Λ. Therefore,

$$\mathcal{R}_i(t) - \overline{\mathcal{R}}_i(t;\widehat{\beta},\widehat{\Lambda}) = \mathcal{R}_i(t) - \mathcal{R}_0(X_i,t;\beta_0,\Lambda_0)$$

$$-(\mathcal{P}_n - \mathcal{P})S_1(O;\beta_0,\Lambda_0,X_i,t) - \mathcal{I}(X_i,t)(\widehat{\beta}-\beta_0,\widehat{\Lambda}-\Lambda_0) + o_p(n^{-1/2}),$$

where O denotes the observed statistic, $S_1(O;\beta_0,\Lambda_0,X_i,t)$ is the influence function given in equation (A.3), and $\mathcal{I}$ is the linear operator as given in equation (A.4).

Consequently, since $\sup_{i,t}|\overline{X}_i(t) - E_0(X_i,t;\widehat{\beta},\widehat{\Lambda})| \to 0$, (A.1) gives

$$\begin{aligned}
& n^{1/2}(\widehat{\gamma}-\gamma_0) \\
&= n^{1/2}\left(\Sigma_X + o(1)\right)^{-1}\left[n^{-1}\sum_{i=1}^{n}\int \omega(t)I(Y_i>t)(X_i - E_0(X_i,t;\widehat{\beta},\widehat{\Lambda}))d\left\{R_i(t) - \mathcal{R}_0(X_i,t;\beta_0,\Lambda_0)\right.\right. \\
&\qquad \left.\left. -(\mathcal{P}_n-\mathcal{P})S_1(O;\beta_0,\Lambda_0,X_i,t) - \mathcal{I}(X_i)(\widehat{\beta}-\beta_0,\widehat{\Lambda}-\Lambda_0)\right\}\right] + o_p(1) \\
&= n^{1/2}\Sigma_X^{-1}(\mathcal{P}_n-\mathcal{P})\left[\int \omega(t)I(Y>t)(X - E_0(X,t;\beta_0,\Lambda_0))d(R(t) - \mathcal{R}_0(X,t;\beta_0,\Lambda_0))\right] \\
& -n^{1/2}\Sigma_X^{-1}(\mathcal{P}_n-\mathcal{P})\widetilde{E}\left[\int \omega(t)I(\widetilde{Y}>t)(\widetilde{X} - E_0(\widetilde{X},t;\beta_0,\Lambda_0))dS_1(O;\beta_0,\Lambda_0,\widetilde{X},t)\right] \\
& -n^{1/2}\Sigma_X^{-1}(\mathcal{P}_n-\mathcal{P})\widetilde{E}\left[\int \omega(t)I(\widetilde{Y}>t)(\widetilde{X} - E_0(\widetilde{X},t;\beta_0,\Lambda_0))d\mathcal{I}(\widetilde{X},t)[S_\beta,S_\Lambda]\right] \\
& +o_p(1).
\end{aligned}$$

Here, $\widetilde{E}$ is the expectation with respect to $(\widetilde{Y},\widetilde{X})$.

The asymptotic distribution for $n^{1/2}(\widehat{\beta}-\beta_0,\widehat{\Lambda}-\Lambda_0,\widehat{\gamma}-\gamma_0)$ thus follows from the above expansion and the expansions in (5).

Proof of Theorem 3

.

We examine Ω_1, Ω_2 and $\Omega_3 - \widehat{\gamma}$ separately. Clearly, using the same notation as in the proof of Theorem 2,

$$\Omega_1 = \left[n^{-1} \sum_{i=1}^{n} \int I(Y_i > t)\omega(t)(X_i - \overline{X}_i(t))^{\otimes 2} dt \right]^{-1} \times$$

$$n^{-1} \left[\sum_{i=1}^{n} \mathcal{Z}_i \int I(Y_i \geq t)\omega(t)(X_i - \overline{X}_i(t)) \left\{ dR_i(t) - d\overline{R}_i(t; \widehat{\beta}, \widehat{\Lambda}) \right\} \right].$$

Since the first term converges to Σ_X almost surely and $\overline{R}_i(t; \widehat{\beta}, \widehat{\Lambda})$ converges to $\mathcal{R}_0(X_i, t; \beta_0, \Lambda_0)$ and belongs to some Donsker class, we use Theorem 3.6.13 in van der Vaart and Wellner (1996) and conclude that conditional on data,

$$\Omega_1 = \Sigma_X^{-1} n^{-1} \left[\sum_{i=1}^{n} \mathcal{Z}_i \int I(Y_i > t)\omega(t)(X_i - E_0(X_i, t; \beta_0, \Lambda_0)) \left\{ d\mathcal{R}_i(t) - d\mathcal{R}_0(X_i, t; \beta_0, \Lambda_0) \right\} \right]$$
$$+ o_p(n^{-1/2}).$$

Similarly, we have

$$\begin{aligned}
\Omega_2 = & -\frac{1}{n}\Sigma_X^{-1} \sum_{j=1}^{n} \mathcal{Z}_j \widetilde{E} \left[\int I(\widetilde{Y} > t)\omega(t)(\widetilde{X} - E_0(\widetilde{X}, t; \beta_0, \Lambda_0)) \right. \\
& \left. \times \frac{(dN_j(t) - X_j^T \gamma) I(\Lambda_0(Y_j) e^{-X_j^T \beta_0} > \Lambda_0(t) e^{-\widetilde{X}^T \beta_0}, X_j^T \beta_0 \geq \widetilde{X}^T \beta_0)}{E\left[I(\Lambda_0(Y) e^{-X^T \beta_0} > \Lambda_0(t) e^{-\widetilde{X}^T \beta_0}, X^T \beta_0 \geq \widetilde{X}^T \beta_0) \right]} \right] \\
& + \frac{1}{n}\Sigma_X^{-1} \sum_{j=1}^{n} \mathcal{Z}_j \widetilde{E} \left[\int \omega(t) I(\widetilde{Y} > t)(\widetilde{X} - E_0(\widetilde{X}, t; \beta_0, \Lambda_0)) \right. \\
& \times I(\Lambda_0(Y_j) e^{-X_j^T \beta_0} > \Lambda_0(t) e^{-\widetilde{X}^T \beta_0}, X_j^T \beta_0 \geq \widetilde{X}^T \beta_0) \\
& \left. \times \frac{E\left[(dN(t) - X^T \gamma_0 dt) I(\Lambda_0(Y) e^{-X^T \beta_0} > \Lambda_0(t) e^{-\widetilde{X}^T \beta_0}, X^T \beta_0 \geq \widetilde{X}^T \beta_0) \right]}{E\left[I(\Lambda_0(Y) e^{-X^T \beta_0} > \Lambda_0(t) e^{-\widetilde{X}^T \beta_0}, X^T \beta_0 \geq \widetilde{X}^T \beta_0) \right]^2} \right] \\
& + o_p(n^{-1/2}) \\
= & \frac{1}{n}\Sigma_X^{-1} \sum_{i=1}^{n} \mathcal{Z}_i \tilde{E} \left[\int \omega(t) I(\widetilde{Y} > t)(\widetilde{X} - E_0(\widetilde{X}, t; \beta_0, \lambda_0)) dS_1(O_i; \beta_0, \Lambda_0, \widetilde{X}, t) \right] + o_p(n^{-1/2}).
\end{aligned}$$

Finally,

$$\Omega_3 - \widehat{\gamma} = \left[\sum_{i=1}^{n} \int I(Y_i > t)\omega(t)(X_i - \overline{X}_i(t))^{\otimes 2} dt\right]^{-1} \times$$

$$\left[\sum_{i=1}^{n} \mathcal{Z}_i \int I(Y_i \geq t)\omega(t)(X_i - \overline{X}_i(t)) \left\{ d(\overline{R}_i(t; \widetilde{\beta}, \widetilde{\Lambda}) - \overline{R}_i(t; \widehat{\beta}, \widehat{\Lambda})) \right\}\right].$$

On the other hand,

$$\overline{\mathcal{R}}_i(t; \widetilde{\beta}, \widetilde{\Lambda}) - \overline{\mathcal{R}}_i(X, t; \widehat{\beta}, \widehat{\Lambda}) = \overline{\mathcal{R}}_i(t; \widetilde{\beta}, \widetilde{\Lambda}) - \mathcal{R}_0(X_i, t; \widetilde{\beta}, \widetilde{\Lambda}) - \left\{ \overline{\mathcal{R}}_i(t; \widehat{\beta}, \widehat{\Lambda}) - \mathcal{R}_0(X_i, t; \widehat{\beta}, \widehat{\Lambda}) \right\}$$

$$+ \{\mathcal{R}_0(X_i, t; \widetilde{\beta}, \widetilde{\Lambda}) - \mathcal{R}_0(X_i, t; \widehat{\beta}, \widehat{\Lambda})\}. \qquad (A.5)$$

Note that

$$\overline{\mathcal{R}}_i(t; \widetilde{\beta}, \widetilde{\Lambda}) - \mathcal{R}_0(X_i, t; \widetilde{\beta}, \widetilde{\Lambda}) - \left\{ \overline{\mathcal{R}}_i(t; \widehat{\beta}, \widehat{\Lambda}) - \mathcal{R}_0(X_i, t; \widehat{\beta}, \widehat{\Lambda}) \right\}$$

$$= (\mathcal{P}_n - \mathcal{P}) \left[S_1(O; \widetilde{\beta}, \widetilde{\Lambda}, X_i, t) - S_1(O; \widehat{\beta}, \widehat{\Lambda}, X_i, t) \right] = o_p(n^{-1/2})$$

and that the last term in (A.5), by the Taylor expansion, is equal to

$$\frac{1}{n} \sum_{i=1}^{n} \mathcal{I}(X_i, t)[\widetilde{\beta} - \widehat{\beta}, \widetilde{\Lambda} - \widehat{\Lambda}] + o_p(n^{-1/2})$$

$$= \frac{1}{n} \sum_{i=1}^{n} \mathcal{Z}_i \mathcal{I}(X_i, t)[S_\beta, S_\Lambda] + o_p(n^{-1/2}).$$

Hence, from the influence function for $\widehat{\gamma}$ as derived in proving Theorem 2, we obtain

$$\Omega_1 + \Omega_2 + (\Omega_3 - \widehat{\gamma}) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{Z}_i S_\gamma(N_i, Y_i, \Delta_i, X_i; \beta_0, \Lambda_0) + o_p(n^{-1/2}).$$

Theorem 3 thus holds from Theorem 3.6.13 in van der Vaart and Wellner (1996).

REFERENCES

Abrams, D. I., Goldman, A. I., Launer, C., Korvick, J. A., Neaton, J. D., Crane, L. R., Grodesky, M., Wakefield, S., Muth, K., Kornegay, S., Cohn, D. L., Harris, A., Luskin-Hwark, R., Markowitz, N., Sampson, J. H., Thompson, M., Deyton, L. (1994). A comparative trial of Didanosine or Zalcitabine after treatment with Zidovudine in patients with human immunodeficiency virus infection. *New England Journal of Medicine* 330, 657-662.

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

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

Ghosh, D. and Lin, D. Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* 56, 554-562.

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

Ghosh, D. and Lin, D. Y. (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* 59, 877-885.

Huang, C. and Wang, M. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* 99, 1153-1165.

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

Lin, D. Y., Robins, J. M. and Wei, L. J. (1996). Comparing two failure time distributions in the presence of dependent censoring. *Biometrika* 83, 381-393.

Lin, D.Y., Wei, L.J., Yang, I. and Ying, Z. (2000). Semiparametric regression fort he mean and rate functions of recurrent events. *Journal of the Royal Statistical Society B* 62, 711-730.

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

Liu, L., Wolfe, R. A., and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* 60, 747-756.

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

Miloslavsky, M., Keles, M., van der Laan, M. J. and Butler, S. (2004). Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *Journal of the Royal Statistical Society*, Serie B 66, 239-257.

Neaton, J. D., Wentworth, D. N., Rhame, F., Hogan, C., Abrams, D. I., and Deyton, L. (1994). Consideration in choice of a clinical endpoint for AIDS clinical trials. *Statistics in Medicine* 13, 2107-2125.

Pepe, M. and Cai, J.(1993). Some graphical displays and marginal regression analysis for recurrent failure times and time dependent covariates, *Journal of the American Statistical Association* 88, 811-820.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* 68, 373-379.

Schaubel, D.E., Zeng, D. and Cai, J. (2006). A semiparametric additive rates model for recurrent event data. *Lifetime Data Analysis* 12, 386-406.

Vlahov, D., Anthony, J. C., Muňoz, A., Margolick, J., Nelson, K.E., Celentano, D. D., Solomon, L., and Polk, B. F. (1991). The ALIVE study: A longitudinal study of HIV-1 infection in intravenous drug users: Description of methods. *Journal of Drug Issues* 21, 758-776.

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

Vlahov, D., Anthony, J. C., Muňoz, A., Margolick, J., Nelson, K.E., Celentano, D. D., Solomon, L., and Polk, B. F. (1991). The ALIVE study: A longitudinal study of HIV-1 infection in intravenous drug users: Description of methods. *Journal of Drug Issues* 21, 758-776.

Wang, M., Qin, J., and Chiang, C. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association* 96, 1057-1065.

Zeng, D. and Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* 93, 627-640.

Zeng, D. and Lin, D. Y. (2009). Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics*, in press.