MODELING LEFT-TRUNCATED AND RIGHT-CENSORED SURVIVAL DATA WITH LONGITUDINAL COVARIATES[1]

BY YU-RU SU AND JANE-LING WANG

University of California, Davis and National Cheng Kung University, and University of California, Davis

There is a surge in medical follow-up studies that include longitudinal covariates in the modeling of survival data. So far, the focus has been largely on right-censored survival data. We consider survival data that are subject to both left truncation and right censoring. Left truncation is well known to produce biased sample. The sampling bias issue has been resolved in the literature for the case which involves baseline or time-varying covariates that are observable. The problem remains open, however, for the important case where longitudinal covariates are present in survival models. A joint likelihood approach has been shown in the literature to provide an effective way to overcome those difficulties for right-censored data, but this approach faces substantial additional challenges in the presence of left truncation. Here we thus propose an alternative likelihood to overcome these difficulties and show that the regression coefficient in the survival component can be estimated unbiasedly and efficiently. Issues about the bias for the longitudinal component are discussed. The new approach is illustrated numerically through simulations and data from a multi-center AIDS cohort study.

1. Introduction. Since the seminal paper by Wulfsohn and Tsiatis (1997), longitudinal covariates have played an increasingly important role in the modeling of survival data. One major challenge to incorporate longitudinal covariates is that simple approaches, such as the partial likelihood method for the Cox proportional hazards model [Cox (1972)], often require knowledge of the entire longitudinal process. This is often not feasible in reality for follow-up checks at discrete and intermittent time points. A common practice is to impute the values of the missing longitudinal processes and then apply the partial likelihood approach to the imputed data. This is called a two-stage approach, where the longitudinal process is imputed at the first stage before the partial likelihood approach is employed to estimate parameters in the survival model at the second stage. The most common imputation method is to use the last and most recent value of the patient to impute a missing value, the so-called last-value-carry-forward method, which has been adopted in standard software such as SAS and R. Additional two-stage procedures

were developed by Tsiatis, DeGruttola and Wulfsohn (1995) and Dafni and Tsiatis (1998).

It is easy to foresee serious biases with such an imputation method if the follow-up schedule is infrequent over time and also when the longitudinal covariates are contaminated by noises or measurement errors. Both scenarios provide strong motivation to find alternative approaches. The approach developed by Wulfsohn and Tsiatis (1997) to model the survival and longitudinal data simultaneously through their joint likelihood is attractive on two counts: (i) the resulting parametric estimators are semiparametrically efficient when the baseline hazard function is unknown, and (ii) the joint likelihood procedure is often insensitive to the normality assumption on the longitudinal data, if there is a reasonable number of repeated measurements available for the longitudinal processes; see Zeng and Cai (2005) and Dupuy, Grama and Mesbah (2006) for (i) and Song, Davidian and Tsiatis (2002), Tsiatis and Davidian (2004) and Hsieh, Tseng and Wang (2006) for (ii).

The above joint likelihood approach not only successfully removes the biases on the survival component but also leads to efficient estimation. A historical example for the joint likelihood approach is the investigation of CD4 T-cell counts as a biomarker of time-to-death or time-to-AIDS [DeGruttola and Tu (1994), Wulfsohn and Tsiatis (1997), Henderson, Diggle and Dobson (2000)]. In these and other works, the survival time is subject to the usual right censoring. However, left truncation is common for studies with delayed entry. Specifically, if the recruitment of patients continues after the onset time of a study, those that have already experienced the event are often excluded from the study, which then results in left truncation of the event-time. Patients who remain in the study are further subject to the usual right censoring, so the sample consists of left-truncated and right-censored (LTRC) survival times. It is well known that left truncation is a biased sampling plan as subjects with shorter survival times tend to be excluded from the sample. As a result, the longitudinal measurements are also sampled with bias.

An example of left-truncated and right-censored longitudinal study is the Italian multi-center HIV (human immunodeficiency virus) study [Rezza et al. (1989), The-Italian-Seroconversion-Study (1992)], where the primary endpoint is the time from HIV positive to AIDS onset, that is, the incubation period of AIDS. In this study, patients who have developed AIDS at the time of recruitment were excluded from the study, resulting in left truncation of the survival data, and CD4 counts for those who were HIV positive but ADIS free were measured at each follow-up visit. As there are no procedures available to handle such data properly, we develop in this paper a semiparametric joint likelihood approach to accommodate LTRC survival data with longitudinal covariates that are measured intermittently.

Although there is a sizable literature to jointly model right-censored survival and longitudinal data [see Wulfsohn and Tsiatis (1997), Henderson, Diggle and Dobson (2000), Song, Davidian and Tsiatis (2002) and the review papers by Tsiatis and Davidian (2004)], the extension to LTRC survival data turns out nontrivial due to the left-truncation feature of the data. To see this, consider first the simpler

case of left-truncated data with time-independent covariates or no covariates at all. Lynden-Bell (1971), Woodroofe (1985) and Wang (1987) investigated estimation of the survival function when subjects come from the same population, that is, there are no covariates involved. Here, one only needs to adjust the risk set for truncated data to reach a suitable extension of the Kaplan–Meier estimator. For time-independent covariates Andersen et al. (1993) considered estimation under the Cox model and showed that the partial likelihood approach for right-censored data still works for LTRC survival data when one conditions on the values of the covariates and truncation times.

For time-dependent covariate, the Andersen et. al. (1993) partial likelihood approach can still be employed if the entire covariate history is available for all subjects. This is not the case for longitudinal covariates that are observed intermittently at discrete time points. Since imputation methods lead to biases of the estimates, bias corrected approaches have been employed in the literature for right-censored data with longitudinal covariates. In particular, Wang (2006) proposed a method to correct the bias through the partial score equation. Such an approach is termed "corrected score" methods, which originates from studies of measurement errors. While corrected score methods typically lead to $\sqrt{n}$-consistent estimators for the regression parameters in the Cox model, they are not efficient and easy to derive. Extension of the corrected score methods to LTRC (left-truncated and right-censored) data might be feasible but have not been explored. In this paper, we adopt the full and joint likelihood approach of the survival and longitudinal data due to its aforementioned efficiency and robustness features. Unfortunately, direct maximization of the full joint likelihood is much more complicated than the cases with no left truncation. We discovered a modified likelihood that is simpler, yet retains the efficiency of the full likelihood approach, as described in Section 2.

The rest of the paper is organized as follows. In Section 2, we introduce a joint model setting for both the survival time and longitudinal processes and propose a modified likelihood approach for statistical inference. An EM algorithm to maximize the modified likelihood is derived in Section 3, along with the large sample properties of the nonparametric maximum modified likelihood estimator (NPMMLE), including consistency, asymptotic normality and efficiency. Numerical performance of the proposed estimating procedure is validated through simulation studies in Section 4 and illustrated through the Italian HIV study in Section 5. Section 6 contains some discussion.

2. Joint modeling under LTRC. We consider the setting that the survival time Y^* of a subject is subject to random left truncation by T^*, so a subject is enrolled in a study only if $Y^* \geq T^*$. Let n be the total number of subjects enrolled in the study. With such a biased sampling plan, to avoid confusion of notation, we denote the survival and truncation time of the ith enrolled subjects as (Y_i, T_i), which are sampled from the joint subpopulation of (Y_i^*, T_i^*), where $Y_i^* \geq T_i^*$.

Upon entering the study, these n subjects are subject to the usual right censor-ship, so the final observed survival data for the ith subject is a triplet (T_i, Z_i, Δ_i), where $Z_i = \min(Y_i, C_i)$ is the time of the endpoint event or drop-out (censoring) time C_i, whichever occurs first, and $\Delta_i = I(Y_i \leq C_i)$ is the censoring indicator.

In reality, drop-out or censoring only occurs when a subject is enrolled into the study. This fact implies that the right-censoring time C_i is greater than the truncation time T_i, for $i = 1, \ldots, n$. Therefore, we introduce a positive random variable U_i to represent the time from entry into the study to drop-out from the study, that is, $U_i = C_i - T_i$.

In addition to the survival data, baseline and longitudinal covariates are col-lected intermittently for the ith subject from the time the subject enters the study until the observational limit Z_i. This results in n_i repeated measurements, denoted by $\vec{W}_i = (W_{i1}, W_{i2}, \ldots, W_{in_i})$, where the measurements are taken at time points $\vec{s}_i = (s_{i1}, s_{i2}, \ldots, s_{in_i})$. It is important to make a note here that the observed $\vec{W}_i$ are also subject to the same biased sampling plan as the survival data, so there is a background longitudinal vector, which we will denote as $\vec{W}_i^*$ for the ith subject enrolled in the study. Therefore, $\vec{W}_i$ is sampled from the subpopulation of $\vec{W}^*$, where $Y_i^* \geq T_i^*$, and values beyond Z_i are not observed. For simplicity of nota-tion, we assume in this section that there is only one longitudinal covariates, but additional longitudinal or baseline covariates can be handled easily and the AIDS data discussed in Section 5 contain two longitudinal covariates, one observed inter-mittently but the complete history of the other one, the time-dependent treatment indicator, is available.

2.1. *The joint models.* Since repeated measurements from the same subjects are likely to be correlated, we introduce a latent $q \times 1$ random vector A_i^* to account for their dependency and assume a common parametric density function $f_A^*(\cdot|\alpha)$ with an unknown parameter α for A_i^*. A linear mixed effects model will be con-sidered for the longitudinal covariate

$$(2.1) \qquad \vec{W}_i^* = X(\vec{s}_i) + \varepsilon_i = g(\vec{s}_i)A_i^* + \varepsilon_i,$$

where $g(\cdot)$ is a known q-dimensional function and the $n_i \times 1$ vector ε_i plays the role of measurement errors, sampled from a multivariate normal distribution with independent marginal distribution $\mathcal{N}(0, \sigma^2)$, and independent of all other afore-mentioned random variables.

For the survival time Y_i^*, a proportional hazards model is employed, and the hazard rate of Y_i^* at time t given A_i^* is

$$(2.2) \qquad \lambda_{Y_i^*}(t|A_i^*) = \lambda_0(t) \exp(\beta X_i(t)),$$

where λ_0 is the baseline hazard rate and β is the regression coefficient. The trun-cation time T_i^* and the time U_i, from entry to drop-out, are assumed to have distri-bution function $F_{T^*}(\cdot)$ and $F_U(\cdot)$, respectively. We adopt the standard assumption

in survival analysis, that Y_i^*, T_i^* and U_i are conditionally independent given the covariates. This is equivalent to assuming conditional independence of Y_i^*, T_i^* and U_i given the value of A_i^*. We also assume that T_i^* and U_i are independent of A_i^*, and the parameters in the models for either the survival or longitudinal parts are noninformative.

2.2. *A modified likelihood approach.* For the model described in the previous subsection, the parameters of interest are (β, α, σ^2 and $\Lambda_0(\cdot)$), where the first three components are in the Euclidean space whereas $\Lambda_0(t) = \int_0^t \lambda_0(u)\, du$, the cumulative hazard function, is in a functional space, hence the model is semiparametric. Since a likelihood approach usually provides the most efficient estimating procedure, we first consider the full likelihood function L_i^O based on the observations $(t_i, z_i, \delta_i, \vec{w}_i)$ from the ith subject. The derivation of the full likelihood from the ith subject is shown below:

$$
L_i^O = f_{(T,Y,\Delta,W)}(t_i, z_i, \delta_i, w_i) = \frac{f_{(T^*,Y^*,\Delta^*,W^*)}(t_i, z_i, \delta_i, w_i)}{P(Y^* \geq T^*)}
$$

$$
= \left\{ \int [f_{Y^*}(z_i|A_i^* = a_i)]^{\delta_i}[S_{Y^*}(z_i|A_i^* = a_i)]^{1-\delta_i} \right.
$$

$$
\left. \times f_{W^*}(w_i|A_i^* = a_i) f_{A^*}(a_i)\, da_i \right\} f_{T^*}(t_i)/P(Y_i^* \geq T_i^*)
$$

(2.3)
$$
= \left\{ \int \frac{[f_{Y^*}(z_i|A_i^* = a_i)]^{\delta_i}[S_{Y^*}(z_i|A_i^* = a_i)]^{1-\delta_i}}{S_{Y^*}(t_i|A_i^* = a_i)} \right.
$$

$$
\left. \times f_{W^*}(w_i|A_i^* = a_i)\frac{S_{Y^*}(t_i|A^* = a_i) f_{A^*}(a_i)}{S_{Y^*}(t_i)}\, da_i \right\} \frac{S_{Y^*}(t_i) f_{T^*}(t_i)}{P(Y_i^* \geq T_i^*)}
$$

$$
= \left\{ \int [f_{Y^*}(z_i|Y_i^* \geq t_i, A_i^* = a_i)]^{\delta_i}[S_{Y^*}(z_i|Y_i^* \geq t_i, A_i^* = a_i)]^{1-\delta_i} \right.
$$

$$
\left. \times f_{W^*}(w_i|A_i^* = a_i) f_{A^*}(a_i|Y_i^* \geq t_i)\, da_i \right\}
$$

$$
\times f_{T^*}(t_i|Y_i^* \geq T_i^*),
$$

where f_V is the density function of the random variable V in the subscript, and S_V is the corresponding survival function. In (2.3), besides the baseline hazard function λ_0, the density function f_{T^*} also serves as a nonparametric component. Because of these two nonparametric components, the full likelihood function is unbounded, so we resort to the nonparametric maximum likelihood approach, which leads to a similar scenario as in conventional survival analysis that the full likelihood is the same as the conditional likelihood given the left-truncation time. This has been explored in the literature [Andersen et al. (1993), Klein and Moeschberger (2003)] for LTRC data with baseline covariates, and was first explored in Wang (1987) for the simpler situation of left-truncated data that came from a single population. Following a similar argument as in Wang (1987), we

found that the full likelihood can be simplified to the following conditional likelihood for the ith subject as

$$
L_i^C = \left\{ \int [f_{Y^*}(z_i|Y_i^* \geq t_i, A_i^* = a_i)]^{\delta_i} [S_{Y^*}(z_i|Y_i^* \geq t_i, A_i^* = a_i)]^{1-\delta_i} \right.
$$

(2.4)

$$
\left. \times f_{W^*}(w_i|A_i^* = a_i) f_{A^*}(a_i|Y_i^* \geq t_i) \, da_i \right\}.
$$

Next, we consider the nonparametric maximum likelihood estimators (NPMLE) of the survival component, which, by a similar argument for joint modeling right-censored data and their longitudinal covariates [Zeng and Cai (2005), Dupuy, Grama and Mesbah (2006)], leads to a piecewise linear baseline cumulative hazard function with jumps at each uncensored event time (i.e., at Y_i, whenever $\Delta_i = 1$). Let n_u denote the total number of uncensored events, the baseline cumulative hazard function is thus re-parameterized as a n_u-dimensional vector.

So far, the derivation of the likelihood function and NPMLE follows a similar path as the much investigated case of a joint modeling setting with right-censored data, where NPMLE's for the parametric component enjoy nice asymptotic properties and are semiparametrically efficient. Despite these similarities, the left truncation feature triggers complications in the estimation of the finite dimensional parameter in the joint LTRC model. First, as shown in the Appendix, the parameter α associated with the latent variable A^* is not identifiable. This is a consequence of the biased sampling plan, since the samples are actually drawn from the subpopulation $Y^* \geq T^*$. Consequently, only $E(A^*|Y^* \geq T^*)$ and $\mathrm{var}(A^*|Y^* \geq T^*)$ could be identified under the normality assumption. Thus, while it is possible to identify the unknown parameters of Y^* and T^* based on the joint conditional distribution of $(Y^*, T^*)|Y^* \geq T^*$, where the notation $(\cdot|Y^* \geq T^*)$ stands for a random variable/vector sampled from the subpopulation with $Y^* \geq T^*$, there is not enough information to recover $E(A^*)$ and $\mathrm{var}(A^*)$ and hence the true longitudinal parameters α.

A second complication is that the score equations for the survival components, β and Λ_0, are much more complicated than the situation under a right-censored only model and, as shown in Appendix A.1, as they require estimation of the expectations of nonlinear functions of the observed data along with the the parameters of interest. This motivates us to modify the likelihood so as to simplify the estimation of all parameters that are identifiable. Our proposal is to aim at the following *modified likelihood*, denoted by L^m, as an alternative of the full, also the conditional, likelihood in (2.4). The modified likelihood is

$$
L^m = \prod_{i=1}^{n} \left\{ \int [f_{Y^*}(z_i|Y_i^* \geq t_i, A_i^* = a_i)]^{\delta_i} \right.
$$

(2.5)

$$
\times [S_{Y^*}(z_i|Y_i^* \geq t_i, A_i^* = a_i)]^{1-\delta_i}
$$

$$
\left. \times f_{W^*}(w_i|A_i^* = a_i) f_{A^*}(a_i) \, da_i \right\} f_{T^*}(t_i|Y_i^* \geq T_i^*),
$$

where the lower case variables denote the values of the corresponding upper case variables, for example, δ_i is the value of Δ_i. The estimators obtained by maximizing the modified likelihood, where the nonparametric cumulative hard function is replaced by a step function will be referred to as the *nonparametric maximum modified likelihood* (NPMMLE) hereafter.

The difference between (2.4) and (2.5) is that $f_{A^*}(a_i|Y_i^* \geq t_i)$ in the full likelihood (2.4) is replaced by $f_{A^*}(a_i)$ in (2.5). This is motivated by the fact that $f_{A^*}(a|Y^* \geq t) = \frac{S_{Y^*}(t|A^*=a)}{S_{Y^*}(t)} f_{A^*}(a)$ and $E[\frac{S_{Y^*}(t|A^*)}{S_{Y^*}(t)}] = 1$, for any t, and that, as shown in Lemma A.1 in the Appendix, the score functions of the survival parameters from (2.5) are asymptotically the same as those from (2.4). Theoretical results in the next section and numerical evidence in Section 4 demonstrate good performance of estimators of all the survival parameters, $(\beta, \Lambda_0(\cdot))$ and of the measurement errors σ^2 of the longitudinal component that we derived from this modified likelihood.

3. EM-algorithm and asymptotic properties. Let $\gamma = (\beta, \alpha, \sigma^2)$ be the finite dimensional parameter in the joint survival and longitudinal model, and Λ be a step function. The log modified likelihood is

$$l^m(\gamma, \Lambda) = \sum_{i=1}^{n} \ln \int [\Lambda\{z_i\} \exp \beta g(z_i) a_i]^{\delta_i}$$
$$\times \exp\left\{-\sum_{j:t_i < y_j^0 \leq z_i} \Lambda\{y_j^0\} \exp\{\beta g(y_j^0) a_i\}\right\}$$
$$\times (2\pi\sigma^2)^{-n_i/2}$$
$$\times \exp\left\{-\sum_{j=1}^{m_i} [w_{ij} - g(s_{ij}) a_i]^2/(2\sigma^2)\right\} f_{A^*}(a_i) \, da_i,$$

where $\Lambda\{\cdot\}$ is the jump size of Λ at the respective time point in the argument, and y_j^0 is the jth sorted observed survival time in increasing order. Moreover, τ_1 and τ_2 denote the lower bound of truncation time and the largest censoring time corresponding to the end of the study.

Since direct maximizing the proposed modified likelihood involves integration of a complex function with respect to the random effects, we employ the expectation-maximization (EM) algorithm [Laird and Ware (1982)] to stabilize the maximization procedure. In the implementation of the EM algorithm, a Monte Carlo integration approach is used to approximate the expectation terms of functions $h(A^*)$ appearing in the E-step. A one-step Newton–Raphson method is applied to solve the nonlinear equations in the M-step. The posterior density of the random effects A_i^* given the observed data from the ith subject, $o_i = (t_i, z_i, \delta_i, \vec{w}_i)$,

is of the form

$$
\begin{aligned}
f_{A^*|O}(a|o_i) &= \frac{f_{(Y,\Delta)|(A,T)}(z_i, \delta_i|a, t_i) \times f_{A^*|W^*}(a|\vec{w}_i)}{\int f_{(Y,\Delta)|(A,T)}(z_i, \delta_i|a, t_i) \times f_{A^*|W^*}(a|\vec{w}_i)\, da} \\
&= [\Lambda\{z_i\}]^{\delta_i} \exp\left\{ -\sum_{j:t_i < y_j^0 \le z_i} \Lambda\{y_j^0\} \exp\{\beta g(y_j^0)a\} \right\} \times f_{A^*|W^*}(a|\vec{w}_i) \\
&\quad \Big/ \int [\Lambda\{z_i\}]^{\delta_i} \exp\left\{ -\sum_{j:t_i < y_j^0 \le z_i} \Lambda\{y_j^0\} \exp\{\beta g(y_j^0)a\} \right\} \\
&\quad \times f_{A^*|W^*}(a|\vec{w}_i)\, da.
\end{aligned}
$$

For a simpler implementation of the algorithm, we shall impose a normal assumption on the random effects and assume that A_i^*, $i = 1, \ldots, n$, follow a normal distribution $N(\mu, \Sigma)$, where (μ, Σ) plays the role of the parameter α.

By taking the first derivative of the log modified likelihood calculated in the E-step with respect to each parameter, the NPMMLE, $\hat{\beta}$, $\{\hat{\lambda}_k, k = 1, \ldots, n_u\}$, $\hat{\sigma}$, $\hat{\mu}$ and $\hat{\Sigma}$, can be obtained through the following formulas, where λ_k is the jump size of Λ at the kth sorted observed survival time:

$$
\hat{\lambda}_k = \frac{1}{\sum_{i:t_i < y_k^0 \le z_i} E[\exp\{\beta g(y_k^0)A_i^*\}|o_i]}, \qquad k = 1, \ldots, n_u,
$$

$$
\hat{\sigma}^2 = \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} E[(w_{ij} - g(s_{ij})A_i^*)^2|o_i],
$$

$$
\hat{\mu} = \frac{1}{n} \sum_{i=1}^n E(A_i^*|o_i), \qquad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n E[(A_i^* - \hat{\mu})(A_i^* - \hat{\mu})^T|o_i]
$$

and $\hat{\beta}$ is the root of the score $s(\beta)$, which is solved by an one-step Newton–Raphson method with the updating rule

$$
\beta_{\text{new}} = \beta_{\text{old}} - \frac{s(\beta_{\text{old}})}{s'(\beta_{\text{old}})},
$$

where

$$
s(\beta) = \sum_{i=1}^n \delta_i \left[g(z_i) E(A_i^*|o_i) - \frac{\sum_{j:t_j < z_i \le z_j} E(g(z_i)A_j^* \exp\{\beta g(z_i)A_j^*\}|o_j)}{\sum_{j:t_j < z_i \le z_j} E(\exp\{\beta g(z_i)A_j^*\}|o_j)} \right],
$$

$$
s'(\beta) = \sum_{i=1}^n \delta_i \left\{ \left[\frac{\sum_{j:t_j < z_i \le z_j} E(g(z_i)A_j^* \exp\{\beta g(z_i)A_j^*\}|o_j)}{\sum_{j:t_j < z_i \le z_j} E(\exp\{\beta g(z_i)A_j^*\}|o_j)} \right]^2 \right.
$$

$$
\left. - \frac{\sum_{j:t_j < z_i \le z_j} E((g(z_i)A_j^*)^2 \exp\{\beta g(z_i)A_j^*\}|o_j)}{\sum_{j:t_j < z_i \le z_j} E(\exp\{\beta g(z_i)A_j^*\}|o_j)} \right\}.
$$

Except for α, the proposed nonparametric maximum modified likelihood estimates (NPMMLE) of the parameters enjoy nice properties that are similar to the NPMLE, as illustrated in the next two theorems. Below, we list some regularity conditions needed for the theorems:

(C1) The parameter space of the finite dimensional parameters, S_γ, is bounded and closed on Euclidean space. The true value γ_0 is an interior point of S_γ.

(C2) On the parameter space of β, $(\exp\{\beta g(S)A^*\}|Y^* \geq T^*)$ is bounded below by m and above by M with probability 1.

(C3) $P(T \leq \tau_1 \text{ and } Y \geq \tau_2) > 0$. This ensures that not all data are truncated or censored.

(C4) $E_{\theta_0}\{\exp[\beta_0 g(u)A^*]I(T^* < u \leq Y^*)|Y^* \geq T^*\}$ is bounded away from 0 on the parameter space of β. Here $E_{\theta_0}(\cdot)$ stands for the expectation taken under the true value of the parameter θ_0.

(C5) $g(t)$ is of uniformly bounded variation on $[\tau_1, \tau_2]$, and there exists a constant D such that $P(n_i \leq D) = 1, \forall i$.

(C6) The distribution $f_{A^*}(\cdot|\alpha)$ is continuous with respect to α and has continuous second derivative with respect to α. Moreover, the Fisher information matrix obtained from f_{A^*} for α is positive definite.

THEOREM 1 (Consistency of the estimators).   *Under the regularity conditions* C1–C5, *the NPMMLE of* $(\beta_0, \sigma_0^2, \Lambda_0)$, *denoted as* $(\hat{\beta}_n, \hat{\sigma}^2, \hat{\Lambda}_n)$, *is consistent under the Euclidean norm* $|\cdot|$ *and supremum norm* $\|\cdot\|_\infty$ *on* $[\tau_1, \tau_2]$, *respectively.*

For $H = \{h = (h_1, h_2, h_3)\}$ and $0 < p < \infty$, let $H_p = \{h \in H : \|h_1\|, |h_2|, \|h_3\|_v \leq p\}$, be a collection of directions that are used in the Appendix. The notation $\|\cdot\|_v$ denotes the the total variation of the function in the norm plus the absolute value of this function evaluated at 0. The next theorem shows that the NPMMLE converges in distribution to a Gaussian element in the parameter space at a $\sqrt{n}$-rate.

THEOREM 2 (Asymptotic normality and efficiency).   *Under the regularity conditions* C1–C6, *the process* $\sqrt{n}(\hat{\alpha}_n - E(\hat{\alpha}_n), \hat{\sigma}_n^2 - \sigma_0^2, \hat{\beta}_n - \beta_0, \hat{\Lambda}_n - \Lambda_0)$ *converges in distribution to a mean zero Gaussian process* G *in the functional space* $l_\infty(H_p)$ *on* H_p. *Moreover, the NPMMLE* $\hat{\beta}$ *is semiparametrically efficient for* β_0.

Proofs of these two theorems are provided in the Appendix.

For estimating the standard errors of the NPMMLE, we recommend to use the bootstrap procedure instead of the profile likelihood approach in Murphy and van der Vaart (2000) and Zeng and Cai (2005), which did not work well for LTRC data due to the high fluctuation of the estimated profile likelihood function and possibly negative estimate of the standard error. The performance of the bootstrap procedure for estimating the standard errors of the NPMLE under joint modeling

with right-censoring cases has been studied by Tseng, Hsieh and Wang (2005) for the accelerated failure time model, and by Hsieh, Tseng and Wang (2006) for the Cox model. The results in these two papers and support the validity of the bootstrap method in the scope of joint modeling. Our simulation results reported in Section 4 also supports the use of the bootstrap approach. In comparison, the bootstrap method is more reliable than the profile likelihood method at a higher computational cost.

4. Simulation study. To verify numerically the validity of the proposed procedure, we conducted simulations under five different settings. Since there is an intrinsic bias on the longitudinal component, the simulations focus on the performance of the estimate of β and how it would be affected by the level of contamination from the measurement errors and the variation of the random effects. As a benchmark setting, we considered a linear trend in time with random effects on the longitudinal covariate and assess the influence of the variance of the random slope on the accuracy of estimating β. The left-truncation times are generated from an exponential distribution with parameter 1, while the right-censoring times are from an exponential distribution with parameter 3. The baseline hazard rate is from an exponential distribution with mean 1. All 5 simulation settings have sample size $n = 200$ with true values $\beta = 1$, $\mu = (2, 0.5)$ and $(\sigma_{11}, \sigma_{12}) = (0.5, -0.001)$. The values of (σ_{22}, σ^2) are different for the five settings and set as: $(0.01, 0.1)$, $(0.01, 0.4)$, $(0.01, 0.025)$, $(0.0025, 0.1)$ and $(0.04, 0.1)$. The first three settings demonstrate the impact of contaminations by measurement errors while the last two illustrate the effect of the variation of the random slope.

Simulation results based on 100 Monte Carlo samples are reported in Table 1. Results under the first three settings suggest that β can be estimated unbiasedly, and measurement errors affect the precision, but not the magnitude of the biases. As expected, higher level of noise contamination leads to less precise estimate of β and higher chance of divergence in the algorithm. In all three settings, the variance of measurement errors can be estimated with high accuracy and precision. Comparing with the results under the first, fourth and fifth setting from Table 1, we observe that the variance of the random slopes has little effect on the performance of $\hat{\beta}$.

The results for the longitudinal part echo the above discussion of the non-identifiability of the parameter α, as the means of the random intercept and random slopes (shown in the second column of Table 1) are consistently underestimated. The actual targets of the estimates are the conditional quantities marked as μ_1^* and μ_2^* etc. in the first column of Table 1. The sizes of the biases vary with the level of truncation probability and size of measurement errors and can be very small for the mean of the random slope, for example, in setting 3, where the measurement error is small. Thus, this bias problem in estimating the longitudinal component may elude researchers, while it is a cause of substantial concern in settings with large error variances.

TABLE 1
Simulation results under five settings with sample size 200 *and varying values of* σ_{22} *and* σ^2. *The actual targets of the longitudinal estimates are conditional quantities marked as* μ_1^* *and* μ_2^* *etc. and are listed next to the true longitudinal value in the first column. The mean and SD of the estimates based on* 100 *Monte Carlo samples are reported in the second and third column*

Case	Parameter	Average of NPMMLE	SE(MC)	MSE	Convergence rate
1	β (1)	0.9923	0.1633	0.0267	98%
	σ^2 (0.1)	0.0998	0.0021	5e−6	
	μ_1/μ_1^* (2/1.73)	1.7461	0.0478	0.0668	
	μ_2/μ_2^* (0.50/0.50)	0.4545	0.0985	0.0118	
	$\sigma_{11}/\sigma_{11}^*$ (0.50/0.45)	0.4634	0.0527	0.0041	
	$\sigma_{12}/\sigma_{12}^*$ (−0.001/−0.001)	−0.0424	0.0453	0.0038	
	$\sigma_{22}/\sigma_{22}^*$ (0.01/0.01)	0.0738	0.0409	0.0057	
2	β (1)	0.9185	0.1765	0.0378	72%
	σ^2 (0.4)	0.4003	0.0086	7e−5	
	μ_1/μ_1^* (2/1.74)	1.7455	0.0531	0.0676	
	μ_2/μ_2^* (0.50/0.50)	0.3801	0.1640	0.413	
	$\sigma_{11}/\sigma_{11}^*$ (0.5/0.45)	0.4730	0.0505	0.0033	
	$\sigma_{12}/\sigma_{12}^*$ (−0.001/−0.001)	−0.1122	0.0917	0.0208	
	$\sigma_{22}/\sigma_{22}^*$ (0.01/0.01)	0.1856	0.1156	0.0442	
3	β (1)	1.0380	0.1548	0.0254	96%
	σ^2 (0.025)	0.0250	4.8283e-4	$\simeq 0$	
	μ_1/μ_1^* (2/1.73)	1.7443	0.0468	0.0676	
	μ_2/μ_2^* (0.50/0.50)	0.4900	0.0643	0.0042	
	$\sigma_{11}/\sigma_{11}^*$ (0.50/0.45)	0.4520	0.0534	0.0052	
	$\sigma_{12}/\sigma_{12}^*$ (−0.0004/−0.0004)	−0.0193	0.0338	0.0015	
	$\sigma_{22}/\sigma_{22}^*$ (0.01/0.01)	0.0571	0.0219	0.0027	
4	β (1)	0.9684	0.1504	0.0236	98%
	σ^2 (0.1)	0.0997	0.0023	5e−6	
	μ_1/μ_1^* (2/1.74)	1.7464	0.0460	0.0664	
	μ_2/μ_2^* (0.50/0.50)	0.4491	0.0948	0.0116	
	$\sigma_{11}/\sigma_{11}^*$ (0.5/0.45)	0.4497	0.0423	0.0043	
	$\sigma_{12}/\sigma_{12}^*$ (−0.001/−0.0007)	−0.0439	0.0518	0.0045	
	$\sigma_{22}/\sigma_{22}^*$ (0.0025/0.0025)	0.0797	0.0479	0.0072	
5	β (1)	0.9934	0.1567	0.0246	95%
	σ^2 (0.1)	0.0996	0.0020	4e−6	
	μ_1/μ_1^* (2/1.74)	1.7522	0.0464	0.0636	
	μ_2/μ_2^* (0.50/0.50)	0.4498	0.1168	0.0162	
	$\sigma_{11}/\sigma_{11}^*$ (0.50/0.45)	0.4559	0.0442	0.0039	
	$\sigma_{12}/\sigma_{12}^*$ (−0.001/−0.002)	−0.0433	0.0692	0.0066	
	$\sigma_{22}/\sigma_{22}^*$ (0.04/0.04)	0.1186	0.0642	0.0159	

TABLE 2
Performance of estimated variance, SE(BT), of $\hat{\beta}$ and $\hat{\sigma}^2$ through bootstrap with 50 resamples

Case	Parameter	SE(MC)	SE(BT)
1	β (1)	0.1633	0.1692
	σ^2 (0.1)	0.0021	0.0020
2	β (1)	0.1765	0.1813
	σ^2 (0.4)	0.0086	0.0091
3	β (1)	0.1548	0.1523
	σ^2 (0.025)	4.8283e−4	5e−4
4	β (1)	0.1504	0.1539
	σ^2 (0.1)	0.0023	0.0020
5	β (1)	0.1567	0.1531
	σ^2 (0.1)	0.0020	0.0021

To make statistical inference about the parameters of interest, it is necessary to get an estimate of the standard error of the NPMMLE, especially for β. We tried the approach in Murphy and van der Vaart (2000) and Louis (1982), but neither works, so we propose to use a bootstrap method [Tseng, Hsieh and Wang (2005)] for estimating the standard error of the NPMMLE and present the results in Table 2. Only the results for estimating the standard errors of $\hat{\beta}$ and $\hat{\sigma}^2$ are shown, since they are estimable. Table 2 supports the use of the bootstrap procedure, as the estimated standard error from the bootstrap method is close to the standard deviation from the 100 Monte Carlo samples, even when the degree of error contamination is large or the random slopes vary widely.

5. Data example: Multi-center HIV study. In this section, we conduct an analysis on the data from a multi-center HIV study in Italy. Details of the study design and a previous analysis can be found in Rezza et al. (1989) and The-Italian-Seroconversion-Study (1992). There were 448 HIV-positive patients in the data. The primary event of interest is the incubation period of acquired immunodeficiency syndrome (AIDS), that is, time (in years) from detection of HIV-infection until the onset of AIDS. There were 140 patients who received the HAART treatment at various times, resulting in a longitudinal treatment indicator that is fully observable, so no modeling of this process is necessary. However, there is a second longitudinal covariate, the CD4 counts, that are observed only intermittently at follow-up visits, motivating the need to model the survival and longitudinal covariates jointly. The main biomedical interest lies in determining the effect of the HAART treatment on reducing the risk of developing AIDS, and the association between the incubation period of AIDS and CD4 T-cell counts in HIV-infected subjects.

For each of the 448 subjects in the study, the longitudinal measurements of CD4 T-cell counts were recorded intermittently along with the time to AIDS or dropout from the study. The total number of longitudinal measurements is 4442 and the average number of longitudinal measurements is 9.92 per patient.

One feature of this data is that the incubation period is subject to left-truncation and right-censoring, since patients were recruited to the study at various times after the study began, and only patients who have not developed AIDS at the time of recruitment are included in the study. Moreover, only 147 out of the 448 patients (about 33%) developed AIDS by the end of the study, so the right censoring rate is quite high for this data.

To model the longitudinal CD4 counts, we adopt a linear mixed effects model on $\log(\text{CD4}+1)$ with changing intercepts and slopes at the time of HAART treatment. Thus,

$$W_i^*(s_{ij}) = X_i(s_{ij}) + \varepsilon_{ij} = A_{i0}^* + A_{i1}^* s_{ij} + A_{i2}^* I(s_{ij} > V_i) + A_{i3}^* s_{ij} I(s_{ij} > V_i) + \varepsilon_{ij},$$

where ε_{ij} is from a normal distribution $N(0, \sigma^2)$, $\vec{A}_i^* = (A_{i0}^*, A_{i1}^*, A_{i2}^*, A_{i3}^*)$ is from a 4-dimensional multivariate normal distribution with a 4×1 mean vector μ and a 4×4 covariance matrix Σ, and V_i represents relative age since HIV-positive of receiving HAART. For those who have never received HAART, V_i is defined to be infinity. For the time-to-AIDS, we assume a Cox model with $X_i(t)$, CD4 counts, as an time-dependent covariate along with another time-dependent treatment indicator, $I(t > V_i)$, which is completely observed. The resulting model is

$$\lambda(t|\vec{A}_i^*) = \lambda_0(t) \exp(\beta_1 X_i(t) + \beta_2 I(t > V_i)).$$

From the EM algorithm with Monte Carlo approximation, the slope, $\hat{\beta}_1$, for the underlying $\log(\text{CD4}+1)$ process is estimated to be -0.5762 (p-value < 0.001), while the slope, $\hat{\beta}_2$, for the longitudinal treatment indicator is estimated to be -1.2189 (p-value < 0.001). As expected, CD4 counts are negatively associated with the risk of AIDS. One unit of decline on $\log(\text{CD4}+1)$ is associated with an increasing risk of AIDS by 78%. In addition to its effect on CD4 counts, HAART has an additional effect on reducing the risk of AIDS. It significantly reduces the risk of developing AIDS by 70% after controlling for the CD4 counts. Through the analysis, we confirm that HAART effectively reduces the risk of developing AIDS both through a positive association with patients' CD4 counts and the risk to develop AIDS.

6. Conclusions and discussion. We have shown, both theoretically and empirically, that joint modeling the time-to-event and longitudinal covariates is an effective modeling approach when the time-to-event is subject to both left truncation and right censoring. However, the extension from right-censorship to LTRC is not trivial. By modifying the joint likelihood, we have shown that NPMMLE

leads to consistent and asymptotically efficient estimation of the survival component and measurement error variance under the setting of a semiparametric Cox model. We have also demonstrated that the corresponding EM algorithm to locate the NPMMLE has good empirical performance and asymptotic properties under the assumption of normal random effects. It is not only computational effective but also robust against departures from the normality assumption.

However, one caveat is the estimability of the longitudinal component. Although we can recover the conditional distribution of the longitudinal parameter, α, given $Y \geq T$, the parameter α itself can not be estimated properly though the modified likelihood due to the biased sampling plan. Additional strong and possibly unverifiable assumptions might be needed in order to recover the parameter α of the random effects. What we have accomplished in this paper is to successfully remove the bias for the estimation of the survival components attributed to the discrete measurement schedule and measurement errors of the longitudinal covariates, thus permitting asymptotically valid and efficient inference for the survival related parameters, which are crucial for the evaluation of therapies.

APPENDIX

A.1. Likelihood and the score equations. By imposing a normality assumption $N(\mu, \Sigma)$ on the random effects A_i^*, the full likelihood in (2.3) from the ith subject becomes

$$L_i^O \propto f_{T^*}(t_i)\sigma^{-n_i}\lambda_0(z_i)^{\delta_i}$$

$$\times \int_{-\infty}^{\infty} \exp\left\{\delta_i \beta g(z_i)a_i - \sum_{j=1}^{n_i}[w_{ij} - g(s_{ij})a_i]^2/(2\sigma^2)\right\} Q_1(z_i, a_i)\,da_i$$

$$\Big/ \int_0^{\infty} \int_{-\infty}^{\infty} Q_1(t, a_i) f_T(t)\,da_i\,dt,$$

where $Q_1(u, a) = \exp\{-\int_0^u \exp[\beta g(t)a]\,d\Lambda_0(t) - (a - \mu)^T \Sigma^{-1}(a - \mu)/2\}$. Following similar arguments as in Wang (1987) and combining with Vardi (1985), we can prove that the NPMLE's of all finite-dimensional parameters are the same as those from the conditional likelihood of $(z_i, \delta_i, \vec{w}_i)$ given $(Y_i^* > t_i)$. Moreover, by a proof similar to that of the classical Cox model for right censored data, the NPMLE from the conditional likelihood is attained by discrete baseline hazard functions that assign positive masses only at uncensored survival times, $(y_1^0, \ldots, y_{n_u}^0)$.

Let $o_i = (t_i, z_i, \delta_i, \vec{w}_i)$ denote the observed data for the ith subject. The first derivative of the log full likelihood leads to the following score functions:

$$s_{\sigma^2}^o = \sum_{i=1}^n \left\{\sum_{j=1}^{n_i} E[w_{ij} - A_i^* g(s_{ij})|o_i]^2 - n_i\sigma^2\right\}\Big/\sigma^{-3},$$

$$s_\mu^o = \Sigma^{-1} \sum_{i=1}^{n} E\{(A_i^* - \mu) - [E(A_i^* | Y_i^* \geq T_i^*) - \mu] | o_i\}$$

$$= \Sigma^{-1} \sum_{i=1}^{n} [E(A_i^* | o_i) - E(A_i^* | Y_i^* \geq T_i^*)],$$

$$s_\Sigma^o = \frac{1}{2} \Sigma^{-1} \sum_{i=1}^{n} \{E[(A_i^* - \mu)(A_i^* - \mu)^T | o_i]$$

$$- E[(A_i^* - \mu)(A_i^* - \mu)^T | Y_i^* \geq T_i^*]\} \Sigma^{-1},$$

$$s_{\Lambda_k}^o = \frac{1}{\Lambda_k} - \sum_{i:t_i < y_k^0 \leq z_i} E\{\exp[\beta g(y_k^0) A^*] | o_i\} - Q_2(y_k^0),$$

$$s_\beta^o = \sum_{i=1}^{n} \delta_i g(y_i) E(A_i^* | o_i)$$

$$- \sum_{i=1}^{n} \sum_{j:t_i < y_j^0 \leq z_i} \Lambda_j E\{g(y_j^0) A_i^* \exp[\beta g(y_j^0) A_i^*] | o_i\} - Q_3,$$

where

$$Q_2(y) = \sum_{i:y \leq t_i} E\{\exp[\beta g(y) A_i^*] | o_i\}$$

$$- n E\{\exp[\beta g(y) A_i^*] I(y \leq T_i) | Y_i^* \geq T_i^*\},$$

$$Q_3 = \sum_{i=1}^{n} \sum_{j:y_j^0 \leq t_i} \Lambda_j E\{g(y_j^0) A_i^* \exp[\beta g(y_j^0) A_i^*] | o_i\}$$

$$- n E\left\{ \sum_{j:y_j^0 \leq t_i} \Lambda_j g(y_j^0) A_i^* \exp[\beta g(y_j^0) A_i^*] \middle| Y_i^* \geq T_i^* \right\}.$$

The score equations, s_μ^o and s_Σ^o, corresponding to the longitudinal data reveal that the estimable terms are the conditional mean and covariance matrix of the random effects given that $Y^* \geq T$ rather than μ and Σ.

The score functions for λ_k, $k = 1, \ldots, n_u$, and β have more complicated forms than those from a partial likelihood under standard Cox model subject to LTRC. The complication is due to the additional terms Q_2 and Q_3, which require estimation of the expectations of nonlinear functions of the observed data along with the the parameters of interest. If we drop these two terms from $s_{\Lambda_k}^o$ and s_β^o, the modified score functions, $s_{\Lambda_k} = s_{\Lambda_k}^o + Q_2(y_k^0)$ and $s_\beta = s_\beta^o + Q_3$, are exactly the score functions from the modified likelihood. The next Lemma validates the use of the modified likelihood (2.5).

LEMMA A.1. (i) $E_{\theta_0}(s_{\Lambda_k}) = E_{\theta_0}(s_{\Lambda_k}^o)$ and $E_{\theta_0}(s_\beta) = E_{\theta_0}(s_\beta^o)$. This provides Fisher consistency of the estimators (2.5).

(ii) Under the regularity conditions for law of large numbers and Slutsky theorem, $n^{-1}(s_{\Lambda_k} - s_{\Lambda_k}^o) = o_p(1)$ and $n^{-1}(s_\beta - s_\beta^o) = o_p(1)$.

PROOF. The proof follows from simple derivation and applications of the law of large numbers along with Slutsky's theorem. □

This lemma demonstrates the asymptotic equivalence of the score functions for the survival-related parameters from (2.3) and (2.5). The latter is computationally simpler to maximize and thus more attractive than the full likelihood.

A.2. Proof of the consistency of the NPMMLE. The proof of consistency includes four major steps and is elaborated below.

Step 1. *Existence of the NPMMLE of* (γ, Λ). We will begin the proof that the candidates for the maximizer, Λ_{n_u}, have a finite and bounded jump at each observed survival time. For simplicity, we use a vector form $\vec{\lambda}_{n_u} = (\lambda_1, \ldots, \lambda_{n_u})$ to express the jump sizes of Λ_{n_u} at ordered survival times. The boundedness of the jump sizes can be demonstrated by proving the existence of an upper bound $B \in \mathbb{R}$ through apagoge. Suppose that for any arbitrary $B \in \mathbb{R}$, there exists $\vec{\lambda}_{n_u, B} = (\lambda_{1,B}, \ldots, \lambda_{n_u, B}) \in \mathbb{R}^{n_u} \backslash [0, B]^{n_u}$ and $\gamma_B \in S_\gamma$ such that $L^m(\gamma_B, \vec{\lambda}_{n_u, B}) > L^m(\gamma, \vec{\lambda}_{n_u})$ for all $(\gamma, \vec{\lambda}_{n_u}) \in S_\gamma \times [0, B]^{n_u}$. The first part in $L^m(\gamma_B, \vec{\lambda}_{n_u, B})$ contributed by the ith subject is bounded above by

$$(\Lambda_{n_u, B}\{z_i\}M)^{\delta_i} \times \exp\left\{-m \sum_{j:t_i < y_j^0 \le z_i} \lambda_{j,B}\right\},$$

where m, M is defined in assumption C2. Since $\vec{\lambda}_{n_u, B} \in \mathbb{R}^{n_u} \backslash [0, B]^{n_u}$, at least one jump size, say $\lambda_{i_0, B}$, is greater than B. It induces that $\sum_{j:t_{i_0} < y_j^0 \le z_{i_0}} \lambda_{j,B} > B$, and then implies that $L^m(\gamma_B, \vec{\lambda}_{n_u, B}) \to 0$ as $B \to \infty$. Thus $L^m(\gamma, \lambda_{n_u}) = 0$, for all $(\gamma, \lambda_{n_u}) \in S_\gamma \times \mathbb{R}^{n_u}$, which is a contradiction. It demonstrates the boundedness of the jump sizes of Λ_{n_u}. Along with the compactness of S_γ provided by assumption C1, we accomplished the existence of the NPMMLE of (γ, Λ).

Step 2. *Almost surely boundedness of* $\hat{\Lambda}(\tau_2)$ *as* $n \to \infty$. For any fixed sample size n, the estimated cumulative hazard function evaluated at the endpoint of the study can be expressed as

$$\hat{\Lambda}(\tau_2) = \sum_{k=1}^n \frac{\delta_k I(z_k \le \tau_2)}{\sum_{i=1}^n E_{\hat{\theta}}[\exp\{\hat{\beta}g(z_k)A_i^*\}|o_i]I(t_i < z_k \le z_i)}$$

(A.1)

$$\le \sum_{k=1}^n \frac{\delta_k I(z_k \le \tau_2)}{m \sum_{i=1}^n I(t_i < z_k \le z_i)} \le \frac{\sum_{k=1}^n \delta_k I(z_k \le \tau_2)}{m \sum_{i=1}^n I(t_i \le \tau_1)I(\tau_2 \le z_i)},$$

where m is the lower bound of $\exp\{\beta g(Y) A^*\}|Y^* \geq T^*$, which exists under assumption C2. By the law of large numbers and the continuous mapping theorem, we have the following two limits as $n \to \infty$:

$$\frac{1}{n} \sum_{k=1}^{n} \delta_k I(z_k \leq \tau_2) \to E\big(\Delta I(Y \leq \tau_2)\big) < 1 \quad \text{and}$$

(A.2)

$$\frac{1}{(1/n) \sum_{i=1}^{n} I(t_i \leq \tau_1 \text{ and } \tau_2 \leq z_i)} \to \frac{1}{P(T \leq \tau_1 \text{ and } Y \geq \tau_2)} < \infty,$$

where the finiteness of the second limit is following assumption C3. Therefore, there exists an upper bound of $\hat{\Lambda}(\tau_2)$ even when n goes to infinity. Moreover, since the terms inside the summation in (A.1) are all strictly positive, $\hat{\Lambda}(\tau_2)$ is always greater than 0. Thus $\hat{\Lambda}(\tau_2)$ has been shown to be bounded almost surely as $n \to \infty$.

   Step 3. *Uniform convergence of* $(\hat{\sigma}_n^2, \hat{\beta}_n, \hat{\Lambda}_n)$ *to* $(\sigma_0^2, \beta_0, \Lambda_0)$. We have shown in Step 2 that $\hat{\Lambda}(\tau_2)$ is finite, combining with the fact that $\hat{\Lambda}$ is a right-continuous and nondecreasing step function along with the Helly selection theorem, there exists a subsequence of $\hat{\Lambda}$ converging pointwisely to a right continuous and monotone function Λ^* with probability 1. Moreover, by the Balzonno–Weierstrass theorem, there is a sub-subsequence of $\hat{\gamma}$ which converges to some γ^*. Therefore, there exists a sub-subsequence of $\hat{\theta}_n$, denoted by $\hat{\theta}_{\eta(n)}$, that converges to $\theta^* = (\gamma^*, \Lambda^*)$. We next show that $\theta^* = (\alpha_0^*, \sigma_0^2, \beta_0, \Lambda_0)$, where α_0^* is the limit of $\hat{\alpha}$. Here a new term, defined as

$$\bar{\Lambda}_n(t) = \frac{1}{n} \sum_{k=1}^{n} \frac{\delta_k I(z_k \leq t)}{(1/n) \sum_{i=1}^{n} E_{\theta_0}[\exp\{\beta_0 g(z_k) A_i^*\}|o_i] I(t_i < z_k \leq z_i)},$$

is introduced to serve as a bridge between $\hat{\Lambda}_n$ and Λ_0.

   We first show the convergence of $\bar{\Lambda}_n$ to Λ_0 as follows. We will use a property that the class of all functions from a closed set to $\mathbb{R}$, which are uniformly bounded and of bounded variation, is Glivenko–Cantelli. Consider the denominator of $\bar{\Lambda}_n$. The assumptions imply that functions of the form $u \to E_{\theta_0}[\exp\{\beta_0 g(u) A^*\} I(T < u \leq Y)|o]$, where o denotes the observed data of a subject, are uniformly bounded and of bounded variation, so the class of these functions is Glivenko–Cantelli. Therefore,

(A.3)

$$\frac{1}{n} \sum_{i=1}^{n} E_{\theta_0}[\exp\{\beta_0 g(u) A_i^*\}|o_i] I(t_i < u \leq z_i)$$

$$\to E_{\theta_0}\{\exp[\beta_0 g(u) A^*] I(T < u \leq Y)|Y^* \geq T^*\}$$

uniformly on $[\tau_1, \tau_2]$. Along with assumption C4, the uniform convergence of the inverse of the right-hand side to the inverse of the left-hand side in (A.3) holds. Moreover, uniform boundedness and bounded variation of functions $t \to \Delta I(Y \leq$

t) imply the Glivenko–Cantelli property of the class consisting of them. Thus we also have

$$(A.4) \qquad \frac{1}{n} \sum_{i=1}^{n} \Delta_i I(Y_i < t) \to E_{\theta_0}[\Delta I(Y < t)]$$

uniformly on $[\tau_1, \tau_2]$. Since

$$\Lambda_0(t) = E\left[\frac{\Delta I(Y \leq t)}{E\{\exp[\beta_0 g(Y) A^*] I(T < u \leq Y) | Y^* \geq T^*\} | u = Y} \right],$$

combining the convergence of the inverse of both sides in (A.3) and (A.4), we obtain $\bar{\Lambda}_n$ converges uniformly to Λ_0 on $[\tau_1, \tau_2]$. By considering the uniform convergence of the ratio of $\hat{\Lambda}\{u\}/\bar{\Lambda}\{u\}$ to $d\Lambda^*(u)/d\Lambda_0(u)$ for $u \in [\tau_1, \tau_2]$, as demonstrated on pages 2146–2147 in Zeng and Cai (2005), the uniform convergence of $\hat{\Lambda}$ to Λ^* is established. The remaining task is to prove the equivalence of $\theta^* = (\gamma^*, \Lambda^*)$ and $\theta_0^* = (\beta_0, \sigma^0, \alpha^*, \Lambda_0)$. This can be done by considering the empirical mean of the distance between $l_i^m(\hat{\theta}_n)$ and $l_i^m(\beta_0, \sigma^0, \alpha^*, \bar{\Lambda}_n)$ and demonstrating that $E_{\theta_0^*}[l^m(\theta^*)/l^m(\theta_0^*)] = 0$ almost surely as shown on page 910 in Dupuy, Grama and Mesbah (2006). Thus $(\hat{\sigma}^2, \hat{\beta}, \hat{\Lambda}_0)$ converges uniformly to $(\sigma_0^2, \beta_0, \Lambda_0)$.

A.3. Proof of asymptotic normality of the NPMMLE. We will apply Theorem 3.3.1 in van der Vaart and Wellner (1996) to prove the asymptotic normality of the NPMMLE $(\hat{\gamma}, \hat{\Lambda})$. The proof consists of four steps to verify each of the four conditions in their theorem.

Step 1. *Fréchet differentiability of the score functions.* For notation simplification, the parameter σ^2 will be combined with α into $\gamma_1 = (\sigma^2, \alpha)$ so that the single parameter γ_1 denotes the parameter of the measurement error ε and the latent random variable A^*. Thus, the new parameter vector is $\theta = (\gamma_1, \beta, \Lambda)$.

Consider a one-dimensional submodel along the direction (h_1, h_2, h_3) of the form

$$\theta_t = (\gamma_1 + th_1, \beta + th_2, \Lambda_t(h_3)),$$

where

$$\Lambda_t(h_3)(\cdot) = \int_0^{\cdot} (1 + th_3(u)) \, d\Lambda(u),$$

$h_1 \in \mathbb{R}^d$, $h_2 \in \mathbb{R}$ and h_3 is a bounded-variation function on $[0, \tau_2]$. Let $H = \{h = (h_1, h_2, h_3)\}$ and $H_p = \{h \in H : \|h_1\|, |h_2|, \|h_3\|_v \leq p\}$. The notation $\|\cdot\|_v$ denotes the absolute value evaluated at 0 plus the total variation of the argument. The imputed log-likelihood contributed by the ith subject evaluated at θ, given the

current value of parameter denoted as $\tilde{\theta}$, is denoted by $l_{\tilde{\theta},i}(\theta)$. The corresponding score function of the local parameter t is

$$\frac{\partial}{\partial t}l_{\tilde{\theta},i}(\theta_t) = h_2 \mathrm{E}_{\tilde{\theta}}\bigg[\delta_i g(z_i) A_i^*$$

$$- \int_{t_i}^{z_i} g(u) A_i^* \exp\{(\beta + th_2)g(u)A_i^*\}(1 + th_3(u))\,d\Lambda(u)\Big|o_i\bigg]$$

$$- \mathrm{E}_{\tilde{\theta}}\bigg[\int_{t_i}^{z_i} h_3(u) \exp\{(\beta + th_2)g(u)A_i^*\}\,d\Lambda(u)\Big|o_i\bigg]$$

$$+ h_1 \mathrm{E}_{\tilde{\theta}}\bigg[\frac{\partial}{\partial t} f_{\varepsilon,A^*}(\varepsilon, A_i^*|\gamma_1 + th_1)\Big|o_i\bigg]$$

$$+ \frac{\delta_i h_3(z_i)}{1 + th_3(z_i)}.$$

Thus the imputed score function of t contributed by the n subjects evaluated at $t = 0$ is

$$S_{n,\tilde{\theta}}(\theta)(h) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial t}l_{\tilde{\theta},i}(\theta_t)\bigg|_{t=0}$$

(A.5)

$$= h_1^T S_{n,\tilde{\theta},1}(\theta) + h_2 S_{n,\tilde{\theta},2}(\theta) + S_{n,\tilde{\theta},3}(\theta)(h_3),$$

where

$$S_{n,\tilde{\theta},1}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}_{\tilde{\theta}}\bigg[\frac{\partial}{\partial \gamma_1} \log f_{\varepsilon,A^*}(\varepsilon_i, A_i^*|\gamma_1)\Big|o_i\bigg],$$

$$S_{n,\tilde{\theta},2}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}_{\tilde{\theta}}\bigg[\delta_i g(z_i) A_i^* - \int_{t_i}^{z_i} g(u) A_i^* \exp\{\beta g(u)A_i^*\}\,d\Lambda u\Big|o_i\bigg],$$

$$S_{n,\tilde{\theta},3}(\theta)(h_3) = \frac{1}{n}\sum_{i=1}^{n}\bigg\{\delta_i h_3(z_i) - \mathrm{E}_{\tilde{\theta}}\bigg[\int_{t_i}^{z_i} h_3(u) \exp\{\beta g(u)A_i^*\}\,d\Lambda(u)\Big|o_i\bigg]\bigg\}.$$

By defining $\theta(h) = (\gamma_1, \beta, \Lambda)(h_1, h_2, h_3) = h_1^T \gamma_1 + h_2\beta + \int_0^{\tau 2} h_3(u)\,d\Lambda(u)$, where $h \in H_p$, the parameter θ can be regarded as a functional on H_p, the parameter space $\Theta = \{\theta\}$ is a subspace of $L^\infty(H_p)$ and the score in (A.5) is a random map from Θ to a Banach space which contains functions (operations) of h.

Besides the above imputed score, we also need the mean imputed score function of t under the true value θ_0 and denote it as

$$S_{\tilde{\theta}}(\theta)(h) = \mathrm{E}_{\theta_0}\bigg[\frac{\partial}{\partial t}l_{\tilde{\theta}}(\theta_t)\bigg|_{t=0}\bigg] = h_1^T S_{n\tilde{\theta},1}(\theta) + h_2 S_{\tilde{\theta},2}(\theta) + S_{\tilde{\theta},3}(\theta)(h_3),$$

where

$$S_{\tilde{\theta},1}(\theta) = \mathrm{E}_{\theta_0}\bigg\{\mathrm{E}_{\tilde{\theta}}\bigg[\frac{\partial}{\partial \gamma_1} \log f_{\varepsilon,A^*}(\varepsilon_i, A_i^*|\gamma_1)\Big|o_i\bigg]\bigg\},$$

$$S_{\tilde{\theta},2}(\theta) = \mathrm{E}_{\theta_0}\left\{\mathrm{E}_{\tilde{\theta}}\left[\Delta_i g(Y_i) A_i^* - \int_{T_i}^{Y_i} g(u) A_i^* \exp\{\beta g(u) A_i^*\} d\Lambda(u)\Big|o_i\right]\right\},$$

$$S_{\tilde{\theta},3}(\theta)(h_3) = \mathrm{E}_{\theta_0}\left\{\Delta_i h_3(Y_i) - \mathrm{E}_{\tilde{\theta}}\left[\int_{T_i}^{Y_i} h_3(u) \exp\{\beta g(u) A_i^*\} d\Lambda(u)\Big|o_i\right]\right\}.$$

To prove the Fréchet differentiability of the map, $\theta \to S_{\theta_0^*}(\theta)$ at θ_0^*, where $\theta_0^* = (\gamma_{10}^*, \beta_0, \Lambda_0)$ with $\gamma_{10}^* = (\sigma_0^2, \alpha^*)$, we need to calculate the corresponding derivative. First, we introduce a notation $\nabla_\theta S_{\tilde{\theta}}(\theta_0^*) = \frac{\partial}{\partial t} S_{\tilde{\theta}}(\theta_0^* + t\theta)|_{t=0}$, where $\theta_0^* + t\theta = (\alpha_0 + t\alpha, \beta_0 + t\beta, \Lambda_0(\cdot) + t\Lambda(\cdot))$. Then

$$\nabla_\theta S_{\tilde{\theta}}(\theta_0^*)(h)$$

$$= \frac{\partial}{\partial t} S_{\tilde{\theta}}(\theta_0^* + t\theta)(h)\bigg|_{t=0}$$

$$= \frac{\partial}{\partial t} \mathrm{E}_{\theta_0^*}\left\{h_1^T \mathrm{E}_{\tilde{\theta}}\left[\frac{\partial}{\partial(\gamma_{10}^* + t\gamma_1)} \log f_{\varepsilon,A^*}(\varepsilon_i, A_i^* | \gamma_{10}^* + t\gamma_1)\Big|o_i\right]\right.$$

$$+ h_2 \mathrm{E}_{\tilde{\theta}}\left[\Delta_i g(Y_i) A_i^*\right.$$

(A.6)
$$\left. - \int_{T_i}^{Y_i} g(u) A_i^* \exp\{\beta_0 + t\beta g(u) A_i^*\}(d\Lambda_0(u) + t\, d\Lambda(u))\Big|o_i\right]$$

$$+ \Delta_i h_3(Y_i)$$

$$- \mathrm{E}_{\tilde{\theta}}\left[\int_{T_i}^{Y_i} h_3(u)\right.$$

$$\left.\left.\times \exp\{(\beta_0 + t\beta)g(u) A_i^*\}(d\Lambda_0(u) + t\, d\Lambda(u))\Big|o_i\right]\right\}\bigg|_{t=0}.$$

Using the chain rule, equation (A.6) can be simplified as

$$-\gamma_1^T \sigma_{\tilde{\theta},1}(h) - \beta \sigma_{\tilde{\theta},2}(h) - \int_0^{\tau_2} \sigma_{\tilde{\theta},3}(h)(u)\, d\Lambda(u),$$

where

(A.7) $$\sigma_{\tilde{\theta},1}(h) = -\mathrm{E}_{\theta_0^*}\left\{h_1^T \mathrm{E}_{\tilde{\theta}}\left[\frac{\partial^2}{\partial\gamma_1\,\partial\gamma_1^T} \log f_{\varepsilon,A^*}(\varepsilon_i, A_i^* | \gamma_{10}^*)\Big|o_i\right]\right\},$$

$$\sigma_{\tilde{\theta},2}(h) = \mathrm{E}_{\theta_0^*}\left\{\mathrm{E}_{\tilde{\theta}}\left[\int_0^{\tau_2} [h_2 g(u) A_i^* + h_3(u)]g(u) A_i^* \exp\{\beta_0 g(u) A_i^*\}\right.\right.$$

(A.8)
$$\left.\left.\times I(T_i < u \leq Y_i)\, d\Lambda_0(u)\Big|o_i\right]\right\},$$

$$\sigma_{\tilde{\theta},3}(h)(u) = \mathrm{E}_{\theta_0^*}\{\mathrm{E}_{\tilde{\theta}}[[h_2 g(u) A_i^* + h_3(u)] \exp\{\beta_0 g(u) A_i^*\}$$

(A.9)
$$\times I(T_i < u \leq Y_i)|o_i]\}.$$

Evaluating (A.6) at the true value θ_0^* leads to

$$(A.10) \quad \nabla_\theta S_{\theta_0^*}(\theta_0^*)(h) = -\gamma_1^T \sigma_{\theta_0^*,1}(h) - \beta\sigma_{\theta_0^*,2}(h) - \int_0^{\tau_2} \sigma_{\theta_0^*,3}(h)(u)\,d\Lambda(u),$$

where each of the σ-function has similar forms as the corresponding function listed in (A.7), (A.8) or (A.9) with the double expectation $E_{\theta_0^*}\{E_{\tilde\theta}[[\cdot|o_i]]\}$ replaced by $E_{\theta_0^*}\{\cdot\}$. Now apply the Taylor expansion of $\exp\{(\beta_0 + t\beta)g(u)A_i^*\}$ at $t = 0$, to get

$$S_{\theta_0^*}(\theta_0^* + t\theta) - S_{\theta_0^*}(\theta_0^*) - \nabla_\theta S_{\theta_0^*}(\theta_0^*) = o(t),$$

where the small-o function does not depend on θ. Therefore,

$$\frac{\|S_{\theta_0^*}(\theta_0^* + t\theta) - S_{\theta_0^*}(\theta_0^*) - \nabla_\theta S_{\theta_0^*}(\theta_0^*)\|_p}{t} \to 0 \qquad \text{as } t \to 0$$

uniformly in $\theta = (\gamma_1, \beta, \Lambda)$. Thus the Fréchet derivative of the mapping $\theta \to S_{\theta_0^*}(\theta)$ evaluated at θ_0^* takes the form (A.10). We will use the notation $\dot{S}_{\theta_0^*}(\theta_0^*)(\theta)$ to denote it.

Step 2. *Continuous invertibility of* $\dot{S}_{\theta_0^*}(\theta_0^*)(\theta)$. The continuous invertibility of the Fréchet derivative can be established by showing that there exists some number $c > 0$ such that

$$(A.11) \qquad \inf_{\theta \in \text{lin}\Theta} \frac{\|\dot{S}_{\theta_0^*}(\theta_0^*)\|_{l^\infty(H)}}{\|\theta\|_{l^\infty(H)}} > c.$$

Since $\dot{S}_{\theta_0^*}(\theta_0^*)(\theta)$ can be expressed as a linear combination of the three σ-operators according to (A.6), it is necessary to check the continuous invertibility of those σ-operators. The proof is similar to the arguments in the Appendix of Zeng and Cai (2005). Through the continuous invertibility of $\sigma_{\theta_0^*}$, the lower bound c can be found as $\frac{q}{3p}$, where q satisfies $\sigma_{\theta_0^*}^{-1}(H_q) \subseteq H_p$. Details to find the lower bound are analogous to the approach in Dupuy, Grama and Mesbah (2006) (page 915). Thus the derivative $\dot{S}_{\theta_0^*}(\theta_0^*)$ is continuously invertible.

Step 3. *Convergence in distribution to a tight element*. In this step, the convergence of $\sqrt{n}(S_{n,\hat\theta_n} - S_{\theta_0^*})(\theta_0^*)$ in distribution will be demonstrated. Since $S_{\theta_0^*}(\theta_0^*)$ is the mean of the score function evaluated at the true value of θ, it is equal to zero. Then

$$[S_{n,\hat\theta_n} - S_{\theta_0^*}](\theta_0^*)(h) = \frac{1}{n}\sum_{i=1}^n [D_{i,1}(h) + D_{i,2}(h) + \delta_i h_3(y_i) + D_{i,3}(h)],$$

where

$$D_{i,1}(h) = h_1^T E_{\hat\theta_n}\left[\frac{\partial}{\partial\alpha}\log f_{\varepsilon,A^*}(\varepsilon_i, A_i^*|\gamma_{10}^*)\Big|o_i\right],$$

$$D_{i,2}(h) = h_2 E_{\hat\theta_n}\left[\delta_i g(z_i)A_i^* - \int_{t_i}^{z_i} g(u)A_i^*\exp\{\beta_0 g(u)A_i^*\}\,d\Lambda_0 u\Big|o_i\right],$$

$$D_{i,3}(h) = -E_{\hat\theta_n}\left[\int_{t_i}^{z_i} h_3(u)\exp\{\beta_0 g(u)A_i^*\}\,d\Lambda_0(u)\Big|o_i\right].$$

The class $\{\frac{1}{n}\sum (D_{i,1} + D_{i,2})(h) : \|h_1\| + |h_2| \le p\}$ is bounded Donsker, since it is a finite dimensional class of measurable score functions. Moreover, since any class of real-valued functions on $[0, \tau_2]$ that are uniformly bounded and bounded in variation is Donsker, the class $\{\delta h_3(y) : h_3 \in BV_p\}$ is Donsker. The Donsker property of the class $\{\frac{1}{n}\sum D_{i,3}(h) : h_3 \in BV_p\}$ also follows from this fact. We have thus shown that the class $\{[S_{n,\hat{\theta}_n} - S_{\theta_0^*}](\theta_0^*)(h) : \|h_1\| + |h_2| \le p, h_3 \in BV_p\}$ is Donsker, since the sum of bounded Donsker classes is also Donsker. This implies

$$\sqrt{n}(S_{n,\hat{\theta}_n} - S_{\theta_0^*})(\theta_0^*) \xrightarrow{D} Z,$$

a tight Gaussian process in $l_\infty(H_p)$.

Step 4. *Verification of conditions* 1 *and* 4. Condition 4 holds by the consistency of the estimator $\hat{\theta}_n$. Condition 1 can be verified by considering the Donsker property of the class $\{S_{\cdot,\theta}(\theta)(h) - S_{\cdot,\theta_0^*}(\theta_0^*)(h) : \|\theta - \theta_0^*\|_p < \nu, h \in H_p\}$ for some $\nu > 0$, where $S_{\cdot,\theta}(\theta)(h)$ is the general form of $S_{i,\theta}(\theta)(h) = \frac{\partial}{\partial t} l_{\theta,i}(\theta_t)|_{t=0}$. We omit the details since they are similar to those for the case of right-censored data, considered in Zeng and Cai (2005).

We have verified the four conditions needed for the asymptotic distribution of the NPMMLE $\hat{\theta}_n$, and therefore

$$\sqrt{n}(\hat{\theta}_n - \theta_0^*) \xrightarrow{D} -\dot{S}_{\theta_0^*}(\theta_0^*)Z,$$

as $n \to \infty$.

Using the form of the Fréchet derivative in (A.6), one finds that there exists a linear operator $\sigma = (\sigma_{\theta_0^*,1}, \sigma_{\theta_0^*,2}, \sigma_{\theta_0^*,3})$ that maps H_p to $\mathbb{R}^{d+1} \times BV_p$, such that

$$\dot{S}_{\theta_0^*}(\theta_0^*)(\theta_1 - \theta_2)(h) = -(\gamma_{11} - \gamma_{12})^T \sigma_{\theta_0^*,1}(h) - (\beta_1 - \beta_2)\sigma_{\theta_0^*,2}(h)$$

$$- \int_0^{\tau_2} \sigma_{\theta_0^*,3}(h)(u)\, d(\Lambda_1 - \Lambda_2)(u).$$

The continuous invertibility of the σ operator has been shown already, so its inverse operator, denoted by σ^{-1}, exists. Since

$$\sqrt{n}\dot{S}_{\theta_0^*}(\theta_0^*)(\hat{\gamma}_1 - \gamma_{10}^*, \hat{\beta} - \beta_0, \hat{\Lambda} - \Lambda_0)(h) = \sqrt{n}\{S_{n,\theta_0^*}(h) - S_{\theta_0^*}(\theta_0^*)(h)\} + o_p(1),$$

by applying the inverse operator σ^{-1} on both sides we obtain that

$$
\text{(A.12)} \quad
\begin{aligned}
&\sqrt{n}\left\{-(\hat{\gamma}_1 - \gamma_{10}^*)^T h_1 - (\hat{\beta} - \beta_0)h_2 - \int_0^{\tau_2} h_3(u)\, d(\hat{\Lambda} - \Lambda_0)(u)\right\} \\
&= \sqrt{n}\{S_{n,\theta_0^*}(\tilde{h}) - S_{\theta_0^*}(\theta_0^*)(\tilde{h})\} + o_p(1),
\end{aligned}
$$

where $\tilde{h} = (\tilde{h}_1, \tilde{h}_2, \tilde{h}_3) = \sigma^{-1}(h)$. If h_1 and h_3 in (A.12) are chosen to be 0, then this reduces to

$$\sqrt{n}\{-(\hat{\beta} - \beta_0)h_2\}$$
$$= \sqrt{n}\{S_{n,\theta_0^*}(\tilde{h}) - S_{\theta_0^*}(\theta_0^*)(\tilde{h})\} + o_p(1),$$

where the latter term is in the form of linear combinations of score functions for the parameters. Since score functions derived from the modified likelihood is asymptotically equivalent to those from the full likelihood by Lemma A.1, the influence function is the same as the efficient influence function for $\beta_0 h_2$ by its uniqueness in the linear span of the scores. Thus the estimator $\hat{\beta}$ is efficient for β_0.

REFERENCES

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. *Springer Series in Statistics*. Springer, New York. MR1198884

COX, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758

DAFNI, U. G. and TSIATIS, A. A. (1998). Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics* **54** 1445–1462.

DEGRUTTOLA, V. and TU, X. (1994). Modeling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50** 1003–1014.

DUPUY, J.-F., GRAMA, I. and MESBAH, M. (2006). Asymptotic theory for the Cox model with missing time-dependent covariate. *Ann. Statist.* **34** 903–924. MR2283397

HENDERSON, R., DIGGLE, P. and DOBSON, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1** 465–480.

HSIEH, F., TSENG, Y.-K. and WANG, J.-L. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics* **62** 1037–1043. MR2297674

KLEIN, J. P. and MOESCHBERGER, M. L. (2003). *Survival Analysis*: *Techniques for Censored and Truncated Data*. Springer, New York.

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

LOUIS, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **44** 226–233. MR0676213

LYNDEN-BELL, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices of the Royal Astronomy Society* **155** 95–118.

MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–485. MR1803168

REZZA, G., LAZZARIN, A., ANGARANO, G., SINICCO, A., PRISTERÁ, R., TIRELLI, U., SALASSA, B., RICCHI, E., AIUTI, F. and MENNITI-LPPOLITO, F. (1989). Tje natural history of HIV infection in intravenous drug users: Risk of disease progression in a cohort of serconverters. *AIDS* **3** 87–90.

SONG, X., DAVIDIAN, M. and TSIATIS, A. A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58** 742–753. MR1945011

THE-ITALIAN-SEROCONVERSION-STUDY (1992). Disease progression and early predictors of AIDS in HIV-seroconverted injecting drug users. *AIDS* **6** 421–426.

TSENG, Y.-K., HSIEH, F. and WANG, J.-L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika* **92** 587–603. MR2202648

TSIATIS, A. A. and DAVIDIAN, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statist. Sinica* **14** 809–834. MR2087974

TSIATIS, A. A., DEGRUTTOLA, V. and WULFSOHN, M. (1995). Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and CD4 counts in patients with AIDS. *J. Amer. Statist. Assoc.* **90** 23–37.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. Springer, New York. MR1385671

VARDI, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.* **13** 178–205. MR0773161

WANG, M.-C. (1987). Product limit estimates: A generalized maximum likelihood study. *Comm. Statist. Theory Methods* **16** 3117–3132. MR0917759

WANG, C. Y. (2006). Corrected score estimator for joint modeling of longitudinal and failure time data. *Statist. Sinica* **16** 235–253. MR2256090

WOODROOFE, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13** 163–177. MR0773160

WULFSOHN, M. S. and TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53** 330–339. MR1450186

ZENG, D. and CAI, J. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *Ann. Statist.* **33** 2132–2163. MR2211082

GRADUATE GROUP IN BIOSTATISTICS   DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, DAVIS   UNIVERSITY OF CALIFORNIA, DAVIS
DAVIS, CALIFORNIA 95616   DAVIS, CALIFORNIA 95616
USA   USA
AND   E-MAIL: jlwang.ucdavis@gmail.com
DEPARTMENT OF STATISTICS
NATIONAL CHENG KUNG UNIVERSITY
TAINAN 701
TAIWAN
E-MAIL: yrsu@stat.ncku.edu.tw