

Joint Modeling of Longitudinal and Cure-survival Data

Sehee Kim, *Department of Biostatistics,*
University of Michigan, Ann Arbor, Michigan 48109, U.S.A.
Email: seheek@umich.edu

Donglin Zeng, *Department of Biostatistics,*
University of North Carolina, Chapel Hill, NC 27599, U.S.A.
Email: dzeng@bios.unc.edu

Yi Li, *Department of Biostatistics,*
University of Michigan, Ann Arbor, Michigan 48109, U.S.A.
Email: yili@umich.edu

Donna Spiegelman, *Departments of Epidemiology and Biostatistics,*
Harvard School of Public Health, Boston, MA 02115, U.S.A.
Email: stdls@hsph.harvard.edu

*Received: *** Revised: ****

Abstract

This article presents semiparametric joint models to analyze longitudinal measurements and survival data with a cure fraction. We consider a broad class of transformations for the cure-survival model, which includes the popular proportional hazards structure and the proportional odds structure as special cases. We propose to estimate all the parameters using the nonparametric maximum likelihood estimators (NPMLE). We provide the simple and efficient EM algorithms to implement the proposed inference procedure. Asymptotic properties of the estimators are shown to be asymptotically normal and semiparametrically efficient. Finally, we demonstrate the good performance of the method through extensive simulation studies and a real-data application.

AMS Subject Classification: Here the AMS Subject Classification <http://www.ams.org/msc/>

Keywords: Cure-survival data; Joint models; Longitudinal data; Nonparametric maximum likelihood; Random effects; Transformation models.

1 Introduction

In cancer research, along with repeatedly measured biomarkers, it is often observed that a certain proportion of patients are cured, immune, or unsusceptible to the event of interest. For example, cancer can be considered cured if all metastasis-competent tumor cells are successfully removed by treatment, and hence no recurrence will be observed. In these situations since the survival curve will never reach to zero due to these cured patients, the standard survival analyses based on non-cure models are inconvenient for inference on cure rates. The existence of this possible cure fraction can be evident from a stable plateau at the right tail of survival curve as in Figure 1 (a).

To analyze such cure-survival data, two classes of cure rate models are commonly used: mixture cure model and promotion time cure model. The former mixture cure model is on the basic concept that the underlying population is a mixture of two subpopulations of the cured and uncured with the respective probabilities $p_c(\mathbf{Z}_i)$ and $1 - p_c(\mathbf{Z}_i)$, that is $S_{pop}(t|\mathbf{Z}_i) = p_c(\mathbf{Z}_i) + \{1 - p_c(\mathbf{Z}_i)\}S_{uc}(t|\mathbf{Z}_i)$, where \mathbf{Z}_i is the vector of covariates, $S_{uc}(t|\mathbf{Z}_i)$ is the conditional survival function of the uncured population (Berkson and Gage, 1952; Laska and Meisner, 1992; Maller and Zhou, 1996; Kuk and Chen, 1992; Sy and Taylor, 2000; Lu and Ying, 2004). It is assumed that all patients in the uncured subpopulation will eventually experience the event while those in the cured subpopulation will never. Lu and Ying (2004) considered a class of transformation models for the event time where generalized estimating equations were used for parameter estimation, and the asymptotic properties were established by the usual counting process and its associated martingale theory. However, their linear transformation approach was limited to time-independent covariates only. Therefore, as alternative to the mixture cure models, we focus on the promotion time cure model that can handle transformation with time-dependent covariates when there is a need for transformations, for example, in cases where the violated proportional hazards assumption is of concern. The promotion time cure model has been proposed under biological assumptions that a patient has N metastatic tumor cells remaining after treatment (Yakovlev et al., 1996; Tsodikov, 1998; Chen et al., 1999; Tsodikov et al., 2003) in cancer clinical trials. Let N_i be the number of metastatic cancerous cells of the i th patient, which is an unobservable latent variable, following a Poisson distribution with mean $\pi(\mathbf{Z}_i)$. We denote the time for the k th metastatic cancer cell to produce a detectable tumor (promotion time) by \tilde{T}_k ($k = 1, \dots, N_i$) and assume that, conditional on N_i , \tilde{T}_k 's are independent and identically distributed (iid) with a common distribution function $F(t)$. Viewing $F(t) = 1 - S_{uc}(t)$, it can be interpreted similarly to the distribution function of the uncured patients in the mixture cure model. Then, the time to relapse of cancer for the i th patient, defined by $T_i = \min\{\tilde{T}_1, \dots, \tilde{T}_{N_i}\}$, has a form of the survival function

$$\begin{aligned}
 S(t|\mathbf{Z}_i) &= P[N_i = 0] + \sum_{k \geq 1} P[\tilde{T}_1 > t, \dots, \tilde{T}_{N_i} > t | N_i = k]P[N_i = k] \\
 &= \exp\{-\pi(\mathbf{Z}_i)\} + \sum_{k \geq 1} \{1 - F(t)\}^k \frac{\pi(\mathbf{Z}_i)^k \exp\{-\pi(\mathbf{Z}_i)\}}{k!} \\
 &= \exp\{-\pi(\mathbf{Z}_i)F(t)\}, \tag{1}
 \end{aligned}$$

where $\pi(\mathbf{Z}_i)$ is a known link function. In the promotion time cure model (1), the survival function is integrated into one formulation regardless of cured or uncured, and we can see that (1) retains the proportional hazards structure when $\pi(\mathbf{Z}_i) = \exp(\beta^T \mathbf{Z}_i)$. Moreover, if the regression coefficients β include an intercept term, say β_0 , the baseline cumulative hazard function is equal to $\exp(\beta_0)F(t)$, which implies that the model (1) becomes the Cox proportional hazards model with a bounded baseline cumulative hazard. If the cured patients exist in the population, the survival rate at $t = \infty$ can naturally be interpreted as the cure rate, i.e., the cure rate is $S(\infty|\mathbf{Z}_i) = \exp\{-\pi(\mathbf{Z}_i)\} \neq 0$, leading to an improper survival function.

Some of these curable diseases are associated with longitudinal biomarkers, and it is often of interest to model these two different types of data as outcomes simultaneously. We will therefore propose a joint model to analyze longitudinal and survival data with a cure fraction. While there has been a great deal of work done on joint modeling of longitudinal and cure-survival data based on the mixture cure models (Law et al., 2002; Yu et al., 2004, 2008), there has been scant literature about joint modeling based on the promotion time cure models. For a more detailed review on the joint mixture cure models see Yu et al. (2004). Instead, we will provide a brief review of the alternative approach, joint promotion time cure models that we adopt here. Brown and Ibrahim (2003) and Chen et al. (2004) proposed joint promotion time cure model with emphasis on two different types of longitudinal models. To model immune response, Brown and Ibrahim (2003) proposed a longitudinal model with a point mass at zero that changed over time and subjects in probability. On the other hand, Chen et al. (2004) considered a true immune response was unobservable, and adopted the longitudinal model in the context of measurement errors. In common, they considered a piecewise constant function to estimate the baseline distribution $F(t)$. For inference, they used Bayesian approaches via Gibbs sampler and Markov chain Monte Carlo sampling, which may be straightforward ways to proceed, but computationally intensive, compared to frequentist analyses. Therefore, we propose a joint cure model based on a frequentist approach to balance out complication brought by relaxing a functional form of the baseline distribution $F(t)$ as well as the proportionality assumption.

The objective of the article is to present a flexible joint promotion time cure model based on a frequentist inference that is a new approach to the existing joint promotion time cure models. To account for the long-term plateau at the tail of survival distribution resulting from the existence of cured patients, we propose a broad class of transformed promotion time cure models, which integrate the popular proportional hazards and odds cure models in to one general form. Inference procedures using the nonparametric maximum likelihood estimation (NPMLE) are developed, a simple and efficient algorithm is provided for its implementation, and the new joint cure model is illustrated with the colorectal cancer data from the Health Professionals Follow-up Study (HPFS). The proposed work here advances to existing joint cure models in: 1) flexibility throughout nonparametric baseline distribution function and transformation models; 2) extended ability to handle time-varying covariates; and 3) well-established asymptotic properties of the NPMLEs.

2 Joint Transformation Models

2.1 Joint Models

Let $Y(t)$ be the longitudinal measurement at time t , T be the time to the survival event, and $\mathcal{Z} = \{Z(t); t \geq 0\}$ be the covariate process, where $Z(t)$ is the vector of external covariates at time t , possibly time-varying. We introduce latent random effects to account for the correlation between longitudinal and survival components on the same subject. Particularly, let b denote the subject-specific random effects following a multivariate normal distribution with mean zeros and covariance matrix Σ_b . We further assume that $Y(t)$ and T are independent, conditional on \mathcal{Z} and b . Then, the proposed joint model for the longitudinal data $Y(\cdot)$ and the population survival function of T with a cure fraction are given by

$$\begin{aligned} Y(t) &= \alpha^T Z_1(t) + b^T \tilde{Z}_1(t) + \epsilon(t), \\ S(t | \mathcal{Z}, b) &= \exp \left\{ -H \left(\int_0^t e^{\beta^T Z_2(u) + (\psi \circ b)^T \tilde{Z}_2(u)} dF(u) \right) \right\}, \end{aligned} \quad (2)$$

where α and β are vectors of unknown regression parameters in the longitudinal and survival components, respectively, $Z_k(t)$ and $\tilde{Z}_k(t)$ ($k = 1, 2$) are subsets of $Z(t)$ plus the unit component, and $F(t)$ is an unspecified distribution function of the event times. In addition, $\epsilon(t)$ is a white noise process with mean zero and variance σ_e^2 , ψ is a set of unknown constants with the same number of elements as b , and $\psi \circ b$ denotes the component-wise product of ψ and b . Note in (2) that the correlation among the longitudinal outcomes is formulated through the latent random effects b , and that the association between longitudinal outcomes and the event time is characterized by ψ with the shared latent variables b . Thus, for a fixed covariate \mathcal{Z} , $\psi > 0$ implies the larger longitudinal measures are, the higher hazard rate of the event is. On the other hand, $\psi = 0$ implies that the association can be fully explained by the common covariates in both longitudinal and survival components. The transformation function $H(\cdot)$ is assumed to be continuously differentiable and strictly increasing, and we will discuss about $H(\cdot)$ in more detail in Section 2.2.

We notice that the survival model for the entire population in (2) encompasses more general regression models, extending to a cure rate defined as an asymptotic value of the population survival function when $t \rightarrow \infty$. This definition does not imply that the observed survival time should be infinite since the censoring time (by death from other causes or the end of study, for example) is finite with probability 1. In practice, a sufficiently long follow-up period from a clinician's perspective can be interpreted as $t = \infty$. That is, the cure rate model can be expressed as

$$\lim_{t \rightarrow \infty} S(t | \mathcal{Z}, b) = \exp \left\{ -H \left(\int_0^\infty e^{\beta^T Z_2(u) + (\psi \circ b)^T \tilde{Z}_2(u)} dF(u) \right) \right\}.$$

Thus, our joint cure-survival model (2) allows us to explore a link between the longitudinal measures and the probability of being cured through the shared random effects as well as covariates. Especially, when $Z_2(t)$ and $\tilde{Z}_2(t)$ are time-independent covariates, z_2 and \tilde{z}_2 , respectively, the cure rate can be simplified to

$$E_b[\exp\{-H(\exp\{\beta^T z_2 + (\psi \circ b)^T \tilde{z}_2\})\}],$$

where E_b is taking expectation with respect to b . In fact, it is always true that the conditional cure rate is $\lim_{t \rightarrow \infty} E_b[S(t | \mathcal{Z}, b)] > 0$ (improper survival function), because $H(\cdot)$ is assumed to be finite.

Let C be the non-informative censoring time which is independent of $(Y(\cdot), T, b)$ given \mathcal{Z} , and let $X = \min(T, C)$ denote the observed event time. The observed data for the i th subject with m_i repeated measurements are defined as $O_i = \{Y_i(t_{ik}), X_i, \Delta_i, Z(t); t_{ik} \leq X_i, t \leq X_i, i = 1, \dots, n, k = 1, \dots, m_i\}$, where $\Delta_i = I(T_i \leq C_i)$ with $I(\cdot)$ being the indicator function. Under the model (2), the log-likelihood function for the observed data is given by

$$\begin{aligned} & \sum_{i=1}^n \log \int_b \prod_{k=1}^{m_i} \left[\frac{1}{\sqrt{2\pi\sigma_e^2}} \exp \left\{ \frac{-(Y_i(t_{ik}) - \alpha^T Z_{1i}(t_{ik}) - b^T \tilde{Z}_{1i}(t_{ik}))^2}{2\sigma_e^2} \right\} \right] \\ & \times \left[f(X_i) e^{\beta^T Z_{2i}(X_i) + (\psi_0 b)^T \tilde{Z}_{2i}(X_i)} H' \left(\int_0^{X_i} e^{\beta^T Z_{2i}(u) + (\psi_0 b)^T \tilde{Z}_{2i}(u)} dF(u) \right) \right]^{\Delta_i} \\ & \times \exp \left\{ -H \left(\int_0^{X_i} e^{\beta^T Z_{2i}(u) + (\psi_0 b)^T \tilde{Z}_{2i}(u)} dF(u) \right) \right\} \times f(b; \Sigma_b) db, \end{aligned} \quad (3)$$

where $f(b; \Sigma_b)$ is the density function of b with the parameters Σ_b , and $f(t) = dF(t)/dt$ and $H'(x) = dH(x)/dx$ are the first derivatives of $F(t)$ and $H(x)$, respectively.

2.2 Transformation of Promotion Time Cure Models

In the model (2), $H(\cdot)$ represents a transformation function of the conditional cumulative hazard function, which is required to be pre-specified in the analysis. For example, $H(x)$ can take a form of the logarithmic transformation,

$$H(x) = \begin{cases} \log(1 + \eta x)/\eta, & \eta > 0 \\ x, & \eta = 0. \end{cases}$$

The choices of $\eta = 0$ and $\eta = 1$ lead to the proportional hazards structure and the proportional odds structure, respectively.

In fact, the transformation $H(\cdot)$ has been derived from a biological explanation. Remind that the promotion time cure model without transformation in (1) is based on the conditional independence assumption of $\{\tilde{T}_k | N_i; k = 1, \dots, N_i\}$. However, this assumption may not be satisfied in practice since there are common features shared by the same patient, such as the patient's underlying health condition or dietary habits. As a solution to adjust the correlated cancer progression times, Zeng et al. (2006) have introduced a subject-specific frailty ζ_i , and have assumed that $\{\tilde{T}_k | N_i, \zeta_i; k = 1, \dots, N_i\}$ are mutually independent with the distribution function $F(t)$. Note that ζ_i can reflect the underlying heterogeneity for the rate of metastatic cancer cells through N_i following the Poisson distribution with mean $\zeta_i \pi(\mathbf{Z}_i)$, conditional on (\mathbf{Z}_i, ζ_i) . Following the similar derivation to (1), the resulting survival function for the time to relapse T takes a form

$$S(t | \mathbf{Z}_i) = E_{\zeta_i} [\exp\{-\zeta_i \pi(\mathbf{Z}_i) F(t)\}], \quad (4)$$

where E_{ζ_i} denotes the expectation with respect to ζ_i . Explicitly specifying the distribution for ζ_i as a gamma distribution with unit mean and variance η , for instance, we can now see a desirable connection between (4) and the transformation $H(\cdot)$, as follows:

$$S(t|\mathbf{Z}_i) = [1 + \eta\pi(\mathbf{Z}_i)F(t)]^{-1/\eta} = \exp\{-H(\pi(\mathbf{Z}_i)F(t))\}.$$

3 Inference Procedure

3.1 NPMLs for Joint Transformation Models

We propose to use the nonparametric maximum likelihood estimation (NPML) for estimating parameters $\theta = (\alpha, \beta, \psi, \sigma_e^2, \text{Vec}(\Sigma_b))$ and infinite-dimensional parameter $F(t)$, where $\text{Vec}(\Sigma_b)$ denotes the vector consisting of the upper triangular elements of Σ_b . To obtain the NPMLs, in the log-likelihood function (3), we treat F as a step function with jumps only at the observed failure times and replace $f(t)$ by the jump size of F at t , which is denoted by $F\{t\}$.

For commonly used transformation functions such as a logarithmic transformation, $\exp\{-H(x)\}$ can be expressed as the Laplace transformation of some function $\phi(t)$, $t \geq 0$, such that

$$\exp\{-H(x)\} = \int_0^\infty \exp(-xt) \phi(t) dt.$$

For example, if we choose $\phi(t) = t^{1/\eta-1} \exp(-t/\eta) / \{\Gamma(1/\eta) \eta^{1/\eta}\}$, then it is true that $H(x) = \log(1 + \eta x) / \eta$. Applying the Laplace transformation with a subject-specific frailty ζ_i and using the fact that

$$H'(x) \exp\{-H(x)\} = \int_0^\infty \zeta \exp(-x\zeta) \phi(\zeta) d\zeta,$$

the observed log-likelihood function (3) can be rewritten as

$$\begin{aligned} & l_n(\theta, F\{\cdot\}) \\ &= \sum_{i=1}^n \log \int_b \prod_{k=1}^{m_i} \left[\frac{1}{\sqrt{2\pi\sigma_e^2}} \exp \left\{ \frac{-(Y_i(t_{ik}) - \alpha^T Z_{1i}(t_{ik}) - b^T \tilde{Z}_{1i}(t_{ik}))^2}{2\sigma_e^2} \right\} \right] \\ & \quad \times \int_{\zeta_i} \left[\zeta_i F\{X_i\} e^{q_{2i}(X_i)} \right]^{\Delta_i} \exp \left\{ - \int_0^{X_i} \zeta_i e^{q_{2i}(u)} dF(u) \right\} \phi(\zeta_i) d\zeta_i \\ & \quad \times f(b; \Sigma_b) db, \end{aligned} \tag{5}$$

where $q_{2i}(t) = \beta^T Z_{2i}(t) + (\psi \circ b)^T \tilde{Z}_{2i}(t)$, and we assume that ζ_i and b are independent. The most attractive feature about taking transformation in this way is that the modified log-likelihood (5) can be seen as the proportional hazards frailty model with the conditional hazard function

$$\lambda(t|Z(t), \zeta_i, b_i) = \zeta_i f(t) \exp\{\beta^T Z_{2i}(t) + (\psi \circ b)^T \tilde{Z}_{2i}(t)\}.$$

This makes the algorithm more stable and computationally efficient.

Now, the computation of the NPMLs is identical to maximizing the modified log-likelihood function with respect to θ and all jump sizes of F at the observed failure times. This maximization can be carried out through the following EM algorithm.

3.2 EM Algorithm

We describe the EM algorithm, treating ζ_i and b as missing data to compute the NPMLEs of $(\theta, F\{\cdot\})$. In the E-step, we calculate the conditional expectation of the log-likelihood function for the complete data, given the observed data O_i and the current parameter estimates. Particularly, we need to evaluate the integration of certain functions of (ζ_i, b) , say $\hat{E}[\zeta_i g_i(b) | O_i]$. Hereafter, we drop the conditional part on the observed data and the current parameter estimates, and abbreviate such expectation $\hat{E}[\zeta_i g_i(b) | O_i]$ as $\hat{E}[\zeta_i g_i(b)]$. Computation of this expectation can become doable by first obtaining the *nested* conditional expectation of ζ_i , given b and the observed data. That is, $\hat{E}[\zeta_i g_i(b)]$ can be calculated as $\hat{E}_b[\hat{E}_{\zeta_i}[\zeta_i | b] g_i(b)]$. With the fact that the conditional distribution of ζ_i given b is proportional to

$$h(\zeta_i, b) = \zeta_i^{\Delta_i} \exp \left\{ - \int_0^{X_i} \zeta_i e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)} dF(u) \right\},$$

and the useful relationships by the Laplace transformation, the conditional expectation of ζ_i given b has the form of

$$\hat{E}_{\zeta_i}[\zeta_i | b] = \int \zeta_i \frac{h(\zeta_i, b) \phi(\zeta_i)}{\int h(\zeta_i, b) \phi(\zeta_i) d\zeta_i} d\zeta_i = H'(\tilde{x}_i(b)) - \left[\frac{H''(\tilde{x}_i(b))}{H'(\tilde{x}_i(b))} \right]^{\Delta_i},$$

where $\tilde{x}_i(b) = \int_0^{X_i} e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)} dF(u)$. Once $\hat{E}_{\zeta_i}[\zeta_i | b]$ is calculated, which is a function of b , the conditional expectation $\hat{E}[\zeta_i g_i(b)]$ can be computed using numerical approximation methods such as the Gaussian quadrature with Hermite orthogonal polynomial. Since the conditional distribution of b given O_i is proportional to $\Gamma(O_i | b) f(b; \Sigma_b)$, the conditional expectation is calculated by

$$\hat{E}[\zeta_i g_i(b)] = \int_b \hat{E}_{\zeta_i}[\zeta_i | b] g_i(b) \frac{\Gamma(O_i | b) f(b; \Sigma_b)}{\int_b \Gamma(O_i | b) f(b; \Sigma_b) db} db,$$

where

$$\begin{aligned} \Gamma(O_i | b) = & \exp \left\{ - \sum_{k=1}^{m_i} \left[(b^T \tilde{Z}_{1i}(t_{ik}))^2 - 2(Y_i(t_{ik}) - \alpha_1^T Z_{1i}(t_{ik})) b^T \tilde{Z}_{1i}(t_{ik})) \right] / (2\sigma_e^2) \right\} \\ & \times \exp \left\{ \Delta_i \left[(\psi \circ b)^T \tilde{Z}_{2i}(X_i) + \log H' \left(\int_0^{X_i} e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)} dF(u) \right) \right] \right\} \\ & \times \exp \left\{ -H \left(\int_0^{X_i} e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)} dF(u) \right) \right\}. \end{aligned}$$

In the M-step, we maximize the following objective function of the expected log-

likelihood for the complete data:

$$\begin{aligned} & \sum_{i=1}^n \sum_{k=1}^{m_i} \left\{ -\log \sigma_e^2/2 - \hat{E} \left[(Y_i(t_{ik}) - \alpha^T Z_{1i}(t_{ik}) - b^T \tilde{Z}_{1i}(t_{ik}))^2 / (2\sigma_e^2) \right] \right\} \\ & + \sum_{i=1}^n \Delta_i \left\{ \hat{E}[\log \zeta_i] + \log F\{X_i\} + \beta^T Z_{2i}(X_i) + \hat{E}[\psi \circ b]^T \tilde{Z}_{2i}(X_i) \right\} \\ & + \sum_{i=1}^n \left\{ -\hat{E} \left[\int_0^{X_i} \zeta_i e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)} dF(u) \right] + \hat{E}[\log \phi(\zeta_i) + \log f(b; \Sigma_b)] \right\}, \end{aligned}$$

under the restriction of $\sum_{i=1}^n \Delta_i F\{X_i\} = 1$. Maximizing the above objective function over $(\alpha, \sigma_e^2, \Sigma_b)$ is simple; whereas the rest of parameters $(\beta, \psi, F\{\cdot\})$ do not yield the closed-form of maximizers, and hence it is required to involve a reliable numerical approach. By introducing the Lagrange multiplier μ , we solve the following equation for β :

$$\sum_{i=1}^n \Delta_i \left\{ Z_{2i}(X_i) - \frac{\sum_{j=1}^n R_j(X_i) Z_{2j}(X_i) \hat{E}[\zeta_j e^{q_{2j}(X_i)}]}{\sum_{j=1}^n R_j(X_i) \hat{E}[\zeta_j e^{q_{2j}(X_i)}] + \mu} \right\} = 0, \quad (6)$$

the following equation for ψ :

$$\sum_{i=1}^n \Delta_i \left\{ \hat{E}[b \circ \tilde{Z}_{2i}(X_i)] - \frac{\sum_{j=1}^n R_j(X_i) \hat{E}[\zeta_j e^{q_{2j}(X_i)} (b \circ \tilde{Z}_{2j}(X_i))]}{\sum_{j=1}^n R_j(X_i) \hat{E}[\zeta_j e^{q_{2j}(X_i)}] + \mu} \right\} = 0, \quad (7)$$

and the following equation for μ :

$$\sum_{i=1}^n \Delta_i F\{X_i\} = 1, \quad (8)$$

where $R_j(t) = I(X_j \geq t)$ and $q_{2j}(t) = \beta^T Z_{2j}(t) + (\psi \circ b)^T \tilde{Z}_{2j}(t)$. The restricted NPMLE along with a random effect induced transformation EM was presented by Tsodikov (2002), and the approach has been used in different models as reviewed in Tsodikov et al. (2003). Hence F is estimated as a step function with the following jump size at X_i :

$$F\{X_i\} = \frac{\Delta_i}{\sum_{j=1}^n R_j(X_i) \hat{E}[\zeta_j e^{q_{2j}(X_i)}] + \mu}. \quad (9)$$

To solve these equations at each M-step, we consider a two-step optimization. In the first step, we estimate μ using the bisection method based on the equation (8) and the fact $F\{X_i\} > 0$ ($i = 1, \dots, n$). Since the left side of (8) is a monotone decreasing function of μ by considering $F\{X_i\}$ as a function of μ in (9), the solution always exists. In the second step, to update β and ψ , we plug the estimates $\hat{\mu}$ into equations (6) and (7), treat them as the functions of $\hat{\mu}$, and solve the equations using one-step

Newton-Raphson algorithm. Updating the jump sizes of F can be easily done by the equation (9) with $\hat{\mu}$.

To obtain the NPMLEs, we iterate the E-step and M-step until the parameter estimates converge. The variances of the NPMLEs can be estimated from the inverse of the observed information matrix for all parameters of $(\theta, F\{\cdot\})$, under the restriction of $\sum_{i=1}^n \Delta_i F\{X_i\} = 1$. The observation information matrix can be computed from the complete data log-likelihood function denoted by ℓ_i^c for the i th subject using the following Louis formula (Louis, 1982) of

$$-\sum_{i=1}^n \hat{E}[\nabla^2 \ell_i^c(b) | O_i] - \sum_{i=1}^n \left\{ \hat{E}[\nabla \ell_i^c(b)^{\otimes 2} | O_i] - \hat{E}[\nabla \ell_i^c(b) | O_i]^{\otimes 2} \right\},$$

where $u^{\otimes 2} = uu^T$, ∇ and ∇^2 denote the first and the second derivatives with respect to parameters, and \hat{E} denotes the conditional expectation of a function of b given the observed data and is evaluated at the NPMLEs.

4 Asymptotic Properties

Let $(\hat{\theta}, \hat{F})$ denote the NPMLEs and (θ_0, F_0) denote the true parameter values of (θ, F) . Under the regularity conditions, we will establish the asymptotic properties of the NPMLEs under the following conditions:

- (A1) The true parameter value θ_0 belongs to the interior of a compact set Θ within the domain of θ .
- (A2) With probability 1, $Z(t)$ is left-continuous with uniformly bounded left and right derivatives in $[0, \infty]$.
- (A3) For some constant δ_0 , $P(C = \infty | \mathcal{Z}) > \delta_0 > 0$ with probability 1.
- (A4) For some positive constant M_0 , $M_0^{-1} < \sigma_{0e}^2 < M_0$ and $M_0^{-1} < c^T \Sigma_{0b} c < M_0$ for any constant vector $\|c\| = 1$.
- (A5) The transformation functions $H(\cdot)$ are four-times differentiable with $H(0) = 0$ and $H'(0) > 0$. In addition, there exist positive constants μ_0 and κ_0 such that

$$(1+x)H'(x)\exp\{-H(x)\} \leq \mu_0(1+x)^{-\kappa_0}.$$

Furthermore, there exists a constant $\rho_0 > 0$ such that

$$\sup_x \left\{ \frac{|H''(x)| + |H^{(3)}(x)| + |H^{(4)}(x)|}{H'(x)(1+x)^{\rho_0}} \right\} < \infty,$$

where $H^{(3)}$ and $H^{(4)}$ are the third and fourth derivatives.

- (A6) For any deterministic function $c(t)$ and a constant v such that $c(t) \neq 0$ or $v \neq 0$, $P\{c(t) + v^T Z(t) = 0; t \in [0, \infty]\} = 0$.
- (A7) With some positive probability, $\tilde{Z}_1^T \tilde{Z}_1$ has full rank, where \tilde{Z}_1 denotes a matrix with each row equal to the observed covariate $\tilde{Z}_1(t)^T$ at the time of each measurement.
- (A8) Let K be the number of repeated measures and let d_b be the dimension of b . With probability one, $P(K > d_b | \mathcal{Z}, X) > 0$.

Conditions (A1) - (A3) are the standard assumptions in survival analysis. Condition (A4) is necessary to prove the existence of the NPMLEs. It can be easily verified that Condition (A5) holds for all transformations commonly used, including the logarithmic transformations described in Section 2. Conditions (A6) - (A7) entail the linear independence of design matrices of covariates for the fixed and random effects. Condition (A8) prescribes that some subjects have at least d_b repeated measures.

Under the above conditions, the following theorem shows the consistency of the NPMLEs $(\hat{\theta}, \hat{F})$.

Theorem 1 *Under Conditions (A1) - (A8),*

$$|\hat{\theta} - \theta_0| \rightarrow 0, \quad \sup_{t \in [0, \infty]} |\hat{F}(t) - F_0(t)| \rightarrow 0, \quad a.s.$$

Theorem 1 then leads to the following results on the asymptotic normality of $(\hat{\theta}, \hat{F})$ and the asymptotic efficiency of $\hat{\theta}$.

Theorem 2 *Under Conditions (A1) - (A8), $\sqrt{n}(\hat{\theta} - \theta_0, \hat{F}(t) - F_0(t))$ weakly converges to a zero-mean Gaussian process in $R^{d_\theta} \times BV[0, \infty]$, where d_θ is the dimension of θ and $BV[0, \infty]$ denotes the space of all functions with bounded variations in $[0, \infty]$. Furthermore, the asymptotic covariance matrix of $\sqrt{n}(\hat{\theta} - \theta_0)$ achieves the semiparametric efficiency bound for θ_0 .*

Furthermore, in Appendix, we show that the inverse of the observed information matrix is a consistent estimator of the asymptotic covariance matrix of the NPMLEs. This result allows us to make inference for any functional of $(\theta, F(t))$. To prove Theorems 1-2, we apply the general asymptotic theory of Zeng and Lin (2007). The desired asymptotic properties of the NPMLEs are established followed by the arguments in Appendix B of Zeng and Lin (2007) if we can verify that their regularity conditions hold for our joint cure-survival model setting. Checking the regularity conditions, however, is challenging in our cases. The detailed proofs are provided in Appendix.

5 Simulation Studies

In this section, we demonstrate the finite sample performance of the proposed method through extensive simulation studies. The longitudinal data are generated from

$$Y(t) = 0.7 + z_1 - 0.5z_2 + b + \epsilon(t),$$

and the survival data with a cure proportion are generated from transformation models

$$S(t | z_1, z_2, b) = \exp \left\{ -H \left(e^{0.5z_1 - z_2 + \psi b} F(t) \right) \right\},$$

where z_1 is a dichotomous covariate taking the value of 0 or 1 with the equal probability of 0.5, z_2 is a continuous covariate generated from a uniform distribution on $[-1, 1]$, and $\epsilon(t) \sim N(0, \sigma_\epsilon^2)$ is assumed with $\sigma_\epsilon^2 = 1$. The true failure distribution function in the uncured subpopulation is set to be $F(t) = 1 - \exp(-t)$.

For each subject, the correlation within repeated measures is reflected by the subject-specific random intercept $b \sim N(0, \sigma_b^2)$ with $\sigma_b^2 = 0.5$, and the negative, no, and positive dependences between the longitudinal measures and the cure-survival rate are simulated through different ψ values of -0.3, 0, and 0.3, respectively. For the cure-survival model, we consider three types of transformations $H(\cdot)$ representing the proportional hazards structure ($\eta = 0$), the proportional odds structure ($\eta = 1$), and a transformation in the middle of them with $\eta = 0.5$.

The non-informative censoring time C_i is generated from a uniform distribution with varying rates, depending on the chosen transformation, to design a 30~45% chance of being right-censored and a 20% chance of being cured. We set longitudinal measures to be observed every 0.2 unit of time so that each individual can have about 3 repeated measures, on average.

The results based on 1000 replications are presented in Tables 1-3 for $n=200$ and $n=400$. Tables 1-3 include the average of the differences between the true parameter and the estimates (Bias), the sample standard deviation of the parameter estimators (SE), and the average of the standard error estimators (SEE), and the coverage probability of 95% confidence intervals (CP). The confidence intervals for σ_e^2 and σ_b^2 are constructed based on the Satterthwaite approximation.

Table 1 shows that the NPMLs under the proportional hazards structure $H(x) = x$ are noticeably unbiased, the standard error estimators calculated via the Louis formula well reflect the true variations of the proposed estimators, and the coverage probabilities are in a reasonable range, even with a moderate sample size of 200. As the sample size increases to 400, the biases slightly increase for some estimates; however, they are still very small comparing to the sizes of true parameter values and the variations of the parameter estimators become smaller, and hence the coverage probabilities still lie in a reasonable range. The simulation results shown in Tables 2-3 are similar to those for Table 1, indicating that the proposed method seems to work well for $H(x) = 2\log(1 + x/2)$ and $H(x) = \log(1 + x)$.

6 Data Application

The proposed method was applied to the data from the Health Professionals Follow-up Study (HPFS), a large observational study of male health professionals living in the United States. The main interest of this analysis was to jointly model the relationship between longitudinal vitamin D intake and the survival-cure rate of colorectal cancer (CRC) as endpoints. Since the focus was on the cure of CRC, we restricted our study population to 810 patients who were diagnosed with colorectal cancer between January 1986 and January 2006, without missing at any of the covariates included in the model.

For each subject, vitamin D intakes were assessed via food frequency questionnaires at approximately every four-year intervals between 1986 and 2002. To identify the cure of CRC, we set colorectal cancer-specific death as the event of interest (Ng et al., 2008), while treating deaths from other than CRC or alive until January 2006 as being censored. Then, the cured patients can be defined as a subpopulation among the censored one who have been followed-up sufficiently long enough to be considered cured. In the HPFS during 20 years of follow-up, 250 (31%) colorectal cancer-specific

Table 1: Simulation results for $H(x) = x$. t_p represents the p th percentile.

	<i>True</i>	<i>N</i> = 200				<i>N</i> = 400			
		<i>Bias</i>	<i>SE</i>	<i>SEE</i>	<i>CP</i>	<i>Bias</i>	<i>SE</i>	<i>SEE</i>	<i>CP</i>
$\psi = -0.3$									
α	0.7	-0.009	0.101	0.099	0.944	-0.012	0.073	0.074	0.952
	1.0	0.002	0.143	0.142	0.952	-0.004	0.104	0.103	0.947
	-0.5	0.002	0.125	0.129	0.954	-0.000	0.093	0.096	0.951
σ_e^2	1.0	-0.001	0.064	0.064	0.956	-0.000	0.042	0.041	0.946
	β	0.5	0.009	0.133	0.136	0.951	0.009	0.096	0.094
ψ	-1.0	-0.020	0.165	0.168	0.946	-0.016	0.120	0.118	0.945
	-0.3	-0.010	0.192	0.204	0.968	-0.026	0.148	0.144	0.952
$F(t_{25})$	0.25	-0.005	0.033	0.035	0.949	-0.003	0.025	0.025	0.946
$F(t_{50})$	0.50	-0.004	0.048	0.051	0.954	-0.005	0.035	0.035	0.951
$F(t_{75})$	0.75	-0.003	0.053	0.054	0.942	-0.002	0.036	0.036	0.952
σ_b^2	0.5	-0.008	0.088	0.090	0.963	0.006	0.066	0.064	0.942
$\psi = 0.0$									
α	0.7	-0.008	0.102	0.098	0.936	-0.010	0.070	0.070	0.941
	1.0	-0.002	0.147	0.141	0.943	-0.001	0.099	0.099	0.959
	-0.5	0.002	0.129	0.127	0.942	0.008	0.089	0.090	0.953
σ_e^2	1.0	0.001	0.062	0.064	0.963	-0.002	0.046	0.045	0.947
	β	0.5	0.007	0.130	0.129	0.955	0.004	0.089	0.090
ψ	-1.0	-0.013	0.166	0.160	0.941	-0.009	0.116	0.112	0.935
	0.0	-0.017	0.186	0.190	0.967	-0.018	0.128	0.130	0.952
$F(t_{25})$	0.25	-0.003	0.034	0.034	0.941	-0.002	0.023	0.024	0.945
$F(t_{50})$	0.50	-0.004	0.050	0.049	0.936	-0.002	0.035	0.034	0.944
$F(t_{75})$	0.75	-0.002	0.054	0.052	0.937	-0.001	0.037	0.037	0.948
σ_b^2	0.5	-0.009	0.084	0.088	0.965	-0.003	0.062	0.062	0.961
$\psi = 0.3$									
α	0.7	-0.011	0.098	0.099	0.955	-0.011	0.071	0.070	0.943
	1.0	0.002	0.141	0.141	0.944	-0.001	0.100	0.099	0.944
	-0.5	0.001	0.122	0.129	0.959	0.006	0.094	0.090	0.944
σ_e^2	1.0	0.001	0.062	0.065	0.960	-0.001	0.046	0.046	0.946
	β	0.5	-0.003	0.133	0.135	0.948	-0.003	0.093	0.093
ψ	-1.0	-0.015	0.166	0.168	0.954	-0.002	0.112	0.115	0.958
	0.3	-0.003	0.200	0.202	0.953	-0.023	0.137	0.136	0.955
$F(t_{25})$	0.25	-0.002	0.033	0.035	0.961	-0.000	0.023	0.025	0.953
$F(t_{50})$	0.50	-0.001	0.050	0.051	0.950	-0.000	0.034	0.035	0.955
$F(t_{75})$	0.75	-0.000	0.053	0.053	0.937	0.000	0.037	0.037	0.950
σ_b^2	0.5	-0.011	0.085	0.089	0.966	-0.007	0.060	0.062	0.958

Table 2: Simulation results for $H(x) = 2 \log(1+x/2)$. t_p represents the p th percentile.

	<i>True</i>	<i>N</i> = 200				<i>N</i> = 400			
		<i>Bias</i>	<i>SE</i>	<i>SEE</i>	<i>CP</i>	<i>Bias</i>	<i>SE</i>	<i>SEE</i>	<i>CP</i>
$\psi = -0.3$									
α	0.7	-0.001	0.096	0.099	0.953	-0.008	0.070	0.070	0.944
	1.0	-0.004	0.137	0.138	0.954	-0.002	0.097	0.098	0.942
	-0.5	0.003	0.126	0.125	0.952	0.002	0.087	0.088	0.956
σ_e^2	1.0	0.004	0.063	0.063	0.943	-0.000	0.045	0.044	0.938
	β	0.5	0.009	0.174	0.174	0.944	0.010	0.121	0.121
ψ	-1.0	-0.021	0.209	0.209	0.950	-0.014	0.149	0.146	0.945
	0.3	-0.034	0.286	0.278	0.955	-0.034	0.199	0.192	0.944
$F(t_{25})$	0.25	-0.005	0.036	0.038	0.949	-0.003	0.027	0.026	0.944
$F(t_{50})$	0.50	-0.005	0.054	0.054	0.948	-0.004	0.038	0.038	0.950
$F(t_{75})$	0.75	-0.004	0.059	0.057	0.932	-0.002	0.040	0.040	0.948
σ_b^2	0.5	-0.008	0.091	0.087	0.949	-0.002	0.064	0.061	0.950
$\psi = 0.0$									
α	0.7	-0.008	0.104	0.099	0.930	-0.005	0.070	0.070	0.952
	1.0	-0.005	0.144	0.138	0.950	-0.002	0.099	0.098	0.943
	-0.5	0.011	0.121	0.124	0.954	0.006	0.084	0.088	0.954
σ_e^2	1.0	-0.001	0.064	0.062	0.945	-0.001	0.044	0.044	0.946
	β	0.5	0.005	0.171	0.170	0.945	0.003	0.120	0.118
ψ	-1.0	-0.012	0.211	0.204	0.951	-0.007	0.152	0.143	0.939
	0.0	-0.031	0.286	0.270	0.936	-0.027	0.191	0.188	0.943
$F(t_{25})$	0.25	-0.005	0.035	0.037	0.945	-0.003	0.027	0.026	0.942
$F(t_{50})$	0.50	-0.007	0.051	0.053	0.953	-0.003	0.038	0.038	0.952
$F(t_{75})$	0.75	-0.004	0.056	0.056	0.939	-0.001	0.040	0.040	0.948
σ_b^2	0.5	-0.009	0.082	0.085	0.966	-0.007	0.059	0.060	0.958
$\psi = 0.3$									
α	0.7	-0.005	0.102	0.098	0.943	-0.006	0.067	0.070	0.961
	1.0	-0.005	0.141	0.138	0.936	-0.004	0.095	0.098	0.951
	-0.5	0.006	0.123	0.124	0.951	0.003	0.086	0.088	0.961
σ_e^2	1.0	-0.002	0.062	0.062	0.952	0.000	0.043	0.044	0.952
	β	0.5	0.000	0.169	0.173	0.954	-0.007	0.121	0.120
ψ	-1.0	-0.018	0.212	0.208	0.949	0.001	0.144	0.145	0.958
	0.3	-0.011	0.287	0.275	0.950	-0.028	0.191	0.190	0.946
$F(t_{25})$	0.25	-0.004	0.035	0.037	0.949	0.000	0.027	0.026	0.939
$F(t_{50})$	0.50	-0.002	0.053	0.054	0.950	0.000	0.038	0.038	0.953
$F(t_{75})$	0.75	-0.001	0.057	0.057	0.949	0.001	0.041	0.040	0.951
σ_b^2	0.5	-0.010	0.083	0.086	0.964	-0.004	0.061	0.061	0.957

Table 3: Simulation results for $H(x) = \log(1 + x)$. t_p represents the p th percentile.

	<i>True</i>	<i>N</i> = 200				<i>N</i> = 400			
		<i>Bias</i>	<i>SE</i>	<i>SEE</i>	<i>CP</i>	<i>Bias</i>	<i>SE</i>	<i>SEE</i>	<i>CP</i>
$\psi = -0.3$									
α	0.7	-0.002	0.099	0.098	0.945	-0.002	0.071	0.070	0.942
	1.0	-0.008	0.137	0.136	0.949	-0.005	0.095	0.096	0.958
	-0.5	0.007	0.125	0.122	0.944	0.000	0.088	0.086	0.947
σ_e^2	1.0	-0.003	0.062	0.061	0.952	-0.001	0.043	0.043	0.953
	β	0.5	0.009	0.213	0.207	0.948	0.003	0.139	0.144
ψ	-1.0	-0.030	0.255	0.245	0.938	-0.008	0.175	0.170	0.950
	0.3	-0.048	0.359	0.348	0.943	-0.028	0.254	0.243	0.944
$F(t_{25})$	0.25	-0.006	0.040	0.040	0.939	-0.003	0.028	0.028	0.951
$F(t_{50})$	0.50	-0.006	0.059	0.058	0.941	-0.000	0.041	0.041	0.950
$F(t_{75})$	0.75	-0.003	0.063	0.061	0.940	-0.000	0.044	0.043	0.942
σ_b^2	0.5	-0.008	0.083	0.085	0.965	-0.004	0.063	0.060	0.940
$\psi = 0.0$									
α	0.7	-0.008	0.102	0.099	0.933	-0.004	0.072	0.070	0.952
	1.0	0.000	0.142	0.136	0.947	0.000	0.097	0.097	0.956
	-0.5	0.004	0.122	0.122	0.957	-0.001	0.087	0.086	0.944
σ_e^2	1.0	0.003	0.060	0.061	0.955	-0.001	0.043	0.043	0.946
	β	0.5	0.005	0.205	0.203	0.954	-0.001	0.144	0.142
ψ	-1.0	-0.016	0.239	0.241	0.948	-0.009	0.166	0.168	0.947
	0.0	-0.031	0.359	0.349	0.942	-0.025	0.244	0.240	0.945
$F(t_{25})$	0.25	-0.005	0.041	0.040	0.940	-0.002	0.027	0.028	0.953
$F(t_{50})$	0.50	-0.003	0.057	0.058	0.944	-0.000	0.041	0.041	0.952
$F(t_{75})$	0.75	-0.001	0.061	0.060	0.934	-0.000	0.043	0.043	0.933
σ_b^2	0.5	-0.015	0.085	0.083	0.957	-0.003	0.061	0.060	0.949
$\psi = 0.3$									
α	0.7	-0.001	0.097	0.099	0.947	-0.008	0.069	0.070	0.949
	1.0	-0.008	0.138	0.136	0.939	0.005	0.096	0.096	0.951
	-0.5	-0.001	0.121	0.122	0.958	0.007	0.085	0.086	0.949
σ_e^2	1.0	0.002	0.062	0.061	0.946	-0.001	0.042	0.043	0.959
	β	0.5	-0.007	0.205	0.205	0.956	-0.004	0.146	0.143
ψ	-1.0	-0.019	0.253	0.243	0.941	0.002	0.173	0.170	0.942
	0.3	-0.035	0.352	0.350	0.956	-0.031	0.246	0.243	0.951
$F(t_{25})$	0.25	-0.004	0.039	0.040	0.943	-0.000	0.029	0.028	0.936
$F(t_{50})$	0.50	-0.004	0.058	0.058	0.951	0.000	0.042	0.041	0.949
$F(t_{75})$	0.75	-0.003	0.062	0.061	0.931	-0.000	0.043	0.043	0.942
σ_b^2	0.5	-0.012	0.083	0.084	0.960	-0.007	0.059	0.060	0.950

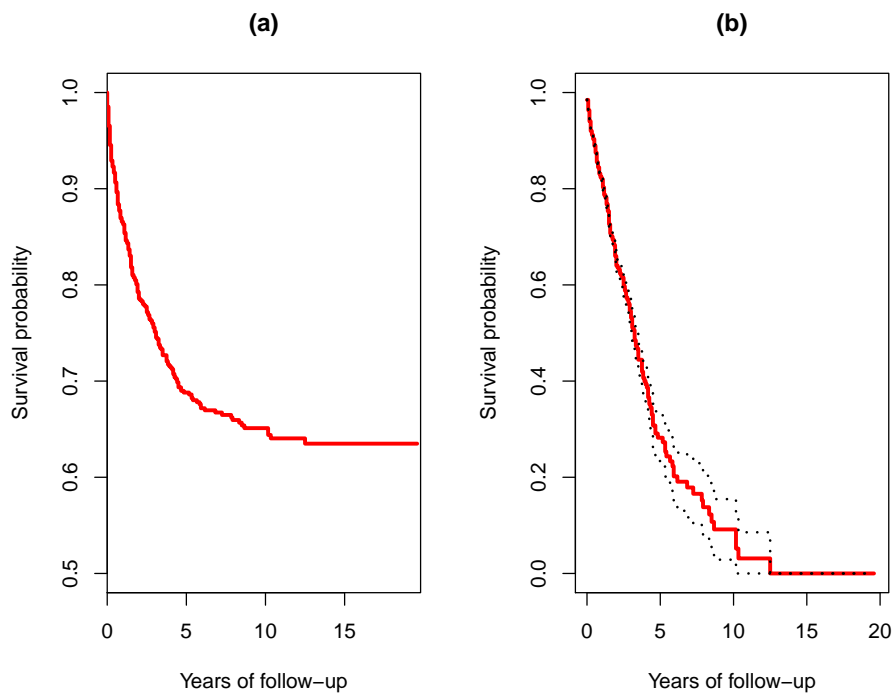


Figure 1: In the HPFS data (N=810) (a) Kaplan-Meier survival curve of the entire study population; (b) Estimated baseline survival function (i.e. $\hat{F}(t)$) of the uncured subpopulation under $H(x) = x$. The solid curves are point estimates, and the dotted curves are 95% confidence intervals.

Table 4: Analysis results for the HPFS study. The 50:50 mixture of χ^2 distributions is used for testing variances.

<i>Effect</i>	Joint Proportional Hazards ($\eta = 0$, AIC= 18008.54)			Joint Proportional Odds ($\eta = 1$, AIC= 18029.07)		
	<i>Estimate</i>	<i>SE</i>	<i>p-value</i>	<i>Estimate</i>	<i>SE</i>	<i>p-value</i>
<i>Longitudinal measures of vitamin D intake</i>						
Intercept	-0.699	0.105	< .001	-0.698	0.105	< .001
Age	0.263	0.088	0.003	0.263	0.088	0.003
BMI	-0.017	0.023	0.453	-0.017	0.023	0.453
Differential grade	0.225	0.175	0.198	0.224	0.176	0.203
Distant metastases	-0.431	0.191	0.024	-0.418	0.192	0.030
σ_e^2	6.115	0.179	< .001	6.116	0.179	< .001
σ_b^2	3.631	0.263	< .001	3.629	0.263	< .001
<i>CRC death</i>						
Age	0.275	0.072	< .001	0.343	0.092	< .001
BMI	-0.010	0.017	0.548	-0.010	0.022	0.652
Differential grade	0.363	0.130	0.005	0.428	0.169	0.012
Distant metastases	2.574	0.130	< .001	3.251	0.172	< .001
ψ	-0.095	0.042	0.026	-0.096	0.055	0.080

deaths were observed. Based on the Kaplan-Meier survival curve in Figure 1 (a), we found that the estimated survival rate at the end of study was very high (64%) even after a sufficient follow-up period (i.e. 20 years), and the earliest point in time that the curve goes flat, 12.5 year of follow-up, was the point at which all remaining disease-free survivors were declared to be cured. There were 116 patients who has been considered cured in the HPFS. We note that some of patients who were right-censored before 12.5 years might indeed have been cured to CRC, but it was inconclusive due to the right-censoring.

We fitted the proposed joint cure model for the longitudinal vitamin D trend and CRC death with the patient's medical information at diagnosis; age, body mass index (BMI), and indicators for tumor differentiation grade (i.e. 1=poor or unspecified, 0=well or moderate) and distant metastases (i.e. 1=yes, 0=no) were included as covariates. Among them, age variable was centered at mean 68 and divided by 10 to represent a decade, and BMI was centered at mean $25kg/m^2$. In addition, a subject-specific random intercept was included in both longitudinal and cure-survival models to account for the correlation between these two outcomes. To explore the possibility of the proportional hazards and the proportional odds structures in cure-survival data, we also applied transformation models $H(x) = \log(1 + \eta)/\eta$ to cure-survival data by varying η values in $[0, 1]$ at every 0.1 increment. We used the Akaike information criterion (AIC) to determine the best form of transformation (i.e. η), and the smallest AIC value was achieved at $\eta = 0$, implying the joint proportional hazards cure (PHC) model was the best fit to the data. Although in the HPFS example the final transformation turned out to be the joint PHC model, in the lack of model-diagnostic tools for the joint modeling, it is valuable to consider transformations to confirm the fit is the best among the class of transformations we consider. To show the impact of transformation on the parameter estimates, Table 4 summarizes the analysis results under the joint PHC model ($\eta = 0$) and the joint proportional odds cure model ($\eta = 1$).

Under the selected best transformation model, Figure 1 (b) displays the estimated baseline survival distribution for the uncured patients ($X_i < \infty$) along with their point-wise 95% confidence intervals. In Figure 1 (b) we note that the tail probability of the estimated baseline survival curve reached zero. The results in Table 4 show that 1) older patients tended to take more vitamin D, and were more likely to be uncured to CRC; and 2) patients with distant metastases appeared to take vitamin D less, and were more likely to be uncured to CRC. The significant negative $\hat{\psi}$ suggested that there was a protective effect of vitamin D intake in relation to the risk of CRC death, which was not explained by the common covariates in both longitudinal and survival components ($p = 0.026$). As an example of quantitative interpretation, the marginal survival rates (≤ 12.5 years) and cure rates (> 12.5 years) for the whole population are given in Figure 2. For instance, when comparing the curves to reference (age of 68, BMI = $25kg/m^2$, well or moderate differentiation, no distant metastases), we can see that the cure rate at age of 78 decreased to 68% from 80% at age of 68, while one for CRC patients with distance metastases decreased to 6% from 80% with no distance metastases. The curves have been obtained by $E_b[S(t|z_2, b)] = E_b[\exp\{-H(e^{\beta^T z_2 + \psi b} F(t))\}]$ evaluated at the NPMLEs for a given covariate z_2 , and their 95% pointwise confidence intervals can also be obtained by applying the functional delta method and evaluating at the NPMLEs.

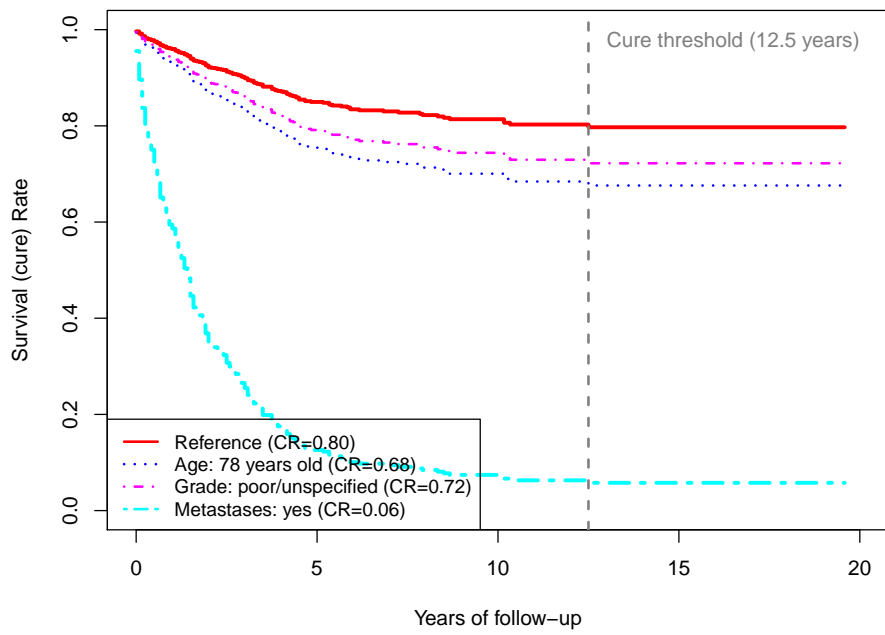


Figure 2: Predicted marginal survival rates of the entire population using the results in Table 4. The rates beyond the cure threshold are interpreted as cure rates (CR). Reference rate is taken for age of 68, BMI = $25\text{kg}/\text{m}^2$, well or moderate differentiation, and no distant metastasis.

7 Concluding Remarks

We have proposed the joint transformation model for longitudinal and survival data which takes the possibility of patients being cured or immune to disease into account. The proposed approach has the advantages of handling time-varying covariates and providing an easier way to explore a large class of cure models in a unified way. We have used the NPMLEs for estimating the model parameters, and the resulting NPMLEs have been shown to be asymptotically normal and efficient. Simulation studies have showed that the proposed estimation procedures produced consistent estimators, and the new EM algorithm enabled to compute the NPMLEs in a simpler and more stable way.

As an example of $H(\cdot)$, we considered a class of logarithmic transformations, which can be misspecified in practice because of limited knowledge or complex relationships between covariates. As an alternative choice, we also explored the performance of a class of Box-Cox transformations,

$$H(x) = \begin{cases} \{(1+x)^\gamma - 1\}/\gamma, & \gamma > 0 \\ \log(1+x), & \gamma = 0, \end{cases}$$

and the selected transformation function was robust to the class of transformations considered. Based on our experiences, it appears that the form of transformations is less important than the problem of selecting the transformation parameter η . We used the AIC to determine the best transformation parameter, but there exist other criteria for model selection such as the Bayes information criterion and cross-validation ('leave-one-subject-out'). The differentiability conditions on H , as in the first part of the Condition (A5), are satisfied with any class of transformations induced by a random effect. Indeed, the validity of our asymptotic properties proven here is not restricted to these frailty class transformations. Other transformations which are not generated by a frailty, for instance, the Box-Cox transformations with $\gamma > 1$, can also satisfy the Condition (A.5). We further note that in this article the frailty representation relating to the Laplace transformation has been introduced to facilitate easy use of EM computation.

In this paper, we assumed that the number of observations of repeated measures are independent of cure-survival data. To account for the informative observation times, our joint cure model can be extended by jointly modeling another recurrent event process. Another promising extension of our joint cure model would be to the context of generalized linear mixed models (GLM) to analyze discrete longitudinal outcomes. It is rather obvious that the general approach presented here is still applicable to GLMs, but some specific parts related to estimation procedures of longitudinal components need to be modified accordingly.

Acknowledgements

The authors would like to thank the referees for helpful comments and gratefully acknowledge use of the Health Professionals Follow-up Study data, funded by NIH/NCI grant P01 CA055075.

References

- Berkson J., Gage R., 1952. Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47, 501–515.
- Brown E., Ibrahim J., 2003. Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. *Biometrics*, 59, 686–693.
- Chen M., Ibrahim J., Sinha D., 1999. A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94, 909–919.
- Chen M., Ibrahim J., Sinha D., 2004. A new joint model for longitudinal and survival data with a cure fraction. *Journal of Multivariate Analysis*, 91, 18–34.
- Kuk A., Chen C., 1992. A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79, 531–541.
- Laska E., Meisner M., 1992. Nonparametric estimation and testing in a cure model. *Biometrics*, 48, 1223–1234.
- Law N., Taylor J., Sandler H., 2002. The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics*, 3, 547–563.
- Louis T., 1982. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 44, 226–233.
- Lu W., Ying Z., 2004. On semiparametric transformation cure models. *Biometrika*, 91, 331–343.
- Maller R., Zhou X., 1996. *Survival analysis with long-term survivors*. Wiley Chichester.
- Ng K., Meyerhardt J., Wu K., Feskanich D., Hollis B., Giovannucci E., Fuchs C., 2008. Circulating 25-hydroxyvitamin d levels and survival in patients with colorectal cancer. *Journal of Clinical Oncology*, 26, 2984–2991.
- Sy J., Taylor J., 2000. Estimation in a Cox proportional hazards cure model. *Biometrics*, 56, 227–236.
- Tsodikov A., 1998. A proportional hazards model taking account of long-term survivors. *Biometrics*, 54, 1508–1516.
- Tsodikov A., 2002. Semiparametric models of long- and short-term survival: an application to the analysis of breast cancer survival in utah by age and stage. *Statistics in medicine*, 21, 895–920.
- Tsodikov A., Ibrahim J., Yakovlev A., 2003. Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, 98, 1063–1078.

- Yakovlev A., Tsodikov A., Asselain B., 1996. *Stochastic models of tumor latency and their biostatistical applications*. New Jersey: World Scientific.
- Yu M., Law N., Taylor J., Sandler H., 2004. Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, 14, 835–862.
- Yu M., Taylor J., Sandler H., 2008. Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model. *Journal of the American Statistical Association*, 103, 178–187.
- Zeng D., Lin D.Y., 2007. Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society, Series B*, 69, 507–564.
- Zeng D., Yin G., Ibrahim J., 2006. Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*, 101, 670–684.

Appendix: Proofs of Asymptotic Properties

This section proves Theorems 1-2 stated in Section 4 by applying the general asymptotic theory of Zeng and Lin (2007). Specifically, it is easy to see that our conditions (A1) - (A8) imply (C1) - (C4), (C6), (C8) of Zeng and Lin (2007), and it remains to prove the two identifiability conditions (C5) and (C7) of Zeng and Lin (2007). The first identifiability is the key step to prove the consistency of the NPMLEs, and the second is to entail the invertibility of the observed information matrix at the true parameters for the proof of the asymptotic normality.

Proof 1 First, we verify the first identifiability condition (C5) in Appendix B of Zeng and Lin (2007). Suppose that the likelihood function for $(\alpha, \beta, \psi, \sigma_e^2, \text{Vec}(\Sigma_b))$ is the same as that for the true parameter values $(\alpha_0, \beta_0, \psi_0, \sigma_{0e}^2, \text{Vec}(\Sigma_{0b}))$. That is, for arbitrary $K > 0$,

$$\begin{aligned}
& \int_b (2\pi\sigma_e^2)^{-K/2} \exp \left\{ -\frac{(\mathbf{Y} - \mathbf{Z}_1\alpha - \tilde{\mathbf{Z}}_1b)^T (\mathbf{Y} - \mathbf{Z}_1\alpha - \tilde{\mathbf{Z}}_1b)}{2\sigma_e^2} \right\} \\
& \quad \times \left[f(x) e^{\beta^T Z_2(x) + (b \circ \psi)^T \tilde{Z}_2(x)} H'(q(x)) \right]^\Delta e^{-H(q(x))} f(b; \Sigma_b) db \\
= & \int_b (2\pi\sigma_{0e}^2)^{-K/2} \exp \left\{ -\frac{(\mathbf{Y} - \mathbf{Z}_1\alpha_0 - \tilde{\mathbf{Z}}_1b)^T (\mathbf{Y} - \mathbf{Z}_1\alpha_0 - \tilde{\mathbf{Z}}_1b)}{2\sigma_{0e}^2} \right\} \\
& \quad \times \left[f_0(x) e^{\beta_0^T Z_2(x) + (b \circ \psi_0)^T \tilde{Z}_2(x)} H'(q_0(x)) \right]^\Delta e^{-H(q_0(x))} f(b; \Sigma_{0b}) db, \quad (10)
\end{aligned}$$

where bold \mathbf{Y} denotes the vector of the observed longitudinal measures at time s_1, \dots, s_K , and \mathbf{Z}_1 and $\tilde{\mathbf{Z}}_1$ in bold type denote matrices with each row equal to the observed covariate $Z_1(s_k)^T$ and $\tilde{Z}_1(s_k)^T$ at $k = 1, \dots, K$, respectively. In addition, $q(t) =$

$\int_0^t e^{\beta^T Z_2(u) + (b \circ \psi)^T \tilde{Z}_2(u)} dF(u)$, and $q_0(t)$ is $q(t)$ evaluated at the true parameter values, and $f(b; \Sigma_b)$ is the density function of the (multivariate) normal distribution with mean zeros and covariance matrix Σ_b . From now, we take the following actions on both sides of (10).

Step 1: For the proof of the identifiability of the longitudinal component, we consider a case $\Delta = 0$ and $X \approx 0$.

Using the fact that $\int b f(b; \Sigma_b) db = \int b f(b; \Sigma_{0b}) db = 0$ and considering $E[Y(s_k)]$ conditional on b , we have $\alpha^T Z_1(s_k) = \alpha_0^T Z_1(s_k)$, for $k = 1, \dots, K$. By Condition (A6), we prove $\alpha = \alpha_0$. Similarly, we consider $E[Y(s_k)Y(s_{k'})]$ and $\text{Var}(Y(s_k))$, given b , and obtain for $k \neq k'$

$$\begin{aligned} & \int_b \left\{ \alpha_0^T Z_1(s_k) + b^T \tilde{Z}_1(s_k) \right\} \left\{ \alpha_0^T Z_1(s_{k'}) + b^T \tilde{Z}_1(s_{k'}) \right\} f(b; \Sigma_b) db \\ &= \int_b \left\{ \alpha_0^T Z_1(s_k) + b^T \tilde{Z}_1(s_k) \right\} \left\{ \alpha_0^T Z_1(s_{k'}) + b^T \tilde{Z}_1(s_{k'}) \right\} f(b; \Sigma_{0b}) db, \end{aligned}$$

followed by the proof of $\Sigma_b = \Sigma_{0b}$ from (A6), and

$$\begin{aligned} & \int_b \left\{ \sigma_e^2 + b^T \tilde{Z}_1(s_k) \tilde{Z}_1(s_k)^T b \right\} f(b; \Sigma_b) db \\ &= \int_b \left\{ \sigma_{0e}^2 + b^T \tilde{Z}_1(s_k) \tilde{Z}_1(s_k)^T b \right\} f(b; \Sigma_{0b}) db, \end{aligned}$$

for $k = 1, \dots, K$. Accordingly, we have that $\sigma_e^2 = \sigma_{0e}^2$.

Step 2: For the survival component, suppose $\Delta = 0$ and $X = t$. Then, (10) implies

$$E_b [\exp \{-H(q(t))\}] = E_b [\exp \{-H(q_0(t))\}],$$

where b follows a normal distribution with mean $\mu_b = V_b \tilde{\mathbf{Z}}_1^T (\mathbf{Y} - \mathbf{Z}_1 \alpha_0) / \sigma_{0e}^2$ and covariance matrix $V_b = [\Sigma_{0b}^{-1} + \tilde{\mathbf{Z}}_1^T \tilde{\mathbf{Z}}_1 / \sigma_{0e}^2]^{-1}$. For fixed \mathbf{Y} , \mathbf{Z}_1 , and $\tilde{\mathbf{Z}}_1$, since b is the complete statistic for μ_b , we can have that

$$\begin{aligned} & \exp \left\{ -H \left(\int_0^t e^{\beta^T Z_2(u) + (b \circ \psi)^T \tilde{Z}_2(u)} dF(u) \right) \right\} \\ &= \exp \left\{ -H \left(\int_0^t e^{\beta_0^T Z_2(u) + (b \circ \psi_0)^T \tilde{Z}_2(u)} dF_0(u) \right) \right\}. \end{aligned}$$

Furthermore, it is followed from the one-to-one mapping of H and exponential function that

$$\log(f(t)) + \beta^T Z_2(t) + b^T (\psi \circ \tilde{Z}_2(t)) = \log(f_0(t)) + \beta_0^T Z_2(t) + b^T (\psi_0 \circ \tilde{Z}_2(t)),$$

with probability 1. By taking the expectation with respect to b for fixed \mathbf{Y} , \mathbf{Z}_1 , and $\tilde{\mathbf{Z}}_1$, we conclude that $\beta = \beta_0$, $f(t) = f_0(t)$ and $\psi = \psi_0$ from the Condition (A6).

Proof 2 Next, we verify the second identifiability condition (C7) in Appendix B of Zeng and Lin (2007). It starts from the score equation along with the path $(\alpha_0 + \xi\nu_1, \beta_0 + \xi\nu_2, \psi_0 + \xi\nu_3, \sigma_{0e}^2 + \xi\nu_4, \text{Vec}(\Sigma_{0b}) + \xi\nu_b, F_0 + \xi \int h dF_0)$. We define D_b as the symmetric matrix such that $\text{Vec}(D_b) = \nu_b$.

Step 1: To make the score equation simple for the proofs of $\nu_1 = 0$, $\nu_4 = 0$ and $D_b = 0$, we consider the same case $\Delta = 0$ and $X \approx 0$ as used in Step 1 of the first identifiability proof. We define

$$V_b^{-1} = \Sigma_{0b}^{-1} + \tilde{\mathbf{Z}}_1^T \tilde{\mathbf{Z}}_1 / \sigma_{0e}^2, \quad \text{and} \quad \mu_b = V_b \tilde{\mathbf{Z}}_1^T (\mathbf{Y} - \mathbf{Z}_1 \alpha_0) / \sigma_{0e}^2,$$

then, the score equation is given by

$$\begin{aligned} 0 = & -\frac{1}{2} \text{Tr}(\Sigma_{0b}^{-1} D_b) + \frac{1}{2} \mu_b^T \Sigma_{0b}^{-1} D_b \Sigma_{0b}^{-1} \mu_b + \frac{1}{2} \text{Tr}(\Sigma_{0b}^{-1} D_b \Sigma_{0b}^{-1} V_b) + \frac{\nu_4}{2\sigma_{0e}^4} \text{Tr}(\tilde{\mathbf{Z}}_1^T \tilde{\mathbf{Z}}_1 V_b) \\ & - \frac{\nu_4 K}{2\sigma_{0e}^2} + \frac{\nu_1^T \tilde{\mathbf{Z}}_1^T (\mathbf{Y} - \mathbf{Z}_1 \alpha_0 - \tilde{\mathbf{Z}}_1 \mu_b)}{\sigma_{0e}^2} \\ & + \frac{\nu_4}{2\sigma_{0e}^4} \left\{ (\mathbf{Y} - \mathbf{Z}_1 \alpha_0)^T (\mathbf{Y} - \mathbf{Z}_1 \alpha_0) - 2(\mathbf{Y} - \mathbf{Z}_1 \alpha_0)^T \tilde{\mathbf{Z}}_1 \mu_b + \mu_b^T \tilde{\mathbf{Z}}_1^T \tilde{\mathbf{Z}}_1 \mu_b \right\}. \end{aligned} \quad (11)$$

By comparing coefficients for the constant, linear and quadratic terms of $(\mathbf{Y} - \mathbf{Z}_1 \alpha_0)$, we have that

$$0 = \frac{\nu_4}{2\sigma_{0e}^2} \left[K - \frac{\text{Tr}(\tilde{\mathbf{Z}}_1^T \tilde{\mathbf{Z}}_1 V_b)}{\sigma_{0e}^2} \right] + \frac{1}{2} \text{Tr}(\Sigma_{0b}^{-1} D_b) - \frac{1}{2} \text{Tr}(\Sigma_{0b}^{-1} D_b \Sigma_{0b}^{-1} V_b), \quad (12)$$

$$0 = \frac{\nu_1^T \tilde{\mathbf{Z}}_1^T}{\sigma_{0e}^2} \left[I - \frac{\tilde{\mathbf{Z}}_1 V_b \tilde{\mathbf{Z}}_1^T}{\sigma_{0e}^2} \right], \quad (13)$$

$$0 = \frac{\nu_4}{2\sigma_{0e}^4} \left[I - \frac{2\tilde{\mathbf{Z}}_1 V_b \tilde{\mathbf{Z}}_1^T}{\sigma_{0e}^2} + \frac{\tilde{\mathbf{Z}}_1 V_b \tilde{\mathbf{Z}}_1^T \tilde{\mathbf{Z}}_1 V_b \tilde{\mathbf{Z}}_1^T}{\sigma_{0e}^4} \right] + \frac{\tilde{\mathbf{Z}}_1 V_b \Sigma_{0b}^{-1} D_b \Sigma_{0b}^{-1} V_b \tilde{\mathbf{Z}}_1^T}{2\sigma_{0e}^4} \quad (14)$$

Since $[I - \tilde{\mathbf{Z}}_1 V_b \tilde{\mathbf{Z}}_1^T / \sigma_{0e}^2]$ is positive definite, we can see that $\nu_1 = 0$ in (13). To simplify (14), we multiply $\tilde{\mathbf{Z}}_1^T$ from the left, $\tilde{\mathbf{Z}}_1$ from the right, and then $[\tilde{\mathbf{Z}}_1^T \tilde{\mathbf{Z}}_1]^{-1}$ from the right on both sides of (14). Using the fact that $\Sigma_{0b}^{-1} D_b = I - \tilde{\mathbf{Z}}_1^T \tilde{\mathbf{Z}}_1 V_b / \sigma_{0e}^2$, the equation (14) becomes

$$\frac{\nu_4}{2\sigma_{0e}^2} \left[I - \frac{\tilde{\mathbf{Z}}_1^T \tilde{\mathbf{Z}}_1 V_b}{\sigma_{0e}^2} \right] + \frac{\tilde{\mathbf{Z}}_1^T \tilde{\mathbf{Z}}_1 V_b \Sigma_{0b}^{-1} D_b}{2\sigma_{0e}^2} = 0, \quad (15)$$

and the equation (12) becomes

$$\frac{\nu_4}{2\sigma_{0e}^2} \left[K - \frac{\text{Tr}(\tilde{\mathbf{Z}}_1^T \tilde{\mathbf{Z}}_1 V_b)}{\sigma_{0e}^2} \right] + \frac{1}{2\sigma_{0e}^2} \text{Tr}(\tilde{\mathbf{Z}}_1^T \tilde{\mathbf{Z}}_1 V_b \Sigma_{0b}^{-1} D_b) = 0. \quad (16)$$

After taking the trace of (15) and subtracting from the equation (16), we obtain that

$$\frac{\nu_4}{2\sigma_{0e}^2} (K - d_b) = 0,$$

where d_b stands for the dimension of b . Based on Condition (A8), we conclude that $\nu_4 = 0$, and hence $D_b = 0$ in (15) by Condition (A7).

Step 2: For the second identifiability of the survival component, we set $\Delta = 0$ and $\bar{X} = t$. Then, the score equation can be written as

$$E_b [\exp\{-H(q_0(t))\} H(q_0(t)) \dot{q}_0(t)] = 0, \quad (17)$$

where $\dot{q}_0(t) = \int_0^t \{h(u) + \nu_2^T Z_2(u) + (\nu_3 \circ b)^T \tilde{Z}_2(u)\} e^{\beta_0^T Z_2(u) + (b \circ \psi_0)^T \tilde{Z}_2(u)} dF_0(u)$, and b is normally distributed with mean μ_b and covariance matrix V_b . By the completeness of the exponential family of b , we can have

$$\exp\{-H(q_0(t))\} H(q_0(t)) \dot{q}_0(t) = 0,$$

for any fixed \mathbf{Y} , \mathbf{Z}_1 and $\tilde{\mathbf{Z}}_1$ with probability 1. Since $H(q_0(t)) > 0$ for $\forall t > 0$ from (A5), we can obtain $\dot{q}_0(t) = 0$, and hence

$$h(t) + \nu_2^T Z_2(t) + (\nu_3 \circ b)^T \tilde{Z}_2(t) = 0.$$

Clearly, we attain $\nu_2 = 0$, $\nu_3 = 0$ and $h = 0$ by (A6).

Finally, we complete the proofs of Theorems 1 - 2 by Theorems 1 - 2 in Zeng and Lin (2007). Let I_n denote the negative Hessian matrix of the observed log-likelihood function with respect to $(\theta, F\{\cdot\})$. As a remark, by following Theorem 3 in Zeng and Lin (2007), we can show that I_n is invertible for large n , and $(\nu^T, U^T) n I_n^{-1} (\nu^T, U^T)^T$ is the consistent estimator of the asymptotic variance of

$$\sqrt{n} \left\{ \nu^T (\hat{\theta} - \theta_0) + \int u(t) d(\hat{F} - F_0) \right\},$$

where U is the vector of $u(\cdot)$ at the observed failure times.