Biostatistics (2003), **4**, 3, *pp.* 385–398
Printed in Great Britain

Semiparametric regression analysis of longitudinal data with informative drop-outs

D. Y. LIN*
Department of Biostatistics, University of North Carolina, CB#7420 McGavran-Greenberg, Chapel Hill, NC 27599-7420, USA
lin@bios.unc.edu

ZHILIANG YING
Department of Statistics, 618 Mathematics, Columbia University, New York, NY 10027, USA
zying@stat.columbia.edu

SUMMARY

Informative drop-out arises in longitudinal studies when the subject's follow-up time depends on the unobserved values of the response variable. We specify a semiparametric linear regression model for the repeatedly measured response variable and an accelerated failure time model for the time to informative drop-out. The error terms from the two models are assumed to have a common, but completely arbitrary joint distribution. Using a rank-based estimator for the accelerated failure time model and an artificial censoring device, we construct an asymptotically unbiased estimating function for the linear regression model. The resultant estimator is shown to be consistent and asymptotically normal. A resampling scheme is developed to estimate the limiting covariance matrix. Extensive simulation studies demonstrate that the proposed methods are suitable for practical use. Illustrations with data taken from two AIDS clinical trials are provided.

Keywords: Artificial censoring; Counting process; Dependent censoring; Linear regression; Missing data; Repeated measures.

1. INTRODUCTION

In longitudinal studies, the response is measured repeatedly through time. Although subjects are normally scheduled to be evaluated at a common set of time points, the actual measurement times may deviate considerably from the schedule and vary from subject to subject. Furthermore, subjects may be withdrawn from the study prematurely so that their subsequent measurements are missing. Informative drop-out or dependent censoring arises if the subject's follow-up time depends on the unobserved values of the response.

As a motivating example, we consider the AIDS Clinical Trials Group (ACTG) Protocol 128 Study, which was a randomized clinical trial comparing high-dose (180 mg m^{-2} every six hours) and low-dose (90 mg m^{-2}) regimens of zidovudine (AZT) in the treatment of children with HIV infection (Brady *et al.*, 1996). There were 216 and 208 children in the high-dose and low-dose groups, respectively. The objective of this study was to determine whether the low-dose regimen would be as effective as the

*To whom correspondence should be addressed

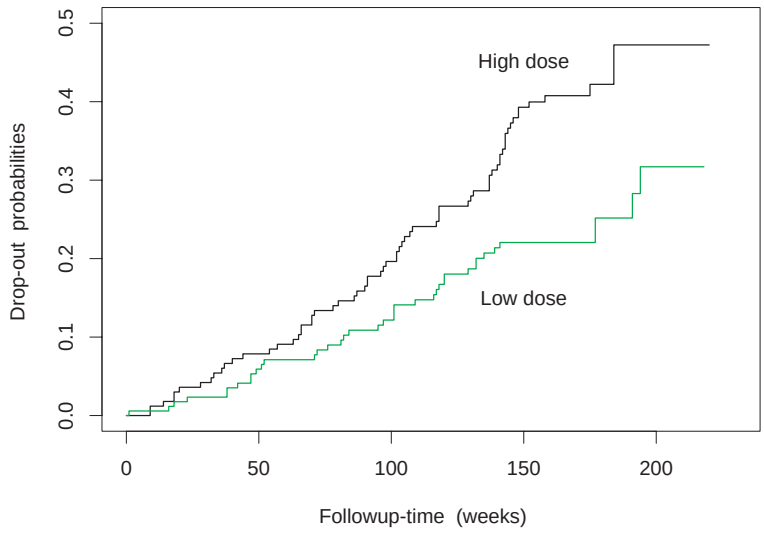

Fig. 1. Kaplan–Meier estimates for the cumulative probabilities of informative drop-out in the ACTG 128 Study.

standard high-dose regimen. One primary efficacy measure was the age-adjusted IQ score, which is a key neuropsychological marker for the HIV-disease status. The patients were scheduled to have their IQ scores measured every six months over a period of four years. The actual measurement times varied substantially from patient to patient. More important, a number of patients were lost to follow-up for reasons that might be related to their observed or unobserved IQ scores. These informative drop-outs include 33 disease progressions, 12 deaths, 14 drug toxicities and 43 voluntary withdrawals. As shown in Figure 1, the patients in the high-dose group tend to drop out earlier than those of the low-dose group; the p-value of the log-rank test is 0.0014.

The issue of informative drop-out/dependent censoring has received tremendous attention since the late 1980s. Most of the existing methods are based on the full parametric specification of the joint distribution of the response and drop-out mechanism. Hogan and Laird (1997) provided an excellent review of this literature, which includes the work of Wu and Bailey (1988), Wu and Carroll (1988), Mori *et al.* (1992), Schluchter (1992), Diggle and Kenward (1994) and Little (1995). As pointed out by Laird (1994), Rubin (1994) and Little (1995), the results under this approach rely heavily on the correct specification of the full parametric model, about which little is often known and which is not testable.

Robins *et al.* (1995) introduced a class of inverse probability of censoring weighted estimators which involves modelling the drop-out probability as a function of observed explanatory variables. This approach was recently extended by Rotnitzky *et al.* (1998) and Scharfstein *et al.* (1999) to adjust for the dependence of the drop-out probability on the unobserved response through sensitivity analysis. These methods as well as some of those based on the full parametric joint model require that the potential measurement times, i.e. the measurement times in the absence of drop-outs, consist of a fixed number of scheduled time points for all subjects; the methods of Rotnitzky *et al.* (1998) and Scharfstein *et al.* (1999) pertain only to a single fixed time point.

Recently, Yao *et al.* (1998) and Glidden and Wei (1999) proposed two-sample rank-based estimators by appealing to the technique of artificial censoring (Robins and Rotnitzky, 1992; Lin *et al.*, 1996). A major attraction of this approach is that the dependence between the drop-out time and response is not parametrized. Glidden and Wei (1999) required that the potential measurement times be fixed whereas

Yao *et al.* (1998) allowed the potential measurement times to be random. Because it involves pairwise comparisons of all observations, the method of Yao *et al.* (1998) is computationally demanding. The methods of Glidden and Wei (1999) and Yao *et al.* (1998) do not allow the potential measurement times to differ systematically between the two groups, and are not readily generalizable beyond the two-sample setting.

This paper deals with the problem of informative drop-outs in general settings. The methods proposed, like those of Yao *et al.* (1998) and Glidden and Wei (1999), do not parameterize the dependence of the drop-out time on the unobserved response and involve artificial censoring. As in Yao *et al.* (1998), we allow the potential measurement times to be random and regard the response over time as a stochastic process with arbitrary dependence structures and distributional forms. Unlike the previous authors, we allow the potential measurement times to depend on covariates. Furthermore, we allow the covariates to be both discrete and continuous, and even time-dependent. In the special two-sample case, the proposed procedure is computationally simpler than that of Yao *et al.* (1998).

2. Proposed methods

Consider a random sample of n subjects. For $i = 1, \ldots, n$, let $Y_i(t)$ be the response at time t and $X_i(t)$ be a $p \times 1$ vector of possibly time-dependent covariates. As in the ACTG 128 Study, the measurements of $Y_i(\cdot)$ are taken at a limited number of possibly irregular and subject-specific time points. Let $N_i(t)$ denote the number of measurements taken on the ith subject by time t, which arises from an arbitrary counting process that is censored at the drop-out time. Specifically, $N_i(t) = N_i^*(t \wedge T_i)$, where $N_i^*(t)$ is a counting process in discrete or continuous time representing the potential measurement times, T_i is the drop-out or censoring time, and $a \wedge b = \min(a, b)$. The process $Y_i(t)$ is observed only at the jump points of $N_i(t)$. The covariate histories $\{X_i(t) : 0 \leqslant t \leqslant T_i\}$ $(i = 1, \ldots, n)$ are assumed to be observed.

In a typical longitudinal study, there are both informative and non-informative drop-outs; the latter may arise from administrative censoring (i.e. study termination) or random loss to follow-up. In the ACTG 128 Study, there were 102 informative drop-outs and 322 non-informative drop-outs. Let D_i and C_i denote, respectively, the latent times to informative drop-out/dependent censoring and non-informative drop-out/independent censoring; D_i is allowed to depend on $Y_i(\cdot)$ even conditional on $X_i(\cdot)$, whereas C_i is independent of $Y_i(\cdot)$ conditional on $X_i(\cdot)$. By definition, $T_i = D_i \wedge C_i$. Write $\delta_i = I(D_i \leqslant C_i)$, where $I(\cdot)$ is the indicator function.

We relate $Y_i(t)$ to $X_i(t)$ through the semiparametric linear regression model:

$$Y_i(t) = \beta_0' X_i(t) + \epsilon_i(t), \quad i = 1, \ldots, n, \tag{1}$$

where β_0 is a $p \times 1$ vector of unknown regression parameters, and $\epsilon_i(\cdot)$ is an arbitrary stochastic process independent of $X_i(\cdot)$. This model is more flexible than the conventional parametric models for longitudinal data in that the evolution of the response over time, i.e. the mean function of $\epsilon_i(\cdot)$ is unspecified. In the absence of informative drop-outs, model (1) has been studied by Zeger and Diggle (1994), Moyeed and Diggle (1994), Lin and Ying (2001) and Lin and Carroll (2001) among others.

We assume that the potential measurement times are non-informative in that

$$E\{\mathrm{d}N_i^*(t) | X_i(t), Y_i(t), C_i, D_i\} = E\{\mathrm{d}N_i^*(t) | X_i(t)\}, \quad i = 1, \ldots, n. \tag{2}$$

In fact, the existing literature on informative drop-outs requires that the potential measurement times be independent of all other random variables so that

$$E\{\mathrm{d}N_i^*(t) | X_i(t), Y_i(t), C_i, D_i\} = E\{\mathrm{d}N_i^*(t)\}, \quad i = 1, \ldots, n. \tag{3}$$

Possibilities of relaxing condition (2) will be discussed in Section 5.

Following Lin and Ying (2001), we specify the semiparametric proportional rates model for $N^*(\cdot)$:

$$E\{\mathrm{d}N_i^*(t)|X_i(t)\} = \mathrm{e}^{\gamma_0' Z_i(t)}\mathrm{d}\Lambda_0(t), \quad i = 1, \ldots, n, \tag{4}$$

where $Z_i(t)$ is the part of $X_i(t)$ that affects $N^*(t)$, γ_0 is a vector of unknown regression parameters, and $\Lambda_0(\cdot)$ is an arbitrary non-decreasing function (Pepe and Cai, 1993; Lawless *et al.*, 1997; Lin *et al.*, 2000). This model is trivially satisfied with $\gamma_0 = 0$ if condition (3) holds.

Let W_i be the subset or function of X_i that may affect the dependent censoring time D_i. We restrict W_i to comprise of bounded covariates only. For simplicity of description, W_i $(i = 1, \ldots, n)$ are assumed to be time-invariant; generalizations to the setting of time-dependent covariates will be discussed in Section 5. In the ACTG 128 Study and other randomized clinical trials, the primary interest lies in the comparison of two treatment groups, in which case X_i, Z_i and W_i all become the treatment indicator, although additional covariates may be needed if there is stratification or clustering.

We suppose that there exist unknown constant vectors η_0 and β_0 such that, for given X_i and t, the bivariate random vectors $\{Y_i(t) - \beta_0' X_i(t),\, D_i \mathrm{e}^{-\eta_0' W_i}\}'$ or $\{Y_i(t) - \beta_0' X_i(t),\, \log D_i - \eta_0' W_i\}'$ $(i = 1, \ldots, n)$ have a common, but completely unspecified joint distribution. Under this assumption, the marginal distribution of $Y_i(t)$ satisfies model (1) while that of D_i satisfies the accelerated failure time model (Kalbfleisch and Prentice, 1980, pp. 32–34; Cox and Oakes, 1984, pp. 64–65). The joint model is stronger than the collection of the two marginal models, although it is much less restrictive than the full parametric model assumed in most of the existing literature. It is implicit in the model statement that the dependence structure between Y and D, which characterizes 'informativeness', must be constant across covariate levels. The independent censoring time C_i is allowed to depend on $X_i(\cdot)$ in an arbitrary manner.

Define $\widetilde{T}_i(\eta) = T_i \mathrm{e}^{-\eta' W_i + d}$ and $\xi_i(t; \eta) = I\{\widetilde{T}_i(\eta) \geqslant t\}$, where $d = \min_{1 \leqslant i \leqslant n} \eta' W_i$. Note that the d so chosen is the largest possible value such that $\widetilde{T}_i(\eta) \leqslant T_i$ for all i. For fixed γ_0 and η_0, we propose to estimate β_0 by the estimating function

$$U_\beta(\beta; \gamma_0, \eta_0) = \sum_{i=1}^n \int_0^\infty \{X_i(t) - \overline{X}(t; \gamma_0, \eta_0)\} \times \left[Y_i(t) - \overline{Y}^*(t; \gamma_0, \eta_0) - \beta' \{X_i(t) - \overline{X}(t; \gamma_0, \eta_0)\} \right] \xi_i(t; \eta_0) \mathrm{d}N_i^*(t),$$

where

$$\overline{X}(t; \gamma, \eta) = \frac{\sum_{i=1}^n \xi_i(t; \eta) \mathrm{e}^{\gamma' Z_i(t)} X_i(t)}{\sum_{i=1}^n \xi_i(t; \eta) \mathrm{e}^{\gamma' Z_i(t)}}, \quad \overline{Y}^*(t; \gamma, \eta) = \frac{\sum_{i=1}^n \xi_i(t; \eta) \mathrm{e}^{\gamma' Z_i(t)} Y_i^*(t)}{\sum_{i=1}^n \xi_i(t; \eta) \mathrm{e}^{\gamma' Z_i(t)}},$$

and $Y_i^*(t)$ is the measurement of Y_i taken at the time point nearest to t. Note that $\xi_i(t; \eta_0) = I\{\widetilde{T}_i(\eta_0) \geqslant t\} = I(T_i \mathrm{e}^{-\eta_0' W_i + d_0} \geqslant t)$, where d_0 is d evaluated at $\eta = \eta_0$. If $\eta_0 = 0$, then $U_\beta(\beta; \gamma_0, \eta_0)$ reduces to $U(\beta; \gamma_0)$ given in Section 2.3 of Lin and Ying (2001); otherwise, $U_\beta(\beta; \gamma_0, \eta_0)$ involves artificial censoring in that the actual censoring times T_i $(i = 1, \ldots, n)$ are replaced by the potentially smaller, artificially created censoring times $\widetilde{T}_i(\eta_0) \equiv T_i \mathrm{e}^{-\eta_0' W_i + d_0}$ in the estimation.

To provide some insights into the artificial censoring, we consider the important special case in which W is a single 0–1 treatment indicator. Then

$$\widetilde{T}_i(\eta_0) = \begin{cases} T_i \mathrm{e}^{-\eta_0 W_i} & \text{if } \eta_0 \geqslant 0, \\ T_i \mathrm{e}^{-\eta_0 W_i + \eta_0} & \text{if } \eta_0 < 0. \end{cases}$$

Thus, we artificially shorten the censoring times by a factor of $\mathrm{e}^{|\eta_0|}$ either in the treatment or control group dependent on whether $\eta_0 > 0$ or $\eta_0 < 0$. In either case, the resultant artificial censoring times

(or more precisely, $D_i \mathrm{e}^{-\eta_0 W_i + d_0}$) are stochastically equal between the two groups under the assumed accelerated failure time model. It is important to realize that, since one cannot uncensor a censored observation, the artificial censoring times cannot be larger than the actual censoring times. For general W, the construction of the $\widetilde{T}_i(\eta)$ is still to make the artificial censoring times independent of covariates subject to the constraints that $\widetilde{T}_i(\eta) \leqslant T_i$ $(i = 1, \ldots, n)$.

We now explain why the artificial censoring yields a valid estimating function for informative drop-outs. Under the assumed joint model for D_i and $Y_i(\cdot)$, the conditional expectations $E\{Y_i(t) - \beta_0' X_i(t) | D_i \mathrm{e}^{-\eta_0' W_i + d_0} \geqslant t, X_i(t)\}$ $(i = 1, \ldots, n)$ have a common value, $\alpha_0(\cdot)$ say, i.e.

$$E\left\{Y_i(t) - \beta_0' X_i(t) \middle| D_i \mathrm{e}^{-\eta_0' W_i + d_0} \geqslant t, X_i(t)\right\} = \alpha_0(t), \quad i = 1, \ldots, n. \tag{5}$$

Define

$$\mathcal{M}_i(t) = \int_0^t \xi_i(s; \eta_0)\{\mathrm{d}N_i^*(s) - \mathrm{e}^{\gamma_0' Z_i(s)} d\Lambda_0(s)\}, \quad i = 1, \ldots, n,$$

$$M_i(t) = \int_0^t \xi_i(s; \eta_0)\left[\left\{Y_i(s) - \beta_0' X_i(s)\right\} \mathrm{d}N_i^*(s) - \mathrm{e}^{\gamma_0' Z_i(s)} \mathrm{d}\mathcal{A}_0(s)\right], \quad i = 1, \ldots, n,$$

where $\mathcal{A}_0(t) = \int_0^t \alpha_0(s)\mathrm{d}\Lambda_0(s)$. Under condition (2) and model (4), $E\{\mathcal{M}_i(t)\} = 0$ $(i = 1, \ldots, n)$ (Lin *et al.*, 2000). Furthermore, by (2), (4), (5) and the assumption that C_i is independent of $Y_i(t)$ conditional on $X_i(t)$, we have

$$\begin{aligned}
E\{M_i(t)\} &= E\int_0^t \xi_i(s; \eta_0)\left[\left\{Y_i(s) - \beta_0' X_i(s)\right\} E\{\mathrm{d}N_i^*(s) | Y_i(s), X_i(s), C_i, D_i\} - \mathrm{e}^{\gamma_0' Z_i(s)} \mathrm{d}\mathcal{A}_0(s)\right] \\
&= E\int_0^t \xi_i(s; \eta_0)\left[\left\{Y_i(s) - \beta_0' X_i(s)\right\} \mathrm{e}^{\gamma_0' Z_i(s)} \mathrm{d}\Lambda_0(s) - \mathrm{e}^{\gamma_0' Z_i(s)} \mathrm{d}\mathcal{A}_0(s)\right] \\
&= E\int_0^t \xi_i(s; \eta_0) \mathrm{e}^{\gamma_0' Z_i(s)}\left[E\left\{Y_i(s) - \beta_0' X_i(s) \middle| T_i \mathrm{e}^{-\eta_0' W_i + d_0} \geqslant s, X_i(s)\right\} \mathrm{d}\Lambda_0(s) - \mathrm{d}\mathcal{A}_0(s)\right] \\
&= E\int_0^t \xi_i(s; \eta_0) \mathrm{e}^{\gamma_0' Z_i(s)} \{\alpha_0(s)\mathrm{d}\Lambda_0(s) - \mathrm{d}\mathcal{A}_0(s)\} = 0.
\end{aligned}$$

Simple algebraic manipulations yield

$$U_\beta(\beta_0; \gamma_0, \eta_0) = \sum_{i=1}^n \int_0^\infty \left\{X_i(t) - \overline{X}(t; \gamma_0, \eta_0)\right\}\left[\mathrm{d}M_i(t) - \{\overline{Y}^*(t; \gamma_0, \eta_0) - \beta_0' \overline{X}(t; \gamma_0, \eta_0)\}\mathrm{d}\mathcal{M}_i(t)\right],$$

which shows that $U_\beta(\beta_0; \gamma_0, \eta_0)$ is a sum of integrals with respect to zero-mean processes. It then follows from the empirical process arguments as given in the appendices of Lin *et al.* (2000) and Lin and Ying (2001) that $U_\beta(\beta_0; \gamma_0, \eta_0)$ is asymptotically centered at 0.

In general, γ_0 and η_0 are unknown and need to be estimated first. Let $N_i^D(t; \eta) = \delta_i I(T_i \mathrm{e}^{-\eta' W_i} \leqslant t)$ $(i = 1, \ldots, n)$. Then a consistent estimator $\widehat{\eta}$ of η_0 can be obtained from the estimating function

$$U_\eta(\eta) = \sum_{i=1}^n \int_0^\infty \{W_i - \overline{W}(t; \eta)\}\mathrm{d}N_i^D(t; \eta),$$

where $\overline{W}(t; \eta) = \sum_{i=1}^n I(T_i \mathrm{e}^{-\eta' W_i} \geqslant t) W_i / \sum_{i=1}^n I(T_i \mathrm{e}^{-\eta' W_i} \geqslant t)$ (Tsiatis, 1990; Wei *et al.*, 1990; Lai and Ying, 1991). In addition, γ_0 can be consistently estimated by $\widehat{\gamma}$, the root of

$$U_\gamma(\gamma) = \sum_{i=1}^n \int_0^\infty \{Z_i(t) - \overline{Z}(t; \gamma)\}\mathrm{d}N_i(t),$$

where $\overline{Z}(t;\gamma) = \sum_{i=1}^n I(T_i \geqslant t)\mathrm{e}^{\gamma' Z_i(t)} Z_i(t) / \sum_{i=1}^n I(T_i \geqslant t)\mathrm{e}^{\gamma' Z_i(t)}$ (Pepe and Cai, 1993; Lawless *et al.*, 1997; Lin *et al.*, 2000). Given $\widehat{\gamma}$ and $\widehat{\eta}$, we estimate β by solving $U_\beta(\beta;\widehat{\gamma},\widehat{\eta}) = 0$. The resulting estimator takes an explicit form

$$\widehat{\beta} = \left[\sum_{i=1}^n \int_0^\infty \{X_i(t) - \overline{X}(t;\widehat{\gamma},\widehat{\eta})\}^{\otimes 2} \xi_i(t;\widehat{\eta})\mathrm{d}N_i(t)\right]^{-1} \times \sum_{i=1}^n \int_0^\infty \{X_i(t) - \overline{X}(t;\widehat{\gamma},\widehat{\eta})\}\{Y_i(t) - \overline{Y}^*(t;\widehat{\gamma},\widehat{\eta})\}\xi_i(t;\widehat{\eta})\mathrm{d}N_i(t),$$

where $a^{\otimes 2} = aa'$. If γ_0 is known to be 0, then one simply solves $U_\beta(\beta;0,\widehat{\eta}) = 0$. In the absence of informative drop-outs, $\widehat{\beta}$ becomes the estimator of Lin and Ying (2001), which further reduces to the ordinary least-squares estimator if there is a single measurement per subject taken at the same time point.

We show in the Appendix that $\widehat{\beta}$ is consistent and asymptotically normal. It is, however, difficult to estimate analytically the limiting covariance matrix for $\widehat{\beta}$ because the $\widehat{\eta}$ involved in $U_\beta(\beta;\widehat{\gamma},\widehat{\eta})$ is a rank-based estimator. We shall estimate the covariance matrix of $\widehat{\beta}$ and construct confidence intervals for individual components of β_0 by adopting a resampling technique due to Parzen *et al.* (1994). Specifically, let

$$\begin{aligned} U_{\beta i} &= \int_0^\infty \left\{X_i(t) - \overline{X}(t;\widehat{\gamma},\widehat{\eta})\right\}\left[\mathrm{d}\widehat{M}_i(t) - \{\overline{Y}^*(t;\widehat{\gamma},\widehat{\eta}) - \widehat{\beta}'\overline{X}(t;\widehat{\gamma},\widehat{\eta})\}\mathrm{d}\widehat{\mathcal{M}}_i(t)\right] \\ &\quad - \widehat{H}\widehat{\Omega}^{-1}\int_0^\infty \{Z_i(t) - \overline{Z}(t;\widehat{\gamma})\}\mathrm{d}\widehat{\mathcal{M}}_i^\dagger(t), \\ U_{\eta i} &= \int_0^\infty \{W_i - \overline{W}(t;\widehat{\eta})\}\mathrm{d}\widehat{\mathcal{M}}_i^D(t), \end{aligned}$$

where $\widehat{\Omega} = -n^{-1}\partial U_\gamma(\widehat{\gamma})/\partial\gamma$,

$$\widehat{H} = n^{-1}\sum_{i=1}^n \int_0^\infty \partial\overline{X}(t;\widehat{\gamma},\widehat{\eta})/\partial\gamma\ \{Y_i(t) - \overline{Y}^*(t;\widehat{\gamma},\widehat{\eta})\}\xi_i(t;\widehat{\eta})\mathrm{d}N_i(t),$$

$$\widehat{M}_i(t) = \int_0^t \xi_i(s;\widehat{\eta})\left[\{Y_i(s) - \widehat{\beta}'X_i(s)\}\mathrm{d}N_i^*(s) - \mathrm{e}^{\widehat{\gamma}'Z_i(s)}\mathrm{d}\widehat{\mathcal{A}}(s)\right],$$

$$\widehat{\mathcal{M}}_i(t) = \int_0^t \xi_i(s;\widehat{\eta})\{\mathrm{d}N_i^*(s) - \mathrm{e}^{\widehat{\gamma}'Z_i(s)}d\widehat{\Lambda}(s)\}, \quad \widehat{\mathcal{M}}_i^\dagger(t) = N_i(t) - \int_0^t I(T_i \geqslant s)\mathrm{e}^{\widehat{\gamma}'Z_i(s)}\mathrm{d}\widehat{\Lambda}(s),$$

$$\widehat{\mathcal{M}}_i^D(t) = N_i^D(t;\widehat{\eta}) - \int_0^t I(T_i\mathrm{e}^{-\widehat{\eta}'W_i} \geqslant s)\mathrm{d}\widehat{\Lambda}^D(s),$$

$$\widehat{\mathcal{A}}(t) = \sum_{i=1}^n \int_0^t \frac{\xi_i(s;\widehat{\eta})\{Y_i(s) - \widehat{\beta}'X_i(s)\}\mathrm{d}N_i(s)}{\sum_{j=1}^n \xi_j(s;\widehat{\eta})\mathrm{e}^{\widehat{\gamma}'Z_j(s)}},$$

$$\widehat{\Lambda}(t) = \sum_{i=1}^n \int_0^t \frac{\mathrm{d}N_i(s)}{\sum_{j=1}^n I(T_j \geqslant s)\mathrm{e}^{\widehat{\gamma}'Z_j(s)}}, \quad \widehat{\Lambda}^D(t) = \sum_{i=1}^n \int_0^t \frac{\mathrm{d}N_i^D(s;\widehat{\eta})}{\sum_{j=1}^n I(T_j\mathrm{e}^{-\widehat{\eta}'W_j} \geqslant s)}.$$

Also, let $(G_1,\ldots,G_n)$ be a random sample of standard normal variables. Define $\widehat{\beta}_G$ to be the solution to $U_\beta(\beta;\widehat{\gamma},\widehat{\eta}_G) = \sum_{i=1}^n U_{\beta i}G_i$, where $\widehat{\eta}_G$ is the solution to $U_\eta(\eta) = \sum_{i=1}^n U_{\eta i}G_i$. If γ_0 is known to be 0, we solve $U_\beta(\beta;0,\widehat{\eta}_G) = \sum_{i=1}^n \widetilde{U}_{\beta i}G_i$ instead, where $\widetilde{U}_{\beta i}$ is the first term of $U_{\beta i}$ with

$\widehat{\gamma}$ replaced by 0. We show in the Appendix that the conditional distribution of $n^{1/2}(\widehat{\beta}_G - \widehat{\beta})$ given the data $\{Y_i(\cdot), X_i(\cdot), N_i(\cdot), T_i, \delta_i\}$ $(i = 1, \ldots, n)$ is asymptotically the same as the unconditional distribution of $n^{1/2}(\widehat{\beta} - \beta_0)$. Thus, to approximate the distribution of $\widehat{\beta}$, we obtain a large number of realizations of $\widehat{\beta}_G$ by repeatedly generating the normal random sample $(G_1, \ldots, G_n)$ while fixing the data $\{Y_i(\cdot), X_i(\cdot), N_i(\cdot), T_i, \delta_i\}$ $(i = 1, \ldots, n)$ at their observed values. The covariance matrix of $\widehat{\beta}$ can then be estimated by the empirical covariance matrix of $\widehat{\beta}_G$. In addition, confidence intervals for individual components of β_0 can be constructed from the empirical percentiles of $\widehat{\beta}_G$ or by the Wald method.

3. Numerical studies

A series of simulation experiments was carried out to assess the properties of the proposed estimators in practical settings. The responses were generated from the random-effect model

$$Y(t) = a_0 + a_1 \sin t + \beta_{01} X_1 + \beta_{02} X_2 + \epsilon^*(t), \tag{6}$$

where X_1 is Bernoulli with 0.5 success probability, X_2 is independent standard normal, $\epsilon^*(t)$ is normal with mean ϕ and variance σ_ϵ^2 for all t, and ϕ is zero-mean normal with variance σ_ϕ^2. The value of σ_ϕ^2 determines the degree of dependence among the responses from the same subject. For randomized clinical trials, X_1 and X_2 represent, respectively, the treatment indicator and baseline covariate. Model (6) is clearly a special case of (1).

We considered both the situations of fixed and random measurement times. For the former, the potential measurement times were set to be integers $1, 2, \ldots$; for the latter, the potential measurement times were generated from a random-effect Poisson process $N^*(\cdot)$ with intensity rate $\psi e^{\gamma_{01} X_1 + \gamma_{02} X_2}$, where ψ is an independent gamma random variable with mean 1 and variance 0.25. Because of the random effect ψ, the measurement times within the same subject are positively correlated. The dependent censoring time D was set to be $D_0 e^{\eta_0 X_1}$, where D_0 is the exponential of a normal random variable with mean $\mu_0 + \phi$ and variance σ_ϵ^2. The dependence between D and Y was induced by the common random effect ϕ. Specifically, the correlation between $\log D$ and Y is $\sigma_\phi^2/(\sigma_\phi^2 + \sigma_\epsilon^2)$, which is also the correlation between two measurements of Y on the same subject. The independent censoring time C was a uniform $(0, \tau)$ variable. The values of μ_0 and τ control the levels of dependent and independent censoring, respectively.

In the situation of fixed measurement times, γ_0 is known to be 0 so that β_0 is estimated by solving the equation $U_\beta(\beta; 0, \widehat{\eta}) = 0$. In the case of random measurement times, the general estimator solving $U_\beta(\beta; \widehat{\gamma}, \widehat{\eta}) = 0$ is used. For comparisons, we also evaluated the estimators of Lin and Ying (2001), which are referred to as the naive estimators because they are valid only under independent censoring.

Table 1 summarizes the main results with $n = 200$, $a_0 = 0$, $a_1 = \beta_{01} = \beta_{02} = 1$, $\eta_0 = 0.5$, $\sigma_\epsilon = 1$ and $\tau = 30$. For random measurement times, $\gamma_{01} = 0.5$ and $\gamma_{02} = -0.5$. The numbers of actual measurement times are approximately 2, 4 and 8, and the percentages of dependent censoring are approximately 90%, 80% and 50% under $\mu_0 = 0$, 1 and 2, respectively. The results in Table 1 pertain to the estimation of β_{01}. The results for β_{02} are similar and thus omitted.

In both the situations of fixed and random measurement times, the proposed parameter estimators are virtually unbiased. The variance estimators reflect well the true variabilities of the parameter estimators, and the confidence intervals have proper coverage probabilities. Note that there is no dependent censoring under $\sigma_\phi = 0$. In this case, both the naive estimators and proposed estimators are valid, and the former are slightly more efficient than the latter. In the presence of dependent censoring, however, the naive estimators are biased and the associated confidence intervals do not have adequate coverage probabilities. The problems worsen as the level of dependent censoring or/and the degree of dependence increases.

Table 1. *Summary statistics for the first series of simulation studies*

			Fixed measurement times				Random measurement times			
μ_0	σ_ϕ	Estimator	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
0	0.0	Proposed	0.005	0.140	0.149	0.96	−0.003	0.135	0.135	0.95
		Naive	0.004	0.125	0.123	0.94	0.002	0.124	0.120	0.94
	0.5	Proposed	0.000	0.189	0.189	0.94	0.000	0.183	0.182	0.94
		Naive	−0.069	0.169	0.158	0.89	−0.060	0.179	0.163	0.90
	1.0	Proposed	−0.002	0.271	0.272	0.95	−0.009	0.280	0.272	0.95
		Naive	−0.165	0.235	0.216	0.85	−0.164	0.254	0.231	0.85
1	0.0	Proposed	0.000	0.084	0.088	0.95	0.001	0.089	0.089	0.94
		Naive	−0.002	0.075	0.074	0.95	0.000	0.081	0.078	0.94
	0.5	Proposed	−0.006	0.136	0.135	0.94	−0.001	0.145	0.146	0.95
		Naive	−0.065	0.128	0.117	0.90	−0.049	0.137	0.131	0.92
	1.0	Proposed	0.002	0.224	0.227	0.96	0.005	0.246	0.243	0.95
		Naive	−0.151	0.191	0.182	0.84	−0.142	0.225	0.205	0.86
2	0.0	Proposed	0.000	0.059	0.062	0.96	−0.004	0.066	0.066	0.95
		Naive	0.000	0.051	0.051	0.95	−0.003	0.060	0.056	0.94
	0.5	Proposed	0.001	0.110	0.113	0.95	−0.003	0.122	0.126	0.95
		Naive	−0.046	0.101	0.099	0.90	−0.046	0.120	0.115	0.92
	1.0	Proposed	0.001	0.194	0.207	0.96	0.008	0.221	0.227	0.95
		Naive	−0.117	0.175	0.163	0.87	−0.110	0.205	0.190	0.89

Note: Bias is the sampling mean of $\widehat{\beta}_1$ minus β_{01}; SSE is the sampling standard error of $\widehat{\beta}_1$; SEE is the sampling mean of the standard error estimator for $\widehat{\beta}_1$; CP is the coverage probability of the 95% Wald confidence interval forβ_{01}. Each entry is based on 1000 simulated data sets. For each data set, 1000 simulation samples were used to estimate the sampling distribution of $\widehat{\beta}_1$. The naive method pertains to Lin and Ying (2001).

Table 2. *Summary statistics for the second series of simulation studies*

Censoring level			$\Gamma_1^* = \Gamma_2^*$			$\Gamma_1^* \neq \Gamma_2^*$		
Group 1	Group 2	β_0	Bias	$\alpha = 0.05$	$\alpha = 0.10$	Bias	$\alpha = 0.05$	$\alpha = 0.10$
20%	20%	0	0.03	0.05	0.09	−0.42	0.06	0.13
		3	0.03	0.88	0.93	−0.42	0.74	0.84
		4	0.03	0.99	1.00	−0.42	0.96	0.98
15%	30%	0	0.03	0.05	0.10	−0.52	0.07	0.14
		3	0.03	0.87	0.93	−0.52	0.70	0.80
		4	0.03	0.99	1.00	−0.52	0.94	0.97

Note: Bias is the sampling mean of $\widehat{\beta}$ minus β_0. The numbers under $\alpha = 0.05$ and $\alpha = 0.10$ pertain to the empirical Type I error/power of the Wald test at the nominal levels of 0.05 and 0.10, respectively.

Incidentally, the results of the simulation do not depend on the actual values of β_0, and are virtually identical when $\sin t$ in (6) is replaced by other functions, such as $t^{1/2}$.

To compare the proposed methods with those of Yao *et al.* (1998) and to assess the sensitivity to the assumption that the degree of dependence between Y and D does not depend on covariates, we considered the simulation set-up of Yao *et al.* Specifically, there were 100 subjects in each of the two groups, and every

subject had two potential measurements at months 24 and 48. We generated $\{Y(24), Y(48), \log D\}$ from the normal distributions with means $(85, 75, \mu_1)$ and $(85+\beta_0, 75+\beta_0, \mu_2)$, and covariance matrices Γ_1^* and Γ_2^* for groups 1 and 2, respectively, where μ_1 and μ_2 were chosen to yield desired levels of censoring. These simulation studies were designed to mimic the ACTG 128 Study. The results are shown in Table 2. Under $\Gamma_1^* = \Gamma_2^*$,

$$\Gamma_1^* = \Gamma_2^* = \begin{bmatrix} 50 & 40 & 0.56 \\ 40 & 50 & 0.84 \\ 0.56 & 0.84 & 0.04 \end{bmatrix};$$

under $\Gamma_1^* \neq \Gamma_2^*$,

$$\Gamma_1^* = \begin{bmatrix} 50 & 40 & 0.56 \\ 40 & 50 & 0.84 \\ 0.56 & 0.84 & 0.04 \end{bmatrix}, \quad \Gamma_2^* = \begin{bmatrix} 50 & 40 & 0.14 \\ 40 & 50 & 0.28 \\ 0.14 & 0.28 & 0.04 \end{bmatrix}.$$

The results in Table 2 are similar to those on W_3 shown in Table 1 of Yao *et al.* (1998). The proposed methods appear to be robust to violation of the assumption of constant dependence.

4. Examples

4.1 *ACTG 128 Study*

We now apply the proposed methods to data taken from the ACTG 128 Study. Since this was a randomized clinical trial, we first consider the two-sample model, in which X is a scalar covariate indicating by the values 1 or 0 whether the patient received low-dose or high-dose AZT. The observed value of $\widehat{\gamma}$ turns out to be -0.017 with an estimated standard error 0.043, which suggests that the potential measurement times were stochastically similar between the two groups. On the other hand, the observed value of $\widehat{\eta}$ is 0.38 with an estimated standard error of 0.13, which shows that the dependent censoring times tend to be shorter for the high-dose patients than the low-dose patients.

In estimating the treatment difference in the IQ score trajectory, we assume that it would take three months for the AZT to take effect. Thus, we exclude the IQ scores measured before month 3 when fitting model (1). The estimation results are displayed in the top panel of Table 3, both for the proposed method and for the naive method of Lin and Ying (2001). Approximately 8% of the measurements are artificially censored under the proposed method. Using either method, one would conclude that the low-dose AZT is at least as effective as the high-dose AZT. One possible explanation for the lack of a dramatic difference between the two methods is that the dependence of the loss to follow-up on the IQ score is weak, especially for the cases of drug toxicities and voluntary withdrawals. It is reassuring, however, that the adjustment for dependent censoring does not alter the conclusion of the analysis.

Standard analysis using generalized estimating equations (GEE) (Liang and Zeger, 1984) or random-effects models (Laird and Ware, 1982) also provided similar results for this study. For example, assuming a linear time trend and an exchangeable correlation matrix, both the GEE analysis and random-intercept model yielded estimated treatment effect of 3.28 with estimated standard error of 1.74.

For further illustration, we add the IQ score at the baseline to the model. The resulting estimates are displayed in the bottom panel of Table 3. The baseline IQ score is highly significant. The estimates for the treatment difference becomes smaller.

4.2 *ACTG 117/118 Study*

The ACTG 117/118 Study was conducted to evaluate the effects of switching to didanosine (ddI) from zidovudine (AZT) for HIV-infected subjects who have tolerated the AZT treatment for at least 16 weeks

Table 3. *Regression analysis for the ACTG 128 Study*

	Proposd Method				Naive Method			
Parameter	Est	SE	Est/SE	95%CI	Est	SE	Est/SE	95%CI
Treatment	3.04	1.89	1.61	(−0.67, 6.75)	3.43	1.88	1.83	(−0.25, 7.10)
Treatment	1.37	1.30	1.05	(−1.19, 3.92)	1.89	1.32	1.43	(−0.70, 4.48)
Baseline IQ	0.65	0.06	10.45	(0.53, 0.78)	0.65	0.06	10.62	(0.53, 0.77)

Note: Est stands for estimate, SE for (estimated) standard error, and 95%CI for the Wald 95% confidence interval. The naive method pertains to Lin and Ying (2001).

Table 4. *Regression analysis for the ACTG 117/118 Study*

	Proposed method				Naive method			
Time	Est	SE	Est/SE	95% CI	Est	SE	Est/SE	95% CI
Week 8	28.22	9.87	2.86	(8.86, 47.58)	27.60	9.58	2.88	(8.82, 46.39)
Week 16	23.53	11.87	1.98	(0.27, 46.79)	19.24	10.47	1.84	(−1.27, 39.76)
Week 24	18.00	12.99	1.38	(−7.47, 43.47)	12.32	10.34	1.19	(−7.95, 32.60)
Overall	23.84	10.48	2.27	(3.29, 44.38)	20.16	9.27	2.17	(1.98, 38.34)

Note: see Note of Table 2.

(Kahn *et al.*, 1992). A total of 304 patients were randomly chosen to continue the AZT therapy while 298 patients were assigned to ddI. The investigators were interested in comparing the CD4 cell counts between the two groups at weeks 8, 16 and 24.

There were 100, 24, 16, 9 and 25 patients in the AZT group who dropped out of the study due to patient's request, physician's decision, toxicities, death and other reasons, respectively; the corresponding numbers in the ddI group were 62, 24, 19, 10 and 32. With X being the indicator for ddI, the observed value of $\widehat{\eta}$ is 0.35 with an estimated standard error of 0.09. This value of $\widehat{\eta}$ results in approximately 8% artificial censoring. A Cox regression analysis reveals that patients with sharp declines in the CD count tend to drop out of the study sooner than those with gradual changes.

Table 4 shows the estimated differences in the mean CD count between the ddI and AZT groups at weeks 8, 16 and 24; the overall estimates pertain to a common treatment difference over the three time points. The proposed method yields larger estimates of the treatment differences than the naive method, especially at weeks 16 and 24. One would expect the naive method to underestimate the effects of ddI because the drop-out times tend to be shorter for the AZT patients than for the ddI patients, and the patients with lower CD4 counts tend to drop out earlier. Assuming a linear time trend and an unstructured correlation matrix, the estimate for the overall treatment difference is approximately 21.8 with an estimated standard error of 8.7 either a GEE analysis or a random-intercept model is used.

5. REMARKS

Informative drop-out is one of the most important and challenging problems in longitudinal data analysis. Much of the existing work in this area relies on full parametric modelling of the dependence between the drop-out time and repeated measures. The approach taken in this paper is more non-parametric: neither the distributional forms nor the dependence structures are parametrized. Thus, the proposed methods should be more robust. For testing the null hypothesis of no treatment difference in a randomized clinical trial, the only assumption required is that the correlation between D and Y does not depend on the treatment group. As shown in Section 3, moderate departures from this assumption have miminal effects on the Type I error.

It is evident from the derivations in Section 2 that, although we model the joint distribution of D_i and $Y_i(t)$, what is really required is the assumption on the conditional means given in (5), which seems weaker than the full joint model. It is possible to check assumption (5) empirically on the basis of the residuals $\widehat{M}_i(\cdot)$ $(i = 1, \ldots, n)$; see Lin and Ying (2001). However, the parameter vector β_0 in (5) is not necessarily the same as β_0 in the marginal distribution of $Y(\cdot)$. The equality of the two parameter vectors is not testable, but is essential to interpreting β_0 as causal effects. A sufficient condition for the equality to hold is assumption (2.8) of Robins (1995), which essentially states that the probability distribution for the unobservable data is the same as that of the observable data.

The proposed estimators of β_0 may be inefficient if the range of $\eta' W$ is large because then there may be excessive artificial censoring. We may alleviate this problem through stratification. Suppose, for example, that the large range of $\eta' W$ is caused by some extreme values in a particular component of W. Then we may stratify the sample on this covariate, one stratum containing the middle values of this covariate and the other containing the rest. We can construct the estimating functions for β_0 separately in the two strata by using different values of d, and then combine the two estimating functions by adding them together. In doing so, the stratum with the middle covariate values is spared from excessive artificial censoring.

The covariate vector W has been restricted to be time invariant. If W is time varying, then we specify that, for given X_i and t, the bivariate random vectors $\{Y_i(t) - \beta_0' X_i(t), \int_0^{D_i} \mathrm{e}^{-\eta_0' W_i(u)} \mathrm{d}u\}'$ $(i = 1, \ldots, n)$ have an arbitrary common joint distribution. We then redefine $\widetilde{T}_i(\eta)$ as $\int_0^{T_i} \mathrm{e}^{-\eta' W_i(t)+h} \mathrm{d}t$, where h is the largest number such that $\widetilde{T}_i(\eta) \leqslant T_i$ for all i, and estimate η by the rank estimators of Robins and Tsiatis (1992) and Lin and Ying (1995). With these modifications, all the results in Section 2 continue to hold.

The proposed methods, as well as all existing methods, require that the potential measurement times are non-informative in the sense of (2). This assumption may be violated if, for example, patients with deteriorating health tend to have their measurements taken less frequently than patients with improving health. One possible way of accommodating such informative measurement times is to assume that there exist constant vectors η_0, γ_0 and β_0 such that $\{D_i \mathrm{e}^{-\eta_0' W_i}, N_i^*(t\mathrm{e}^{\gamma_0' Z_i}), Y_i(t) - \beta_0' X_i(t); t \geqslant 0\}$ $(i = 1, \ldots, n)$ have a common, but completely arbitrary joint distribution. Note that the marginal distribution of $N_i^*(\cdot)$ satisfies the accelerated time model for counting processes (Lin *et al.*, 1998). It can be shown that, with minor modifications, the asymptotic results of Section 2 continue to be valid under this joint model if γ_0 is known to be 0. Inference with general γ_0 is currently under investigation.

APPENDIX A

Asymptotic properties of $\widehat{\beta}$

In this appendix, we derive the asymptotic distribution of $\widehat{\beta}$ and in fact the asymptotic joint distribution of $\widehat{\beta}$ and $\widehat{\eta}$. The resampling method is also justified. The proofs involve arguments previously used by Lin *et al.* (1996) and Lin and Ying (2001), and shall be kept very brief. By Lemma 1 given in Appendix A.1 of Lin and Ying (2001),

$$n^{-1/2} U_\beta(\beta_0; \gamma_0, \eta_0) = n^{-1/2} \sum_{i=1}^{n} \int_0^\infty \{X_i(t) - \overline{x}(t)\} \left[\mathrm{d}M_i(t) - \left\{\overline{y}^*(t) - \beta_0' \overline{x}(t)\right\} \mathrm{d}\mathcal{M}_i(t)\right] + o_p(1), \tag{A.1}$$

$$n^{-1/2} U_\gamma(\gamma_0) = n^{-1/2} \sum_{i=1}^{n} \int_0^\infty \{Z_i(t) - \overline{z}(t)\} \mathrm{d}\mathcal{M}_i^\dagger(t) + o_p(1), \tag{A.2}$$

$$n^{-1/2} U_\eta(\eta_0) = n^{-1/2} \sum_{i=1}^{n} \int_0^\infty \{W_i - \overline{w}(t)\} \mathrm{d}\mathcal{M}_i^D(t) + o_p(1), \tag{A.3}$$

where

$$\mathcal{M}_i^{\dagger}(t) = N_i(t) - \int_0^t I(T_i \geqslant s)\mathrm{e}^{\gamma_0' Z_i(s)}\mathrm{d}\Lambda_0(s), \quad i = 1, \ldots, n,$$

$$\mathcal{M}_i^{D}(t) = N_i^{D}(t;\eta_0) - \int_0^t I(T_i\mathrm{e}^{-\eta_0' W_i} \geqslant s)\mathrm{d}\Lambda_0^{D}(s), \quad i = 1, \ldots, n.$$

$\Lambda_0^D(\cdot)$ is the common cumulative hazard function for $D_i\mathrm{e}^{-\eta_0' W_i}$ $(i = 1, \ldots, n)$, and $\overline{x}(t)$, $\overline{y}^*(t)$, $\overline{w}(t)$ and $\overline{z}(t)$ are the limits of $\overline{X}(t;\gamma_0,\eta_0)$, $\overline{Y}^*(t;\gamma_0,\eta_0)$, $\overline{W}(t;\eta_0)$ and $\overline{Z}(t;\gamma_0)$. It then follows from the Taylor series expansion and the law of large numbers that

$$n^{-1/2}U_\beta(\beta_0;\widehat{\gamma},\eta_0) = n^{-1/2}\sum_{i=1}^n \int_0^\infty \{X_i(t) - \overline{x}(t)\}\left[\mathrm{d}M_i(t) - \left\{\overline{y}^*(t) - \beta_0'\overline{x}(t)\right\}\mathrm{d}\mathcal{M}_i(t)\right]$$

$$-H\Omega^{-1}n^{-1/2}\sum_{i=1}^n \int_0^\infty \{Z_i(t) - \overline{z}(t)\}\mathrm{d}\mathcal{M}_i^{\dagger}(t) + o_p(1), \tag{A.4}$$

where H and Ω are the limits of $\widehat{H}$ and $\widehat{\Omega}$, respectively.

Let $\theta = (\beta', \eta')'$, and θ_0, $\widehat{\theta}$ and $\widehat{\theta}_G$ are similarly defined. Also, write $U(\theta) = \{U_\beta'(\beta;\widehat{\gamma},\eta), U_\eta'(\eta)\}'$. Because the right-hand sides of (A.2) and (A.4) consist of sums of n independent random vectors, the multivariate central limit theorem, together with the Cramér–Wold device, implies that the random vector $n^{-1/2}U(\theta_0)$ is asymptotically zero-mean normal with covariance matrix $B \equiv \lim_{n\to\infty} n^{-1}\sum_{i=1}^n E(u_i^{\otimes 2})$, where $u_i = (u_{\beta i}', u_{\eta i}')'$,

$$u_{\beta i} = \int_0^\infty \{X_i(t) - \overline{x}(t)\}\left[\mathrm{d}M_i(t) - \left\{\overline{y}^*(t) - \beta_0'\overline{x}(t)\right\}\mathrm{d}\mathcal{M}_i(t)\right] - H\Omega^{-1}\int_0^\infty \{Z_i(t) - \overline{z}(t)\}\mathrm{d}\mathcal{M}_i^{\dagger}(t),$$

$$u_{\eta i} = \int_0^\infty \{W_i - \overline{w}(t)\}\mathrm{d}\mathcal{M}_i^{D}(t).$$

The replacements of the unknown quantities in $u_{\beta i}$ and $u_{\eta i}$ with their sample estimators yield $U_{\beta i}$ and $U_{\eta i}$ given in Section 2.

By the arguments of Ying (1993, proof of Theorem 2), we can show that, for θ in a small neighborhood of θ_0,

$$n^{-1/2}U(\theta) = n^{-1/2}U(\theta_0) + An^{1/2}(\theta - \theta_0) + o_p(1), \tag{A.5}$$

where A is a deterministic matrix. As a result, $n^{1/2}(\widehat{\theta} - \theta_0)$ is asymptotically zero-mean normal with covariance matrix $A^{-1}BA^{-1}$.

Clearly, $U(\widehat{\theta}_G) = \sum_{i=1}^n U_i G_i$, where $U_i = (U_{\beta i}', U_{\eta i}')'$. Thus, it follows from (A.5) that

$$n^{1/2}(\widehat{\theta}_G - \widehat{\theta}) = A^{-1}n^{-1/2}\sum_{i=1}^n U_i G_i + o_p(1).$$

The conditional distribution of $n^{-1/2}\sum U_i G_i$ given the data $\{Y_i(\cdot), X_i(\cdot), N_i(\cdot), T_i)$ $(i = 1, \ldots, n)$ is zero-mean normal with covariance matrix $n^{-1}\sum_{i=1}^n U_i^{\otimes 2}$, which converges in probability to B. Consequently, the conditional distribution of $n^{1/2}(\widehat{\theta}_G - \widehat{\theta})$ is asymptotically equivalent to the unconditional distribution of $n^{1/2}(\widehat{\theta} - \theta_0)$.

REFERENCES

BRADY, M. T., MCGRATH, N., BROUWERS, P., GELBER, R., FOWLER, M. G., WEINTRUB, P. S. AND THE PEDIATRIC AIDS CLINICAL TRIALS GROUP (1996). Randomized study of the tolerance and efficacy of high- versus low-dose zidovudine in HIV-infected children with mild-to-moderate symptoms (ACTG 128). *Journal of Infectious Diseases* **173**, 1097–1106.

COX, D. R. AND OAKES, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.

DIGGLE, P. J. AND KENWARD, M. G. (1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics* **43**, 49–93.

GLIDDEN, D. V. AND WEI, L. J. (1999). Rank estimation of treatment differences based on repeated measurements subject to dependent censoring. *Journal of the American Statistical Association* **94**, 888–895.

HOGAN, J. W. AND LAIRD, N. M. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* **16**, 259–272.

KAHN, J. O. *et al.* (1992). A controlled trial comparing continued zidovudine with didanosine in human immunodeficiency virus infection. *The New England Journal of Medicine* **327**, 581–587., ,

KALBFLEISCH, J. D. AND PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

LAI, T. L. AND YING, Z. (1991). Rank regression methods for left-truncated and right-censored data. *The Annals of Statistics* **19**, 531–556.

LAIRD, N. M. (1994). Discussion of the paper by P. J. Diggle and M. G. Kenward, "Informative dropout in longitudinal data analysis". *Applied Statistics* **43**, 84.

LAIRD, N. M. AND WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

LAWLESS, J. F., NADEAU, C. AND COOK, R. J. (1997). Analysis of mean and rate functions for recurrent events. Lin, D. Y. and Fleming, T. R. (eds), *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*. New York: Springer, pp. 37–49.

LIANG, K. Y. AND ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

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 for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society,* Series B **62**, 711–730.

LIN, D. Y., WEI, L. J. AND YING, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**, 605–618.

LIN, D. Y. AND YING, Z. (1995). Semiparametric inference for the accelerated life model with time-dependent covariates. *Journal of Statistical Planning and Inference* **44**, 47–63.

LIN, D. Y. AND YING, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *Journal of the American Statistical Association* **96**, 103–113.

LIN, X. AND CARROLL, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association* **96**, 1045–1056.

LITTLE, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90**, 1112–1121.

MORI, M., WOODWORTH, G. G. AND WOOLSON, R. F. (1992). Application of empirical Bayes inference to estimation of rate change in the presence of informative right censoring. *Statistics in Medicine* **11**, 621–631.

MOYEED, R. A. AND DIGGLE, P. J. (1994). Rates of convergence in semi-parametric modeling of longitudinal data. *Australian Journal of Statistics* **36**, 75–93.

PARZEN, M. I., WEI, L. J. AND YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–350.

PEPE, M. S. AND CAI, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association* **88**, 811–820.

ROBINS, J. M. (1995). An analytic method for randomized trials with informative censoring. *Lifetime Data Analysis* **1**, 241–254 (part I); 417-434 (part II).

ROBINS, J. M. AND ROTNITZKY, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. Jewell, N. P., Dietz, K. and Farewell, V. T. (eds), *AIDS Epidemiology: Methodological Issues*. Boston, MA: Birkhäuser, pp. 297–331.

ROBINS, J. M., ROTNITZKY, A. AND ZHAO, L. P. (1995). Analysis of semiparametric regression-models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106–121.

ROBINS, J. M. AND TSIATIS, A. A. (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika* **79**, 311–319.

ROTNITZKY, A., ROBINS, J. M. AND SCHARFSTEIN, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1311–1339.

RUBIN, D. B. (1994). Discussion of the paper by P. J. Diggle and M. G. Kenward, "Informative dropout in longitudinal data analysis". *Applied Statistics* **43**, 80–82.

SCHARFSTEIN, D., ROTNITZKY, A. AND ROBINS, J. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association* **94**, 1096–1120.

SCHLUCHTER, M. D. (1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine* **11**, 1861–1870.

TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics* **18**, 354–372.

WEI, L. J., YING, Z. AND LIN, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–851.

WU, M. C. AND BAILEY, K. R. (1988). Analyzing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine* **7**, 337–346.

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.

YAO, Q., WEI, L. J. AND HOGAN, J. W. (1998). Analysis of incomplete repeated measurements with dependent censoring times. *Biometrika* **85**, 139–149.

YING, Z. (1993). A large sample study of rank estimation for censored regression data. *The Annals of Statistics* **21**, 76–99.

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.

[*Received April 8, 2002; first revision September 4, 2002; second revision November 11, 2002; accepted for publication November 12, 2002*]