Semiparametric Regression Analysis of Longitudinal Data With Informative Observation Times

Jianguo SUN, Do-Hwan PARK, Liuquan SUN, and Xingqiu ZHAO

Statistical analysis of longitudinal data has been discussed by many authors, and a number of methods have been proposed. Most of the research have focused on situations where observation times are independent of or carry no information about the response variable and therefore rely on conditional inference procedures given the observation times. This article considers a different situation, where the independence assumption may not hold; that is, the observation times may carry information about the response variable. For inference, estimating equation approaches are proposed, and both large-sample and final-sample properties of the proposed methods are established. The methodology is applied to a bladder cancer study that motivated this investigation.

KEY WORDS: Estimating equation; Informative observation times; Longitudinal data; Nonhomogeneous Poisson process.

1. INTRODUCTION

This article discusses regression analysis of longitudinal data, which commonly occur in many types of studies, including medical follow-up studies and observational investigations. In this analysis, one major difficulty is that observation times often differ from subject to subject. Most of the proposed methods focus on the situation where observation times are independent of response variables completely or given covariates (Diggle, Liang, and Zeger 1994; Laird and Ware 1982; Zhang 2002). Here we consider a different situation, where the independence assumption may not hold, in other words, the observation times may be dependent on the response variables, and the question is how to carry out the analysis. Estimating equation approaches are proposed for the analysis.

A common situation where informative observation times occur is that observation times are subject- or response variable–dependent. For example, they may be hospitalization times of subjects in the study (Wang, Qin, and Chiang 2001). Such an example is given by a set of longitudinal data arising from a bladder cancer follow-up study conducted by the Veterans Administration Cooperative Urological Research Group (Sun and Wei 2000; Zhang 2002). All patients had superficial bladder tumors when they entered the study, and these tumors were removed transurethrally. Many patients had multiple recurrences of tumor during the study, and these recurrent tumors were also removed at clinical visits. The observed data include the numbers of recurrent tumors between clinical visits, and one objective of the study was to compare tumor recurrence rates. One problem with the dataset is that some patients in the study had significantly more clinical visits than others (Sun and Wei 2000); this indicates that the number of clinical visits may contain some information about the tumor occurrence rate. Thus an important question is how to take into account or make use of this information for inference about the tumor recurrence rate. More details about the study are given later in the article.

A number of methods have been proposed for the analysis of longitudinal data if observation times are independent of the response variable completely or given covariates. In this case, two commonly used approaches are the estimating equation and random-effects model approaches (Diggle et al. 1994; Laird and Ware 1982; Zeger and Diggle 1994). Among others, Diggle et al. (1994) gave an excellent review of these methods in addition to other methods. More recently, Hoover, Rice, Wu, and Yang (1998), Lin and Ying (2001), and Welsh, Lin, and Carroll (2002) discussed general semiparametric analysis of longitudinal data.

One situation that is similar to that considered here, but different, is the analysis of longitudinal data in the presence of informative or nonignorable dropouts or informative censoring times (Little 1995; Roy and Lin 2002; Wang and Taylor 2001; Wu and Carroll 1988; Wulfsohn and Tsiatis 1997). In this case there exists an event time representing dropout or censoring time, such as death, that is related to the underlying longitudinal variable of interest and must be modeled together with the longitudinal variable to obtain valid inference. For the analysis, a common approach is to use either selection models (Diggle and Kenward 1994) or pattern-mixture models (Little 1995), which model the longitudinal variable or the event variable marginally, depending on the objective of the study, and then use the conditional model for the other variable given the marginally modeled variable. Another approach is to jointly model the two variables together (Wulfsohn and Tsiatis 1997). In both approaches, normality is often assumed for the longitudinal variable, and the likelihood function is commonly used for inference. Note that for the situation discussed later, instead of dealing with longitudinal and survival processes together, we have general longitudinal and counting processes, and there do not seem to exist established statistical methods for their joint analysis. For inference, we focus on semiparametric methods that do not require normality assumption.

In what follows, we begin in Section 2 by introducing notation and assumptions and describing models as well as some motivations for the presented models. For the analysis, we treat observation times as realizations of counting processes. For longitudinal processes of interest, we generalize general linear models used by Hoover et al. (1998) and Lin and Ying (2001), among others, to allow the dependence of response process on the counting process characterizing the observation times. In Section 3 we present inference procedures and establish the consistency and asymptotic normality of the proposed

Jianguo Sun is Associate Professor (E-mail: *sunj@missouri.edu*) and Do-Hwan Park is Doctoral Candidate, Department of Statistics, University of Missouri, Columbia, MO 65211. Liuquan Sun is Research Associate, Institute of Applied Mathematics, Academia Sinica, Beijing, Peoples Republic of China 100080. Xingqiu Zhao is Postdoctoral Fellow, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1. The research was supported in part by a National Institutes of Health grant (to J. Sun) and National Natural Science Foundation of China grant 10071092 (to L. Sun). The authors thank the editor, an associate editor, and a referee for their many helpful comments and suggestions.

Journal of the American Statistical Association
September 2005, Vol. 100, No. 471, Theory and Methods
DOI 10.1198/016214505000000060

estimators. We report simulation results on the proposed methods in Section 4 and they indicate that the methods work well for situations considered. In Section 5 we apply the proposed methodology to the bladder cancer study. We conclude with some discussion in Section 6.

2. NOTATION, MOTIVATIONS, AND STATISTICAL MODELS

Consider a longitudinal study that consists of a random sample of n subjects. For subject i, let $Y_i(t)$ denote the response variable and $\mathbf{X}_i(t)$ denote a p-dimensional vector of covariates that may depend on time t, $i = 1, \dots, n$. Suppose that $Y_i(t)$ is observed at time points $T_{i1} < T_{i2} < \cdots < T_{i,K_i}$, where K_i is the total number of observations on subject i. In the following, we regard these observation times arising from an underlying counting process $N_i^*(t)$ characterized by $N_i(t) = \sum_{j=1}^{K_i} I(T_{ij} \le t) = N_i^*(\min(t, C_i))$, where C_i is the follow-up or censoring time for subject i, $i = 1, \dots, n$. Then the process $Y_i(t)$ is observed only at the time points where $N_i(t)$ jumps. We assume that the covariate history $\{\mathbf{X}_i(t) : 0 \le t \le C_i\}$ is observed for each subject.

Before presenting the model for $Y_i(t)$ and $N_i^*(t)$, we note that for inference about $Y_i(t)$, if they are independent completely or conditional on covariates, then a marginal approach is usually used for inference (Lin and Ying 2001). For the situation where $Y_i(t)$ and $N_i^*(t)$ may depend on each other, one could have three choices: modeling them jointly, modeling $Y_i(t)$ marginally and then $N_i^*(t)$ conditional on $Y_i(t)$, or modeling $N_i^*(t)$ marginally and then $Y_i(t)$ conditional on $N_i^*(t)$. As for the case of longitudinal data with informative dropouts discussed earlier, to apply the first two approaches, one usually needs a normality assumption for inference, and it is difficult to establish asymptotic theory for the inference procedure. Also, it seems natural, as in the case where there exist both longitudinal and survival processes (Hogan and Laird 1997), that the second approach (corresponding to the selection model approach) should be used if the counting process is of primary interest, whereas the third approach (corresponding to the mixture model approach) should be chosen if the longitudinal process is of primary interest, as is the case here. If $Y_i(t)$ and $N_i^*(t)$ are independent, then the second and third approaches are equivalent.

Note that another basis for choosing a model is the purpose of the analysis. In addition to the evaluation of covariate effects on the longitudinal process, prediction and the independence test between the two processes are also often of interest. For these, the foregoing third, or conditional, approach is much more natural and convenient than the second, or marginal, approach. This is partly because it is hard or not straightforward to directly incorporate the observation process into marginal models about the longitudinal process, such as the model proposed by Lin and Ying (2001). Motivated by all of these reasons, we propose to use the conditional model approach, which allows direct testing of the independence and prediction in addition to easy derivation of the marginal covariate effects. In contrast, the marginal model approach only allows inference about the marginal covariate effects. The presented model is a generalization of the marginal model given by Lin and Ying (2001) and thus allows direct comparison of the results with and without taking into account the dependence.

Define $\mathcal{F}_{it} = \{N_i(s), 0 \le s < t\}$. For the analysis, we assume that $Y_i(t)$ follows the marginal model

$$E\{Y_i(t)|\mathbf{X}_i(t), \mathcal{F}_{it}\} = \mu_0(t) + \boldsymbol{\beta}'\mathbf{X}_i(t) + \boldsymbol{\alpha}'\mathbf{H}(\mathcal{F}_{it}) \qquad (1)$$

given $\mathbf{X}_i(t)$ and $\mathcal{F}_{it}$, where $\mu_0(t)$ is an unspecified smooth function of t, $\boldsymbol{\beta}$ is a vector of unknown regression parameters, $\boldsymbol{\alpha}$ is a q-dimensional vector of regression coefficients, and $\mathbf{H}(\cdot)$ is a vector of known functions of the counting process $N_i(t)$ up to time $t-$. For the observation process, we assume that $N_i(t)$ may depend on $\mathbf{X}_i(t)$ and let $N_i^*(t)$ is a nonhomogeneous Poisson process with

$$E\{dN_i^*(t)|\mathbf{X}_i(t)\} = e^{\boldsymbol{\gamma}'\mathbf{X}_i(t)}\, d\Lambda_0(t) \qquad (2)$$

for $i = 1, \dots, n$. In the foregoing, $\boldsymbol{\gamma}$ is a vector of unknown regression parameters and $\Lambda_0(t)$ is an arbitrary nondecreasing function representing the mean cumulative number of observations by time t. We give some comments about model (2) later.

As mentioned earlier, the main interest here is on the longitudinal process, rather than the observation process, which makes the foregoing conditional approach a natural choice. As suggested by the associate editor, one way to see the current situation is that one faces two confounding processes in terms of covariate effects, and it is necessary to adjust for the effect on the observation process to correctly evaluate the effect on the longitudinal process. By directly including the former effect, the foregoing conditional models allow one to actually correct for it. In contrast to the foregoing approach, a marginal approach would use, for example, model (1) without the third term on the right side and model $E\{dN_i^*(t)|\mathbf{X}_i(t), Y_i(t)\}$ instead of $E\{dN_i^*(t)|\mathbf{X}_i(t)\}$. A joint model approach would use some latent variables to connect $E\{Y_i(t)|\mathbf{X}_i(t)\}$ and $E\{dN_i^*(t)|\mathbf{X}_i(t)\}$. For both cases, it seems difficult to directly adjust for the covariate effect on the observation process and establish sound inference procedures.

The model (1) specifies that the process $Y_i(t)$ depends on the process $N_i(t)$ in a linear fashion through function H, which can be chosen according to situations. A natural and simple choice for H may be $H(\mathcal{F}_{it}) = N_i(t-)$, which means that $Y_i(t)$ and $\mathcal{F}_{it}$ are related through or all information about $Y_i(t)$ in $\mathcal{F}_{it}$ is given by the total number of observations. An alternative is that $Y_i(t)$ depends on $\mathcal{F}_{it}$ only through a recent number of observations, say, in u time units, and this corresponds to $H(\mathcal{F}_{it}) = N_i(t-) - N_i(t-u)$. One could define $\mathbf{H}$ as a vector given by the foregoing two choices if both the total and recent numbers of observations may contain information about $Y_i(t)$. If $\boldsymbol{\alpha} = \mathbf{0}$, then model (1) reduces to the model considered by Lin and Ying (2001) and Zeger and Diggle (1994). In the following, we assume that the censoring time C_i may depend on covariates $\mathbf{X}_i(t)$ in an arbitrary fashion, but is independent of $N_i^*(t)$ and $Y_i(t)$ given $\mathbf{X}_i(t)$.

3. INFERENCE PROCEDURES

In this section we present inference procedures for models (1) and (2) with the focus on estimation of regression parameters. To motivate the proposed estimators given herein, first consider situations in which $\boldsymbol{\gamma} = \mathbf{0}$; that is, observation times are independent of covariates. In this case, inference about model (1) can be made conditional on observation times, and a natural way for estimating $\boldsymbol\beta$ and $\boldsymbol\alpha$ is to use the least squares principle by minimizing

$$L(\boldsymbol\beta,\boldsymbol\alpha,\mu_0)=\sum_{i=1}^n\sum_{j=1}^{K_i}W(T_{ij})\{Y_i(T_{ij})-\mu_0(T_{ij})-\boldsymbol\beta'\mathbf{X}_i(T_{ij})-\boldsymbol\alpha'\mathbf{H}(\mathcal{F}_{i,T_{ij}})\}^2$$
$$=\sum_{i=1}^n\int_0^\infty W(t)\{Y_i(t)-\mu_0(t)-\boldsymbol\beta'\mathbf{X}_i(t)-\boldsymbol\alpha'\mathbf{H}(\mathcal{F}_{it})\}^2\,dN_i(t),$$

where $W(t)$ is a possibly data-dependent weight function.

For estimation of $\mu_0(t)$, note that

$$E\{Y_i(t)-\mu_0(t)-\boldsymbol\beta'\mathbf{X}_i(t)-\boldsymbol\alpha'\mathbf{H}(\mathcal{F}_{it})\}=0,\qquad i=1,\dots,n.$$

Thus a natural estimator of $\mu_0(t)$ is given by $\hat\mu_0(t)=\bar Y(t)-\boldsymbol\beta'\bar{\mathbf{X}}(t)-\boldsymbol\alpha'\bar{\mathbf{H}}(\mathcal{F}_t)$ given $\boldsymbol\beta$ and $\boldsymbol\alpha$, where

$$\bar{\mathbf{X}}(t)=\frac{\sum_{i=1}^n\xi_i(t)\mathbf{X}_i(t)}{\sum_{i=1}^n\xi_i(t)},\qquad \bar{\mathbf{H}}(\mathcal{F}_t)=\frac{\sum_{i=1}^n\xi_i(t)\mathbf{H}(\mathcal{F}_{it})}{\sum_{i=1}^n\xi_i(t)},$$

and

$$\bar Y(t)=\frac{\sum_{i=1}^n\xi_i(t)Y_i^*(t)}{\sum_{i=1}^n\xi_i(t)},$$

where $\xi_i(t)=I(C_i\ge t)$ and $Y_i^*(t)$ is the measurement of Y_i at the time point nearest to t.

By replacing $\mu_0(t)$ in $L(\boldsymbol\beta,\boldsymbol\alpha,\mu_0)$ with $\hat\mu_0(t)$, we have

$$L(\boldsymbol\beta,\boldsymbol\alpha,\hat\mu_0)=\sum_{i=1}^n\int_0^\infty W(t)\big[Y_i(t)-\bar Y(t)-\boldsymbol\beta'\{\mathbf{X}_i(t)-\bar{\mathbf{X}}(t)\}-\boldsymbol\alpha'\{\mathbf{H}(\mathcal{F}_{it})-\bar{\mathbf{H}}(\mathcal{F}_t)\}\big]^2\,dN_i(t).$$

The resulting estimating function for $\boldsymbol\beta$ and $\boldsymbol\alpha$ has the form

$$\mathbf{U}(\boldsymbol\beta,\boldsymbol\alpha)=\sum_{i=1}^n\int_0^\infty W(t)\begin{pmatrix}\mathbf{X}_i(t)-\bar{\mathbf{X}}(t)\\ \mathbf{H}(\mathcal{F}_{it})-\bar{\mathbf{H}}(\mathcal{F}_t)\end{pmatrix}\times\big[Y_i(t)-\bar Y(t)-\boldsymbol\beta'\{\mathbf{X}_i(t)-\bar{\mathbf{X}}(t)\}-\boldsymbol\alpha'\{\mathbf{H}(\mathcal{F}_{it})-\bar{\mathbf{H}}(\mathcal{F}_t)\}\big]\,dN_i(t)=\mathbf{0}.$$

Now we consider the general situation where observation times depend on covariate processes through model (2). For given $\boldsymbol\gamma$, define

$$\bar{\mathbf{X}}(t;\boldsymbol\gamma)=\frac{\sum_{i=1}^n\xi_i(t)\mathbf{X}_i(t)\exp\{\boldsymbol\gamma'\mathbf{X}_i(t)\}}{\sum_{i=1}^n\xi_i(t)\exp\{\boldsymbol\gamma'\mathbf{X}_i(t)\}},$$
$$\bar{\mathbf{H}}(\mathcal{F}_t;\boldsymbol\gamma)=\frac{\sum_{i=1}^n\xi_i(t)\mathbf{H}(\mathcal{F}_{it})\exp\{\boldsymbol\gamma'\mathbf{X}_i(t)\}}{\sum_{i=1}^n\xi_i(t)\exp\{\boldsymbol\gamma'\mathbf{X}_i(t)\}},$$

and

$$\bar Y(t;\boldsymbol\gamma)=\frac{\sum_{i=1}^n\xi_i(t)Y_i^*(t)\exp\{\boldsymbol\gamma'\mathbf{X}_i(t)\}}{\sum_{i=1}^n\xi_i(t)\exp\{\boldsymbol\gamma'\mathbf{X}_i(t)\}}.$$

Motivated by $\mathbf{U}(\boldsymbol\beta,\boldsymbol\alpha)$, we can estimate $\boldsymbol\beta$ and $\boldsymbol\alpha$ using the estimating equation $\mathbf{U}(\boldsymbol\beta,\boldsymbol\alpha;\boldsymbol\gamma)=\mathbf{0}$ given $\boldsymbol\gamma$, where

$$\mathbf{U}(\boldsymbol\beta,\boldsymbol\alpha;\boldsymbol\gamma)=\sum_{i=1}^n\int_0^\infty W(t)\begin{pmatrix}\mathbf{X}_i(t)-\bar{\mathbf{X}}(t;\boldsymbol\gamma)\\ \mathbf{H}(\mathcal{F}_{it})-\bar{\mathbf{H}}(\mathcal{F}_t;\boldsymbol\gamma)\end{pmatrix}\times\big[Y_i(t)-\bar Y(t;\boldsymbol\gamma)-\boldsymbol\beta'\{\mathbf{X}_i(t)-\bar{\mathbf{X}}(t;\boldsymbol\gamma)\}-\boldsymbol\alpha'\{\mathbf{H}(\mathcal{F}_{it})-\bar{\mathbf{H}}(\mathcal{F}_t;\boldsymbol\gamma)\}\big]\,dN_i(t).$$

The parameter $\boldsymbol\gamma$ can be consistently estimated by the solution to

$$\sum_{i=1}^n\int_0^\infty\{\mathbf{X}_i(t)-\bar{\mathbf{X}}(t;\boldsymbol\gamma)\}\,dN_i(t)=\mathbf{0}$$

(Andersen, Borgan, Gill, and Keiding 1993). Let $\hat{\boldsymbol\gamma}$ denote the solution to the foregoing equation. Given $\hat{\boldsymbol\gamma}$, the solution to $\mathbf{U}(\boldsymbol\beta,\boldsymbol\alpha;\hat{\boldsymbol\gamma})=\mathbf{0}$ has a closed form,

$$\begin{pmatrix}\hat{\boldsymbol\beta}\\ \hat{\boldsymbol\alpha}\end{pmatrix}=\left\{\sum_{i=1}^n\int_0^\infty W(t)\begin{pmatrix}\mathbf{X}_i(t)-\bar{\mathbf{X}}(t;\hat{\boldsymbol\gamma})\\ \mathbf{H}(\mathcal{F}_{it})-\bar{\mathbf{H}}(\mathcal{F}_t;\hat{\boldsymbol\gamma})\end{pmatrix}^{\otimes2}dN_i(t)\right\}^{-1}\times\sum_{i=1}^n\int_0^\infty W(t)\begin{pmatrix}\mathbf{X}_i(t)-\bar{\mathbf{X}}(t;\hat{\boldsymbol\gamma})\\ \mathbf{H}(\mathcal{F}_{it})-\bar{\mathbf{H}}(\mathcal{F}_t;\hat{\boldsymbol\gamma})\end{pmatrix}\times\{Y_i(t)-\bar Y(t;\hat{\boldsymbol\gamma})\}\,dN_i(t),$$

where $\mathbf{a}^{\otimes2}=\mathbf{a}\mathbf{a}'$. It is easy to show that the foregoing estimates are consistent.

To establish the asymptotic normality of $\hat{\boldsymbol\beta}$ and $\hat{\boldsymbol\alpha}$, define

$$\hat{\mathcal{A}}(t)=\sum_{i=1}^n\int_0^t\frac{[Y_i(s)-\hat{\boldsymbol\beta}'\mathbf{X}_i(s)-\hat{\boldsymbol\alpha}'\mathbf{H}(\mathcal{F}_{is})]\,dN_i(s)}{\sum_{j=1}^n\xi_j(s)\exp\{\hat{\boldsymbol\gamma}'\mathbf{X}_j(s)\}},$$
$$\hat\Lambda(t)=\sum_{i=1}^n\int_0^t\frac{dN_i(s)}{\sum_{j=1}^n\xi_j(s)\exp\{\hat{\boldsymbol\gamma}'\mathbf{X}_j(s)\}},$$
$$\hat M_i(t)=\int_0^t\big[\{Y_i(s)-\hat{\boldsymbol\beta}'\mathbf{X}_i(s)-\hat{\boldsymbol\alpha}'\mathbf{H}(\mathcal{F}_{is})\}\,dN_i(s)-\xi_i(s)e^{\hat{\boldsymbol\gamma}'\mathbf{X}_i(s)}\,d\hat{\mathcal{A}}(s)\big],$$
$$\hat{\mathcal{M}}_i(t)=N_i(t)-\int_0^t\xi_i(s)e^{\hat{\boldsymbol\gamma}'\mathbf{X}_i(s)}\,d\hat\Lambda(s),$$
$$\hat R_i(t)=\hat M_i(t)-\int_0^t[\bar Y(s;\hat{\boldsymbol\gamma})-\hat{\boldsymbol\beta}'\bar{\mathbf{X}}(s;\hat{\boldsymbol\gamma})-\hat{\boldsymbol\alpha}'\bar{\mathbf{H}}(\mathcal{F}_s;\hat{\boldsymbol\gamma})]\,d\hat{\mathcal{M}}_i(s),$$
$$\hat E_{\mathbf{X}}(t)=\frac{\sum_{i=1}^n\xi_i(t)\mathbf{X}_i(t)\mathbf{X}_i'(t)\exp\{\hat{\boldsymbol\gamma}'\mathbf{X}_i(t)\}}{\sum_{i=1}^n\xi_i(t)\exp\{\hat{\boldsymbol\gamma}'\mathbf{X}_i(t)\}},$$
$$\hat E_{\mathbf{H}}(t)=\frac{\sum_{i=1}^n\xi_i(t)\mathbf{H}(\mathcal{F}_{it})\mathbf{X}_i'(t)\exp\{\hat{\boldsymbol\gamma}'\mathbf{X}_i(t)\}}{\sum_{i=1}^n\xi_i(t)\exp\{\hat{\boldsymbol\gamma}'\mathbf{X}_i(t)\}},$$

and

$$\hat E_Y(t)=\frac{\sum_{i=1}^n\xi_i(t)Y_i(t)\mathbf{X}_i'(t)\exp\{\hat{\boldsymbol\gamma}'\mathbf{X}_i(t)\}}{\sum_{i=1}^n\xi_i(t)\exp\{\hat{\boldsymbol\gamma}'\mathbf{X}_i(t)\}}.$$

Then $\hat{\mathcal{A}}(t)$ and $\hat\Lambda(t)$ are consistent estimators of $\mathcal{A}_0(t)=\int_0^t\mu_0(s)\,d\Lambda_0(s)$ and $\Lambda_0(t)$. Let $\boldsymbol\beta_0$ and $\boldsymbol\alpha_0$ denote the true values of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. It can be shown that as $n \to \infty$, $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and $n^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$ have asymptotically a joint normal distribution with mean 0 and covariance matrix that can be consistently estimated by $\hat{\mathbf{D}}^{-1}\hat{\mathbf{V}}\hat{\mathbf{D}}^{-1}$, where

$$\hat{\mathbf{D}} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty} W(t)\begin{pmatrix}\mathbf{X}_i(t) - \bar{\mathbf{X}}(t;\hat{\boldsymbol{\gamma}})\\ \mathbf{H}(\mathcal{F}_{it}) - \bar{\mathbf{H}}(\mathcal{F}_t;\hat{\boldsymbol{\gamma}})\end{pmatrix}^{\otimes 2} dN_i(t)$$

and

$$\hat{\mathbf{V}} = \frac{1}{n}\sum_{i=1}^{n}\Bigg[\int_0^{\infty} W(t)\begin{pmatrix}\mathbf{X}_i(t) - \bar{\mathbf{X}}(t;\hat{\boldsymbol{\gamma}})\\ \mathbf{H}(\mathcal{F}_{it}) - \bar{\mathbf{H}}(\mathcal{F}_t;\hat{\boldsymbol{\gamma}})\end{pmatrix} d\hat{R}_i(t) - \hat{\mathbf{P}}\hat{\boldsymbol{\Omega}}^{-1}\int_0^{\infty}\{\mathbf{X}_i(t) - \bar{\mathbf{X}}(t;\hat{\boldsymbol{\gamma}})\}\,d\hat{\mathcal{M}}_i(t)\Bigg]^{\otimes 2},$$

where

$$\begin{aligned}\hat{\mathbf{P}} = {} & \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty} W(t)\begin{pmatrix}\hat{E}_{\mathbf{X}}(t) - \bar{\mathbf{X}}(t;\hat{\boldsymbol{\gamma}})\bar{\mathbf{X}}'(t;\hat{\boldsymbol{\gamma}})\\ \hat{E}_{\mathbf{H}}(t) - \bar{\mathbf{H}}(\mathcal{F}_t;\hat{\boldsymbol{\gamma}})\bar{\mathbf{X}}'(t;\hat{\boldsymbol{\gamma}})\end{pmatrix} \times \{Y_i(t) - \bar{Y}(t;\hat{\boldsymbol{\gamma}})\}\,dN_i(t)\\ & - \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty} W(t)\begin{pmatrix}\hat{E}_{\mathbf{X}}(t) - \bar{\mathbf{X}}(t;\hat{\boldsymbol{\gamma}})\bar{\mathbf{X}}'(t;\hat{\boldsymbol{\gamma}})\\ \hat{E}_{\mathbf{H}}(t) - \bar{\mathbf{H}}(\mathcal{F}_t;\hat{\boldsymbol{\gamma}})\bar{\mathbf{X}}'(t;\hat{\boldsymbol{\gamma}})\end{pmatrix} \times \hat{\boldsymbol{\alpha}}'\{\mathbf{H}(\mathcal{F}_{it}) - \bar{\mathbf{H}}(\mathcal{F}_t;\hat{\boldsymbol{\gamma}})\}\,dN_i(t)\\ & + \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty} W(t)\begin{pmatrix}\mathbf{0}\\ \mathbf{H}(\mathcal{F}_{it}) - \bar{\mathbf{H}}(\mathcal{F}_t;\hat{\boldsymbol{\gamma}})\end{pmatrix} \times \big[\hat{E}_Y(t) - \bar{Y}(t;\hat{\boldsymbol{\gamma}})\bar{\mathbf{X}}'(t;\hat{\boldsymbol{\gamma}}) - \hat{\boldsymbol{\beta}}'\{\hat{E}_{\mathbf{X}}(t) - \bar{\mathbf{X}}(t;\hat{\boldsymbol{\gamma}})\bar{\mathbf{X}}'(t;\hat{\boldsymbol{\gamma}})\}\big]\,dN_i(t)\\ & - \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty} W(t)\begin{pmatrix}\mathbf{0}\\ \mathbf{H}(\mathcal{F}_{it}) - \bar{\mathbf{H}}(\mathcal{F}_t;\hat{\boldsymbol{\gamma}})\end{pmatrix} \times \hat{\boldsymbol{\alpha}}'\{\hat{E}_{\mathbf{H}}(t) - \bar{\mathbf{H}}(\mathcal{F}_t;\hat{\boldsymbol{\gamma}})\bar{\mathbf{X}}'(t;\hat{\boldsymbol{\gamma}})\}\,dN_i(t)\end{aligned}$$

and

$$\hat{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{i=1}^{n}\{\hat{E}_{\mathbf{X}}(t) - \bar{\mathbf{X}}(t;\hat{\boldsymbol{\gamma}})\bar{\mathbf{X}}'(t;\hat{\boldsymbol{\gamma}})\}\,dN_i(t).$$

The proof is given in Appendix A.

Note that model (1) gives only the conditional mean of the longitudinal process $Y_i(t)$. Given $\hat{\mu}_0$, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\alpha}}$, and $\hat{\boldsymbol{\gamma}}$, one can easily estimate the marginal mean of $Y_i(t)$, $E\{Y_i(t)|\mathbf{X}_i(t)\}$, and test treatment effects using models (1) and (2). For example, assuming that $H(\mathcal{F}_{it}) = N_i(t-) - N_i(t-u)$ and that $\mathbf{X}_i(t)$ is time-independent, we then have

$$E\{Y_i(t)|\mathbf{X}_i\} = \mu_0(t) + \boldsymbol{\beta}'\mathbf{X}_i + \alpha e^{\boldsymbol{\gamma}'\mathbf{X}_i}\{\Lambda_0(t-) - \Lambda_0(t-u)\}. \tag{3}$$

Also note that in the foregoing, only the joint asymptotic distribution of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ is presented. Using the same approach, one can easily show that the joint distribution of $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\alpha}}$, and $\hat{\boldsymbol{\gamma}}$ can be asymptotically approximated by a normal distribution, which is given in Appendix B.

4. SIMULATION STUDIES

This section reports some results from simulation studies conducted to evaluate the performance of the methods proposed in the previous section. In these studies, following Lin and Ying (2001), we considered the situation where there exist two covariates and first generated X_{i1}'s and X_{i2}'s from Bernoulli distribution with success probability .5 and the standard normal distribution. Given the X_{i1}'s and X_{i2}'s, the observation times T_{ij}'s were generated from model (2) with $\Lambda_0(t) = \lambda_0 t$, and the censoring times C_i were assumed to follow the uniform distribution over interval $(\tau/2, \tau)$, where τ was selected to give the desired number of real observation times. Finally, observations of the response variable were generated from the random-effects model

$$Y_i(t) = \mu_0(t) + \beta_1 X_{i1} + \beta_2 X_{i2} + \alpha N_i(t-) + b_i + \epsilon_i(t),$$

$i = 1, \ldots, n$. In the foregoing, we took $\mu_0(t) = t^{1/2}$ and assumed that the b_i's and $\epsilon_i(t)$'s were independent normal variables with mean 0 and standard deviations 1 and 5. The results presented here are based on $\lambda_0 = 1$, 5,000 replications and sample size $n = 100$ or 300.

In the simulation studies, we considered both covariate-independent ($\gamma = 0$) and covariate-dependent ($\gamma \neq 0$) observation time situations. Table 1 presents the simulation results for estimation of β_1 and α given by the proposed method with the true values of α as -1, 0, or 1 and both β_1 and β_2 equal to 1 for covariate-independent observation times ($\gamma_1 = \gamma_2 = 0$). The results on β_2 are similar and omitted. Here we took the weight function $W(t) = 1$. The table includes the biases (Bias) given by the sample means of the point estimates $\hat{\beta}_1$ and $\hat{\alpha}$ minus the true values, the sampling means of the estimated standard errors of $\hat{\beta}_1$ and $\hat{\alpha}$ (SEE), the sampling standard errors of $\hat{\beta}_1$ and $\hat{\alpha}$ (SSE), and the 95% empirical coverage probabilities for β_1 and α (CP). For the results in the table, τ was set to be 6 or 15, which gave about 5 or 11 observations on average. It can be seen from the table that the estimates $\hat{\beta}_1$ and $\hat{\alpha}$ are basically unbiased and that the proposed variance estimation method seems to work well. Also, with increasing sample sizes and/or the numbers of observations, as expected, the variances decrease and the coverage probabilities are more accurate.

The results for covariate-dependent observation times are given in Table 2, where we took $\gamma_1 = -.25$ and $\gamma_2 = .5$. The other setups in Table 2 are the same as those in Table 1. It can be seen that Table 2 gives basically the same conclusions as Table 1. To assess the bias given using Lin and Ying (2001)'s model [the model (1) without the third term at the right], in the presence of the dependence, we also estimated regression parameters using the approach given by Lin and Ying (2001) and obtained Bias = .0081 for β_1 for the situation corresponding to $\alpha = 0$, $n = 100$, and $\tau = 6$ in Table 2. With $\alpha = 1$ and -1 and other parameters the same, we got Bias = $-.2797$ and .7004 again for β_1, indicating that ignoring the dependence could yield significant biases. We also conducted simulation studies with large sample sizes or different values of β and γ and obtained similar results, with biases and estimated variances getting smaller when the sample size is larger.

Table 1. Simulation Results With Covariate-Independent Observation Times

	n = 100				n = 300			
	τ = 6		τ = 15		τ = 6		τ = 15	
	β_1	α	β_1	α	β_1	α	β_1	α
$\alpha = 0$								
Bias	.02069	.00103	−.00193	−.00347	.00620	−.00011	−.00212	−.00044
SSE	.53892	.16869	.38995	.07764	.30280	.09953	.21867	.04525
SEE	.51798	.16591	.37311	.07635	.30106	.09621	.21683	.04440
CP	.93660	.94460	.93600	.94260	.94920	.93940	.94500	.94680
$\alpha = 1$								
Bias	−.00693	−.00303	.00798	−.00267	.00295	−.00233	−.00080	−.00053
SSE	.52985	.17478	.39138	.07933	.30365	.09817	.21986	.04509
SEE	.51878	.17574	.37332	.07642	.30126	.09637	.21709	.04437
CP	.94200	.93440	.93720	.93560	.94600	.94340	.94520	.94400
$\alpha = -1$								
Bias	.00763	.00674	−.00298	.00163	.00514	−.00030	.00016	.00230
SSE	.54070	.16924	.38768	.07759	.30965	.09666	.21982	.04532
SEE	.51846	.16948	.37366	.07825	.30172	.09846	.21738	.04558
CP	.93680	.94540	.93940	.94780	.94440	.95240	.94900	.95120

5. ANALYSIS OF THE BLADDER CANCER STUDY

Now we apply the proposed approach to the bladder cancer study discussed earlier. The dataset, given by Sun and Wei (2000), includes the clinical visit or observation times and the number of bladder tumors that occurred between clinical visits for 85 patients in the placebo group (47) and the thiotepa treatment group (38). The unit for observation times is months, with the largest observation time being 53 months. The dataset also gave two baseline covariates, the number of initial tumors before entering the study and the size of the largest initial tumor. Several authors have analyzed the dataset, but all of the analyses assumed that observation times were independent of tumor recurrence completely or given covariates (Sun and Wei 2000; Zhang 2002). The goal here is to test the dependence between tumor recurrence and clinical visits and to compare the two treatments in terms of tumor recurrence rates with adjustment for the possible informative clinical visit times.

To analyze the dataset, for subject i we define $Y_i(t)$ as the natural logarithm of the number of observed tumors at time t on the subject plus 1 to avoid 0. We also set X_{i1} as 0 if the patient is in the placebo group and as 1 if the patient is in the thiotepa group, and set X_{i2} to be the number of initial tumors, $i = 1, \dots, 85$. Note that here we do not consider the size of the largest initial tumor, because several analyses have suggested that it had no effect on either tumor recurrences or observation times. For the analysis, it seems natural to assume that $Y_i(t)$ depends on $\mathcal{F}_{it}$ through the number of observations during the 6-month period before t, because it is usually the most recent visits that may carry information about the response variable; that is, $H(\mathcal{F}_{it}) = N_i(t-) - N_i(t-6)$. Applying the method with $W(t) = 1$ yielded $\hat{\alpha} = -.0317$ with an estimated standard error of .0096, indicating that the tumor recurrence process and observation process are significantly negatively correlated. One explanation for this finding is that the more often the patient visited the clinic, had tumors removed and received treatment, the lower the tumor recurrence rate. In other words, this means that the recurrence of bladder tumors may depend on the number of existing tumors. Another reason for the negative correlation could be that more visits means less time for tumor growth.

With respect to the effects of treatment and the number of initial tumors, we obtained $\hat{\beta}_1 = -.1350$, $\hat{\beta}_2 = .0472$, $\hat{\gamma}_1 = .5023$,

Table 2. Simulation Results With Covariate-Dependent Observation Times

	n = 100				n = 300			
	τ = 6		τ = 15		τ = 6		τ = 15	
	β_1	α	β_1	α	β_1	α	β_1	α
$\alpha = 0$								
Bias	−.00745	−.00445	−.00287	−.00153	.00034	−.00155	−.00650	.00034
SSE	.55022	.13505	.40567	.05450	.31954	.07476	.22936	.03032
SEE	.53737	.12622	.39281	.05088	.31171	.07317	.22920	.02984
CP	.93920	.92140	.93700	.92500	.94300	.94000	.94740	.94200
$\alpha = 1$								
Bias	.01232	−.00878	−.00994	−.00224	−.00423	−.00769	.00181	−.00158
SSE	.55367	.13621	.40776	.05296	.31578	.07670	.22881	.02914
SEE	.53465	.12647	.39273	.04776	.31241	.07360	.22882	.02777
CP	.93260	.92420	.93920	.91360	.94660	.93560	.94240	.93140
$\alpha = -1$								
Bias	.01025	.00762	.00605	.00142	.00768	.00553	−.00098	.00108
SSE	.54546	.13468	.40147	.05330	.31542	.07570	.23169	.02853
SEE	.53702	.12835	.39255	.04868	.31259	.07436	.22957	.02822
CP	.94380	.92920	.94240	.92060	.94900	.94260	.94880	.94060

and $\hat{\gamma}_2 = -.0089$, with estimated standard errors of .0501, .0132, .1197, and .0334. Note that here $\hat{\beta}_1$ represents the direct treatment effect on tumor recurrence and $\hat{\gamma}_1$ denotes the indirect treatment effect on tumor recurrence through the observation process. The same is true for $\hat{\beta}_2$ and $\hat{\gamma}_2$ with respect to the effect of the number of initial tumors. To test the treatment effect on tumor recurrence, we considered the hypothesis $\beta_1 = \gamma_1 = 0$ based on model (3) and used the chi-squared statistic based on the joint normal distribution of $\hat{\beta}_1$ and $\hat{\gamma}_1$, which gave a p value $<.001$. We used the same method for testing the effect of the number of initial tumors on tumor recurrence and obtained a p value of .001. The results suggest that both thiotepa treatment and the number of initial tumors significantly affect the tumor recurrence rate. Also, the observation process seems to be related to thiotepa treatment, but not to the number of initial tumors.

For comparison, we further analyzed the data by assuming that there is no direct relationship between tumor recurrence and the observation process, except that both depend on covariates as in the simulation study; that is, we set $\alpha = 0$ in model (2). In this case we obtained $\hat{\beta}_1 = -.1946$ and $\hat{\beta}_2 = .0492$, with estimated standard errors of .0456 and .0131. It is interesting to note that ignoring informative observation times does significantly affect the treatment effect on the tumor recurrence rate, but not the effect of the number of initial tumors on tumor recurrence rate. This could be because the number of initial tumors had no effect on the observation process.

As suggested by a referee, corresponding to the aforementioned number of recent visits, we also considered the situation where $Y_i(t)$ may depend on $N_i(t)$ through the time since the previous visit. In this case we obtained $\hat{\beta}_1 = -.1746$, $\hat{\beta}_2 = .0493$, and $\hat{\alpha} = .0150$ with estimated standard errors of .0484, .0133, and .0078, and the same $\hat{\gamma}$. It can be seen that all results are similar to these given earlier except $\hat{\alpha}$, which again indicates that the tumor recurrence process and the observation process are related, but in a different way. The results here suggest that, as expected, the longer the time since the last visit, the more tumors would occur.

6. CONCLUDING REMARKS

This article has discussed the analysis of longitudinal data in the presence of informative observation times, for which there do not seem to exist methods to our knowledge. To estimate regression parameters, we have proposed an estimating equation that yields consistent and asymptotically normal estimators. The approach can be considered a generalization of the method given by Lin and Ying (2001) for longitudinal data with noninformative observation times. Note that in the foregoing we have assumed that the observation process is a nonhomogeneous Poisson process for simplicity of presentation, and it should be straightforward to generalize the proposed method to general point processes.

In the proposed method, we have adopted a conditional approach. One may ask why we did not use the marginal approach. As discussed earlier, there are several reasons for this; the key is that except for the longitudinal process of interest, we face an additional dependent or confounding process. As for the longitudinal analysis in the presence of informative dropout times, it is hard or impossible to use a simple marginal model for the analysis. Another reason is that in addition to allowing one to easily make inference about the marginal mean of the longitudinal process, the conditional approach presented here allows one to directly test the dependence or confounding effect between the two processes and to perform prediction based on one's recent history. Also, one can actually correct for the confounding effect using the conditional approach. In contrast, the marginal approach gives results only about the marginal mean. In other words, for the situation considered here, the conditional approach provides more information than the marginal approach. Chen and Cook (2003), among others, have proposed a similar approach for a similar problem, but in a different context.

As commented before, a situation that is similar to, but not the same as, that discussed here and has been considered throughout the literature is the analysis of longitudinal data with informative dropout times. Another similar situation was considered by Sun and Wei (2000), who discussed the analysis of longitudinal data where both observation process and dropout process may depend on covariates but are independent of the underlying longitudinal process given covariates. It would be useful and interesting to generalize these existing and proposed methods to situations where the longitudinal process may be directly related to both observation process and dropout process instead of through covariate process.

APPENDIX A: JOINT ASYMPTOTIC NORMALITY OF $\hat{\beta}$ AND $\hat{\alpha}$

In this appendix we use the same notation defined earlier and take all limits at $n \to \infty$. Assume that the $\mathbf{X}_i(t)$'s are external covariates and have bounded variations. Define

$$S^{(0)}(t;\gamma) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(t) e^{\gamma' \mathbf{X}_i(t)},$$

$$S_x^{(1)}(t;\gamma) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(t) e^{\gamma' \mathbf{X}_i(t)} \mathbf{X}_i(t),$$

$$S_y^{(1)}(t;\gamma) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(t) e^{\gamma' \mathbf{X}_i(t)} Y_i^*(t),$$

$$S_h^{(1)}(t;\gamma) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(t) e^{\gamma' \mathbf{X}_i(t)} \mathbf{H}(\mathcal{F}_{it}),$$

$$S_x^{(2)}(t;\gamma) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(t) e^{\gamma' \mathbf{X}_i(t)} \mathbf{X}_i^{\otimes 2}(t),$$

$$S_{yx}^{(2)}(t;\gamma) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(t) e^{\gamma' \mathbf{X}_i(t)} Y_i^*(t) \mathbf{X}_i'(t),$$

and

$$S_{hx}^{(2)}(t;\gamma) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(t) e^{\gamma' \mathbf{X}_i(t)} \mathbf{H}(\mathcal{F}_{it}) \mathbf{X}_i'(t).$$

Also define

$$dM_i(t) = \{Y_i(t) - \boldsymbol{\beta}_0' \mathbf{X}_i(t) - \boldsymbol{\alpha}_0' \mathbf{H}(\mathcal{F}_{it})\}\, dN_i(t) - \xi_i(t) e^{\boldsymbol{\gamma}_0' \mathbf{X}_i(t)}\, d\mathcal{A}_0(t)$$

and

$$d\mathcal{M}_i(t) = dN_i(t) - \xi_i(t) e^{\boldsymbol{\gamma}_0' \mathbf{X}_i(t)}\, d\Lambda_0(t),$$

$i = 1, \ldots, n$. It is easy to show that $M_i(t)$ and $\mathcal{M}_i(t)$ are mean-0 stochastic processes. Let $s^{(0)}$, $s_x^{(1)}$, $s_y^{(1)}(t)$, $s_h^{(1)}$, $s_x^{(2)}$, $s_{yx}^{(2)}$, and $s_{hx}^{(2)}$ denote the limits of $S^{(0)}(t;\boldsymbol\gamma_0)$, $S_x^{(1)}(t;\boldsymbol\gamma_0)$, $S_y^{(1)}(t;\boldsymbol\gamma_0)$, $S_h^{(1)}(t;\boldsymbol\gamma_0)$, $S_x^{(2)}(t;\boldsymbol\gamma_0)$, $S_{yx}^{(2)}(t;\boldsymbol\gamma_0)$, and $S_{hx}^{(2)}(t;\boldsymbol\gamma_0)$. Also, let $\bar{x}(t) = s_x^{(1)}(t)/s^{(0)}(t)$, $\bar{y}(t) = s_y^{(1)}(t)/s^{(0)}(t)$, and $\bar{h}(t) = s_h^{(1)}(t)/s^{(0)}(t)$.

To see the joint asymptotic distribution of $\hat{\boldsymbol\beta}$ and $\hat{\boldsymbol\alpha}$, note that the Taylor series expansion of $\mathbf{U}(\hat{\boldsymbol\beta},\hat{\boldsymbol\alpha};\hat{\boldsymbol\gamma})$ at $\boldsymbol\beta=\boldsymbol\beta_0$, $\boldsymbol\alpha=\boldsymbol\alpha_0$, and $\boldsymbol\gamma=\hat{\boldsymbol\gamma}$ yields

$$n^{1/2}\begin{pmatrix}\hat{\boldsymbol\beta}-\boldsymbol\beta_0\\ \hat{\boldsymbol\alpha}-\boldsymbol\alpha_0\end{pmatrix} = \mathbf{D}^{-1}\{n^{-1/2}\mathbf{U}(\boldsymbol\beta_0,\boldsymbol\alpha_0;\hat{\boldsymbol\gamma})\} + o_p(1)$$

asymptotically, where

$$\mathbf{D} = E\left\{\int_0^\infty w(t)\begin{pmatrix}\mathbf{X}_1(t)-\bar{\mathbf{x}}(t)\\ \mathbf{H}(\mathcal{F}_{1t})-\bar{\mathbf{h}}(t)\end{pmatrix}^{\otimes 2} dN_1(t)\right\}$$

is the limit of $n^{-1}\partial\mathbf{U}(\boldsymbol\beta,\boldsymbol\alpha;\hat{\boldsymbol\gamma})/\partial(\boldsymbol\beta,\boldsymbol\alpha)$ and can be consistently estimated by $\hat{\mathbf{D}}$ given in Section 3, where $w(t)$ is the limit of $W(t)$. Thus it is sufficient to show that $n^{-1/2}\mathbf{U}(\boldsymbol\beta_0,\boldsymbol\alpha_0;\hat{\boldsymbol\gamma})$ has an asymptotic normal distribution with mean 0 and covariance matrix that can be consistently estimated by $\hat{\mathbf{V}}$ given in Section 3.

For $n^{-1/2}\mathbf{U}(\boldsymbol\beta_0,\boldsymbol\alpha_0;\hat{\boldsymbol\gamma})$, again using the Taylor series expansion of $\mathbf{U}(\boldsymbol\beta_0,\boldsymbol\alpha_0;\hat{\boldsymbol\gamma})$ at $\boldsymbol\beta=\boldsymbol\beta_0$, $\boldsymbol\alpha=\boldsymbol\alpha_0$, and $\boldsymbol\gamma=\boldsymbol\gamma_0$, we have, asymptotically,

$$\begin{aligned} &n^{-1/2}\mathbf{U}(\boldsymbol\beta_0,\boldsymbol\alpha_0;\hat{\boldsymbol\gamma})\\ &\quad= n^{-1/2}\mathbf{U}(\boldsymbol\beta_0,\boldsymbol\alpha_0;\boldsymbol\gamma_0)\\ &\qquad - \mathbf{P}\boldsymbol\Omega^{-1}n^{-1/2}\sum_{i=1}^n\int_0^\infty\{\mathbf{X}_i(t)-\bar{\mathbf{x}}(t)\}\,d\mathcal{M}_i(t) + o_p(1), \qquad (A.1)\end{aligned}$$

based on the consistency of $\hat{\boldsymbol\gamma}$ and equation (A.5) of Lin, Wei, Yang, and Ying (2000), where

$$\boldsymbol\Omega = E\left[\int_0^\infty\{\mathbf{X}_1(t)-\bar{\mathbf{x}}(t)\}^{\otimes 2}\xi_1(t)e^{\boldsymbol\gamma_0'\mathbf{X}_1(t)}\,d\Lambda_0(t)\right]$$

and

$$\begin{aligned}\mathbf{P} = {} & E\left[\int_0^\infty w(t)\begin{pmatrix} s_x^{(2)}(t)/s^{(0)}(t) - s_x^{(1)}(t)s_x^{(1)}(t)'/\{s^{(0)}(t)\}^2\\ s_{hx}^{(2)}(t)/s^{(0)}(t) - s_h^{(1)}(t)s_x^{(1)}(t)'/\{s^{(0)}(t)\}^2\end{pmatrix}\right.\\ &\qquad\left.\times\{Y_1(t)-\bar{y}(t)\}\,dN_1(t)\vphantom{\int}\right]\\ &- E\left[\int_0^\infty w(t)\begin{pmatrix} s_x^{(2)}(t)/s^{(0)}(t) - s_x^{(1)}(t)s_x^{(1)}(t)'/\{s^{(0)}(t)\}^2\\ s_{hx}^{(2)}(t)/s^{(0)}(t) - s_h^{(1)}(t)s_x^{(1)}(t)'/\{s^{(0)}(t)\}^2\end{pmatrix}\right.\\ &\qquad\left.\times\boldsymbol\alpha'\{\mathbf{H}(\mathcal{F}_{1t})-\bar{\mathbf{h}}(t)\}\,dN_1(t)\vphantom{\int}\right]\\ &+ E\left[\int_0^\infty w(t)\begin{pmatrix}\mathbf{0}\\ \mathbf{H}(\mathcal{F}_{1t})-\bar{\mathbf{h}}(t)\end{pmatrix}\right.\\ &\qquad\times\{s_{yx}^{(2)}(t)/s^{(0)}(t) - \bar{\mathbf{h}}(t)\bar{\mathbf{x}}'(t)\\ &\qquad\qquad\left. - \boldsymbol\beta'(s_x^{(2)}(t)/s^{(0)}(t) - \bar{\mathbf{x}}(t)\bar{\mathbf{x}}'(t))\}\,dN_1(t)\vphantom{\int}\right]\\ &- E\left[\int_0^\infty w(t)\begin{pmatrix}\mathbf{0}\\ \mathbf{H}(\mathcal{F}_{1t})-\bar{\mathbf{h}}(t)\end{pmatrix}\right.\\ &\qquad\left.\times\boldsymbol\alpha'\{s_{hx}^{(2)}(t)/s^{(0)}(t) - \bar{\mathbf{h}}(t)\bar{\mathbf{x}}'(t)\}\,dN_1(t)\vphantom{\int}\right],\end{aligned}$$

to which $-n^{-1}\partial\mathbf{U}(\boldsymbol\beta_0,\boldsymbol\alpha_0;\boldsymbol\gamma)/\partial\boldsymbol\gamma$ at $\boldsymbol\gamma=\boldsymbol\gamma_0$ converges in probability. Note that $\mathbf{U}(\boldsymbol\beta_0,\boldsymbol\alpha_0;\boldsymbol\gamma_0)$ can be rewritten as

$$\begin{aligned}&\mathbf{U}(\boldsymbol\beta_0,\boldsymbol\alpha_0;\boldsymbol\gamma_0)\\ &\quad=\sum_{i=1}^n\int_0^\infty W(t)\begin{pmatrix}\mathbf{X}_i(t)-\bar{\mathbf{X}}(t;\boldsymbol\gamma_0)\\ \mathbf{H}(\mathcal{F}_{it})-\bar{\mathbf{H}}(\mathcal{F}_t;\boldsymbol\gamma_0)\end{pmatrix}\\ &\qquad\times\left[dM_i(t) - \{\bar{Y}(t;\boldsymbol\gamma_0) - \boldsymbol\beta_0'\bar{\mathbf{X}}(t;\boldsymbol\gamma_0) - \boldsymbol\alpha_0'\bar{\mathbf{H}}(\mathcal{F}_t;\boldsymbol\gamma_0)\}\,d\mathcal{M}_i(t)\right].\end{aligned}$$

Furthermore, following the arguments similar to those given in appendix 2 of Lin and Ying (2001), we have, asymptotically,

$$\begin{aligned}&n^{-1/2}\mathbf{U}(\boldsymbol\beta_0,\boldsymbol\alpha_0;\boldsymbol\gamma_0)\\ &\quad= n^{-1/2}\sum_{i=1}^n\int_0^\infty w(t)\begin{pmatrix}\mathbf{X}_i(t)-\bar{\mathbf{x}}(t)\\ \mathbf{H}(\mathcal{F}_{it})-\bar{\mathbf{h}}(t)\end{pmatrix}\\ &\qquad\times\left[dM_i(t) - \{\bar{y}(t) - \boldsymbol\beta_0'\bar{\mathbf{x}}(t) - \boldsymbol\alpha_0'\bar{\mathbf{h}}(t)\}\,d\mathcal{M}_i(t)\right] + o_p(1), \qquad (A.2)\end{aligned}$$

a sum of n independent mean-0 random vectors plus an asymptotically negligible term.

It then follows from the multivariate central limit theorem and (A.1) and (A.2) that $n^{-1/2}\mathbf{U}(\boldsymbol\beta_0,\boldsymbol\alpha_0;\hat{\boldsymbol\gamma})$ converges in distribution to a mean-0 normal random vector with covariance matrix

$$\begin{aligned}\mathbf{V} = E\Bigg[&\int_0^\infty w(t)\begin{pmatrix}\mathbf{X}_1(t)-\bar{\mathbf{x}}(t)\\ \mathbf{H}(\mathcal{F}_{1t})-\bar{\mathbf{h}}(t)\end{pmatrix}dR_1(t)\\ &- \mathbf{P}\boldsymbol\Omega^{-1}\int_0^\infty\{\mathbf{X}_1(t)-\bar{\mathbf{x}}(t)\}\,d\mathcal{M}_1(t)\Bigg]^{\otimes 2},\end{aligned}$$

where

$$dR_1(t) = dM_1(t) - \{\bar{y}(t) - \boldsymbol\beta_0'\bar{\mathbf{x}}(t) - \boldsymbol\alpha_0'\bar{\mathbf{h}}(t)\}\,d\mathcal{M}_1(t).$$

By using the method in appendix A.3 of Lin et al. (2000), it can be shown that $\hat{\mathbf{V}}$ is a consistent estimate of $\mathbf{V}$. This proves the joint asymptotic normality of $\hat{\boldsymbol\beta}$ and $\hat{\boldsymbol\alpha}$.

APPENDIX B: NORMAL APPROXIMATION TO THE DISTRIBUTION OF $\hat{\boldsymbol\beta}$, $\hat{\boldsymbol\alpha}$, AND $\hat{\boldsymbol\gamma}$

Using the same method as that used in Appendix A, we can show that the joint distribution of $n^{1/2}(\hat{\boldsymbol\beta}-\boldsymbol\beta_0)$, $n^{1/2}(\hat{\boldsymbol\alpha}-\boldsymbol\alpha_0)$, and $n^{1/2}(\hat{\boldsymbol\gamma}-\boldsymbol\gamma_0)$ can be asymptotically approximated by the normal distribution with mean 0 and covariance matrix $\hat{\mathbf{D}}^{-1}\hat{\boldsymbol\Sigma}\hat{\mathbf{D}}^{-1}$, where, using the same notation,

$$\hat{\boldsymbol\Sigma} = \begin{pmatrix}\hat{\boldsymbol\Sigma}_{11} & \hat{\boldsymbol\Sigma}_{12}\\ \hat{\boldsymbol\Sigma}_{12}' & \hat{\boldsymbol\Sigma}_{22}\end{pmatrix}$$

with

$$\hat{\boldsymbol\Sigma}_{11} = \hat{\mathbf{D}}^{-1}\hat{\mathbf{V}}\hat{\mathbf{D}}^{-1},$$

$$\hat{\boldsymbol\Sigma}_{22} = \hat{\boldsymbol\Omega}^{-1}n^{-1}\sum_{i=1}^n\left(\int_0^\infty\{\mathbf{X}_i(t)-\bar{\mathbf{x}}(t)\}\,d\hat{\mathcal{M}}_i(t)\right)^{\otimes 2}\hat{\boldsymbol\Omega}^{-1},$$

and

$$\begin{aligned}\hat{\boldsymbol\Sigma}_{12} = \hat{\mathbf{D}}^{-1}n^{-1}\sum_{i=1}^n\Bigg[\Bigg\{&\int_0^\infty W(t)\begin{pmatrix}\mathbf{X}_i(t)-\bar{\mathbf{X}}(t;\hat{\boldsymbol\gamma})\\ \mathbf{H}(\mathcal{F}_{it})-\bar{\mathbf{H}}(\mathcal{F}_t;\hat{\boldsymbol\gamma})\end{pmatrix}d\hat{R}_i(t)\\ &- \hat{\mathbf{P}}\hat{\boldsymbol\Omega}^{-1}\int_0^\infty\{\mathbf{X}_i(t)-\bar{\mathbf{X}}(t;\hat{\boldsymbol\gamma})\}\,d\hat{\mathcal{M}}_i(t)\Bigg\}\\ &\times\int_0^\infty\{\mathbf{X}_i(t)-\bar{\mathbf{X}}(t;\hat{\boldsymbol\gamma})\}'\,d\hat{\mathcal{M}}_i(t)\Bigg]\hat{\boldsymbol\Omega}^{-1}.\end{aligned}$$

[Received December 2003. Revised December 2004.]

REFERENCES

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

Chen, E. B., and Cook, R. J. (2003), "Regression Modeling With Recurrent Events and Time-Dependent Interval-Censored Marker Data," *Lifetime Data Analysis*, 9, 275–291.

Diggle, P. J., and Kenward, M. G. (1994), "Informative Dropout in Longitudinal Data Analysis" (with discussion), *Applied Statistics*, 43, 49–94.

Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1994), *The Analysis of Longitudinal Data*, Oxford, U.K.: Oxford University Press.

Hogan, J. W., and Laird, N. M. (1997), "Model-Based Approaches to Analyzing Incomplete Longitudinal and Failure Time Data," *Statistics in Medicine*, 16, 259–272.

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L. P. (1998), "Nonparametric Smoothing Estimates of Time-Varying Coefficient Models With Longitudinal Data," *Biometrika*, 85, 809–822.

Laird, N. M., and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963–974.

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

Lin, D. Y., and Ying, Z. (2001), "Semiparametric and Nonparametric Regression Analysis of Longitudinal Data," *Journal of the American Statistical Association*, 96, 103–126.

Little, R. J. A. (1995), "Modeling the Drop-Out Mechanism in Repeated-Measures Studies," *Journal of the American Statistical Association*, 90, 1112–1121.

Roy, J., and Lin, X. (2002), "Analysis of Multivariate Longitudinal Outcomes With Nonignorable Dropouts and Missing Covariates: Changes in Methadone Treatment Practices," *Journal of the American Statistical Association*, 97, 40–52.

Sun, J., and Wei, L. J. (2000), "Regression Analysis of Panel Count Data With Covariate-Dependent Observation and Censoring Times," *Journal of Royal Statistical Society*, Ser. B, 62, 293–302.

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

Wang, Y., and Taylor, M. G. (2001), "Jointly Modeling Longitudinal and Event Time Data With Application to Acquired Immunodeficiency Syndrome," *Journal of the American Statistical Association*, 96, 895–905.

Welsh, A. H., Lin, X., and Carroll, R. J. (2002), "Marginal Longitudinal Nonparametric Regression: Locality and Efficiency of Spline and Kernel Methods," *Journal of the American Statistical Association*, 97, 482–493.

Wu, M. C., and Carroll, R. J. (1988), "Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process," *Biometrics*, 44, 175–188.

Wulfsohn, M. S., and Tsiatis, A. A. (1997), "A Joint Model for Survival and Longitudinal Data Measured With Error," *Biometrics*, 53, 330–339.

Zeger, S. L., and Diggle, P. J. (1994), "Semiparametric Models for Longitudinal Data With Application to CD4 Cell Numbers in HIV Seroconverters," *Biometrics*, 50, 689–699.

Zhang, Y. (2002), "A Semiparametric Pseudolikelihood Estimation Method for Panel Count Data," *Biometrika*, 89, 39–48.