Regression Analysis for Recurrent Events Data under Dependent Censoring

Jin-Jian Hsieh,1∗ A. Adam Ding,2∗∗ and Weijing Wang3∗∗∗

1Department of Mathematics, National Chung Cheng University, Chia-Yi, Taiwan, Republic of China
2Department of Math, Northeastern University, Boston, Massachusetts 02115, U.S.A.
3Institute of Statistics, National Chiao-Tung University, Hsin-Chu, Taiwan, Republic of China
∗email: jjhsieh@math.ccu.edu.tw
∗∗email: ding@neu.edu
∗∗∗email: wjwang@stat.nctu.edu.tw

Summary. Recurrent events data are commonly seen in longitudinal follow-up studies. Dependent censoring often occurs due to death or exclusion from the study related to the disease process. In this article, we assume flexible marginal regression models on the recurrence process and the dependent censoring time without specifying their dependence structure. The proposed model generalizes the approach by Ghosh and Lin (2003, Biometrics 59, 877–885). The technique of artificial censoring provides a way to maintain the homogeneity of the hypothetical error variables under dependent censoring. Here we propose to apply this technique to two Gehan-type statistics. One considers only order information for pairs whereas the other utilizes additional information of observed censoring times available for recurrence data. A model-checking procedure is also proposed to assess the adequacy of the fitted model. The proposed estimators have good asymptotic properties. Their finite-sample performances are examined via simulations. Finally, the proposed methods are applied to analyze the AIDS linked to the intravenous experiences cohort data.

Key words: Artificial censoring; Dependent censoring; Longitudinal study; Multiple events; Pairwise comparison; Recurrent event data; Survival analysis; U-statistics.

1. Introduction

Multiple-events data are commonly seen in medical applications. Recurrence data are a special type of multiple-events data in which the same type of events may occur more than once. Examples include sequences of asthmatic attacks, bleeding incidents, epileptic seizures, infection episodes, tumor recurrences, or hospitalization care, just to name a few. A number of recurrent events models have appeared in the literature. The majority of analyses assume independent censoring. Existing approaches differ in the chosen model quantities and whether the recurrence history is incorporated into the model. Another important aspect is how covariates affect the model quantity. The framework of counting processes has been adopted by many authors. Andersen and Gill (1982); Prentice, Williams, and Peterson (1981); Pepe and Cai (1993); Chang and Wang (1999); and Zeng and Lin (2007a) assumed that the intensity rate of the recurrence process is proportionally affected by covariates. The mean function was modeled by Lawless and Nadeau (1995) assuming multiplicative covariate effects and by Lin, Wei, and Ying (1998) assuming a time transformation that corresponds to the accelerated failure time (AFT) model. Another alternative is to model the gap times between adjacent recurrences. For example, Huang (2000) considered an AFT model on the gap times and Schaubel and Cai (2004) proposed a proportional hazards (PH) model. It is worth mentioning that, even under the independent censorship, the second and subsequent gap times are subject to induced dependent censoring.

In practical applications, a recurrence process may be censored by two types of events. One type of censoring happens when the study period ends or a patient withdraws from the study for reasons unrelated to the disease status. The other type of censoring occurs when a patient dies or is excluded from the study due to biological reasons related to the disease process. How to handle the association between the recurrence process and the dependent censoring event is the key feature of this research, which has received increasing attention in recent years. The frailty approach has been adopted by Wang, Qin, and Chiang (2001); Huang and Wang (2004); Liu, Wolfe, and Huang (2004); Miloslavsky et al. (2004); Ye, Kalbfleisch, and Schaubel (2007); and Zeng and Lin (2007b). The above papers differ in the chosen model quantity for the recurrence processes (i.e., intensity, mean, or occurrence rate functions) and that for the dependent censoring variable (i.e., hazard or the failure time). Also, the mechanism of how the latent frailty variable and observed covariates affect the model quantities may be different. For example, Huang and Wang (2004) assumed multiplicative frailty and covariate effects on the intensity function of the recurrent process and on the hazard of the dependent censoring variable. Cook and Lawless (1997) proposed to modify the occurrence rate function, which also incorporates the effect of dependent censoring. On the other
hand, some authors proposed inference methods for assessing marginal covariate effects without specifying the underlying dependence structure. For example, Chang (2000) assumed AFT models on the gap time between recurrences and dependent censoring time. Ghosh and Lin (2003) assumed AFT models on the total time (measured from the beginning to a recurrence event) and dependent censoring time.

In this article, we also focus on marginal regression analysis without specifying the form of the dependent censoring. We broaden the approach of Ghosh and Lin (2003) to allow more flexible modeling of the recurrence process and the dependent censoring time. Notations and model assumptions are described in Section 2. Section 3 presents the proposed methods and asymptotic analysis. A model-checking procedure for assessing the goodness of fit of the imposed assumption and a related model-selection procedure are discussed in Section 4. Simulation results and data analysis are provided in Sections 5 and 6, respectively. Section 7 contains some concluding remarks.

2. Notations and Models

Let \( N^*(t) \) be the number of recurrent events that occur over the time interval \([0, t]\). \( D \) be the dependent censoring time, \( C \) be the independent censoring time, and \( Z \) be the vector of covariates. Throughout the article, we assume that \( C \) may depend on \( Z \) but \( C \) is independent of \( N^*(t) \) and \( D \) given \( Z \). On the other hand, \( D \) is correlated with the process \( N^*(t) \), despite \( Z \). The recurrence process can also be expressed in terms of recurrence times. Define \( T_k \) as the time from the origin to the \( k \)-th recurrent event \((k = 1, 2, \ldots)\). Hence \( N^*(t) = \sum_{k=1}^{\infty} I(T_k \leq t) \). Observed variables can be denoted as \( N(t) = N^*(t \wedge D \wedge C), X = D \wedge C, \) and \( \delta = I(D \leq C) \). There are \( K = N(X) \) events observed successively at times \( T_k = \min \{t: N(t) \geq k\} \) for \( k = 1, \ldots, K \).

In this article, the effect of covariates on \( N^*(t) \) is the major interest and that on \( D \) is of secondary interest, and the dependence between \( \{N^*(t), D\} \) is nuisance. Accordingly we assume the following regression models:

\[
\begin{align*}
    h_1(T_k) &= \beta_0^* Z + \varepsilon_k, \\
    h_2(D) &= \eta_0^* Z + \xi,
\end{align*}
\]

(1)

where \( h_1(\cdot) \) is a known monotone function, \( h_2(\cdot) \) is a monotone function that may be known or unknown, \( \theta_0 = (\beta_0^*, \eta_0^*) \) are the true parameters, \( \varepsilon_k \) and \( \xi \) are the error variables, \( \beta_0 \) and \( \xi \) may be correlated with each other but both are independent of \( Z \). The marginal distribution of \( \varepsilon_k \) is unspecified. The marginal distribution of \( \xi \) is unspecified for known \( h_2(\cdot) \) and is specified for unknown \( h_2(\cdot) \). When \( h_1(t) = h_2(t) = \log(t) \) (i.e., an AFT assumption for both recurrences and dependent censoring), the models in (1) reduce to the case studied by Ghosh and Lin (2003). This setting allows for more flexible assumptions on \( D \), say the PH model with \( h_2(t) \) unknown and \( \xi \) following the extreme value distribution.

The first model can also be expressed in terms of counting processes. Define \( N_i^*(t) = \sum_{k=1}^{\infty} I(\varepsilon_k \leq t) \) as the number of occurrences for \( \varepsilon_k \) over the time interval \([0, t] \). The models are equivalent to the following

\[
\begin{align*}
    \left[ N^*(h_1^{-1}(t + \beta_0^* Z)) \right] &\overset{d}{=} \left[ N^*_i(t) \right] \\
    h_2(D) &= \eta_0^* Z
\end{align*}
\]

(2)

where \( \overset{d}{=} \) means “with the same distribution.”

Estimation of \( \eta_0 \) is straightforward by applying existing methods that assume independent censorship. Estimation of \( \beta_0 \) is complicated due to dependent censoring. To avoid specifying the dependent relationship, we propose to modify non-parametric statistics originally constructed based on random replicates of \( \{N^*_i(t), \xi\} \).

3. The Proposed Inference Methods

Let \( \{N^*_i(t), T_k, C_i, D, Z_i\} \) be independent realizations of \( \{N^*(t), T_k, C, D, Z\} \) for \( k = 1, \ldots, i = 1, 2, \ldots, n \). Let \( \varepsilon_{ik}(\beta) = h_1(T_{ik}) - \beta^* Z_{ik} \), and \( \xi(\eta) = h_2(D_i) - \eta(Z_i) \). It is easy to see that \( \varepsilon_{ik}(\beta, \xi(\eta)) \) for \( i = 1, \ldots, n \) are independent and identical replications of \( \varepsilon_k(\beta, \xi) \). Define \( N_{ij}(t) = N^*[h_1^{-1}(t + \beta^* Z_{ij})] \). Accordingly \( \{N_{ij}(t), \xi_i(\eta)(i = 1, \ldots, n)\} \) constitute a bivariate random sample with the joint distribution independent of \( Z \). In the presence of censoring, the observed counting process becomes \( N_i(t) = N^*(t \wedge X_i) \), which can also be represented in terms of the error scale by defining \( N_{ij}(t) = N[h_1^{-1}(t + \beta^* Z_{ij})] \). Note that \( N_{ij}(t) \) and \( N_{ij}(t) \) no longer have the same distribution for \( Z_i \neq Z_j \) due to the dependence between \( \varepsilon_{ik} \) and \( \xi_i \).

3.1 Estimation of \( \theta_0 \) When \( h_2(\cdot) \) Is Known

We suggest estimating \( \theta_0 = (\eta_0^*, \beta_0^*) \) by first estimating \( \eta_0 \) and then \( \beta_0 \). Estimation of the latter is the challenge. To motivate the proposed idea, we first discuss estimation of \( \eta_0 \) when \( h_2(\cdot) \) is known. Define \( \xi_i(\eta) = h_2(X_i) - \eta Z_i \). When \( \delta_i = 1, \xi_i(\eta) = \xi_i(\eta) \); whereas when \( \delta_i = 0, \xi_i(\eta) = \xi_i(\eta) = h_2(C_i) - \eta Z_i \). Two statistics can be applied. The first one is the log-rank type statistic:

\[
U^L_i(\eta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\infty \left[ \sum_{i=1}^{n} I(\hat{\xi}_i(\eta) \geq t) Z_i \right] dN_{ci}(t; \eta),
\]

(3)

where \( N_{ci}(t; \eta) = I(\hat{\xi}_i(\eta) \leq t, \delta_i = 1) \). The other is the Gehan-type statistics:

\[
U^G_i(\eta) = 2\sqrt{n} \sum_{1 \leq i < j \leq n} \frac{(Z_i - Z_j) \Phi_{ij}(\eta)}{n(n-1)},
\]

(4)

where

\[
\Phi_{ij}(\eta) = I(\hat{\xi}_i(\eta) \leq \hat{\xi}_j(\eta), \delta_i = 1) - I(\hat{\xi}_i(\eta) \leq \hat{\xi}_j(\eta), \delta_j = 1).
\]

(5)

Note that \( \Phi_{ij}(\eta) \) and \( \Phi_{ji}(\eta) \) have the same distribution for \( i \neq j \). Estimators of \( \eta \) are the zero-crossing points of the corresponding estimating functions.

It is well known that the Gehan statistics can be expressed as weighted log-rank statistics, which can further be represented as a martingale integral asymptotically. In Appendix A, we explore the other direction by writing the log-rank statistics in terms of pairwise notations. The new expression
will provide us some insight for estimating $\beta$ in the presence of dependent censoring.

Under the error scale, $\varepsilon_{ik}$ is subject to censoring by $\varepsilon_{ik}^C = h_1[h_2^{-1}((\xi, \xi^C) + \eta_z^C, Z)] - \beta_z Z$. As a result, $(\varepsilon_{ik}, \varepsilon_{ik}^C)$ no longer have the same distribution for $Z_i \neq Z_j$, due to dependence between $\varepsilon_{ik}$ and $\xi$. The technique of artificial censoring has been adopted by several authors to remove the bias arising from dependent censoring. It provides a way to create the homogeneity for observations under comparison. Now we illustrate how this idea is applied to the two types of statistics.

For the log-rank type statistics, we replace $e_{ik}$ by a new censoring variable $\varepsilon_{ik}^C(\theta) = H_\theta(\xi, \eta) \wedge \xi^C(\theta)$, where $H_\theta(t) = \inf_{t > \zeta} h_1(h_2(t + \eta_z^C)) - \beta_z Z$. The corresponding estimating function for $\beta$, which also depends on $\eta$, can be written as

$$U_2^C(\beta, \eta) = \sum_{i=1}^{n} \int_0^\infty \left\{ \sum_{j=1}^{n} I\{\varepsilon_{ij}^C(\theta) \geq t\} Z_j - \sum_{j=1}^{n} I\{\varepsilon_{ij}^C(\theta) \geq t\} \right\} d\hat{N}_i(t; \theta),$$

where $\hat{N}_i(t; \theta) = \sum_{k=1}^{\infty} I(\varepsilon_{ik}(\beta) \leq t \wedge \varepsilon_{ik}(\theta))$. It follows that $(\varepsilon_{ik}, \varepsilon_{ik}^C(\theta))$ are identically and independently distributed for all $i = 1, \ldots, n$. When $h_1(t) = h_2(t) = \log(t)$, this approach reduces to the proposal of Ghosh and Lin (2003). Ding et al. (2009) considered log-rank type statistics under flexible forms of $h_i(\cdot)$ $(i = 1, 2)$ for semicompeting risks data (i.e., $k = 1$). Here because $\varepsilon_{ik}^C(\theta) \leq \varepsilon_{ik}^C$, some observations are artificially censored. Consequently if such extra artificial censoring is heavy, which may occur when $Z$ has a wide range, the resulting estimator of $\beta$ becomes inefficient.

The Gehan statistics requires the homogeneity only between a pair rather than for the whole sample. Define a different censoring variable $\varepsilon_{ij}^C(\theta) = H_\theta^C(\xi, \eta) \wedge \xi^C(\theta)$, where $H_\theta^C(t) = \inf_{t > \zeta} z_i h_1(h_2(t + \eta_z^C)) - \beta_z Z$. Let $\varepsilon_{ij}^C(\theta) = \varepsilon_{ik}(\beta) \wedge \varepsilon_{ij}^C(\theta)$ and $\delta_{ij}(\theta) = I(\varepsilon_{ik}(\beta) \leq \varepsilon_{ij}^C(\theta))$. We may consider the following Gehan-type statistic

$$U_2^G(\beta, \eta) = \sum_{1 \leq i < j \leq n} \frac{(Z_i - Z_j)\Psi_{ij}(\theta)}{n(n-1)},$$

where

$$\Psi_{ij}(\theta) = \sum_k \left\{ I\{\varepsilon_{ij}(\theta) \leq \varepsilon_{ik}(\theta), \delta_{ij}(\theta) = 1\} - I\{\varepsilon_{ik}(\theta) \leq \varepsilon_{ij}(\theta), \delta_{ij}(\theta) = 1\} \right\}.$$

Note that the kernel $\Psi_{ij}(\theta)$ can be viewed as a modification of (5). When $k = 1$, $U_2^G(\beta, \eta)$ reduces to the method developed for semicompeting risks data proposed by Peng and Fine (2006).

In light of Appendix A, we may also apply artificial censoring to the reexpressed log-rank statistics in terms of pairwise notations. As in Appendix A, we can write

$$U_2^C(\beta, \eta) = \sum_{i=1}^{n} \sum_{j=1}^{n} (Z_i - Z_j) \int_0^\infty \left\{ \sum_{i=1}^{n} I\{\varepsilon_{ij}^C(\theta) \geq t\} \right\} d\hat{N}_i(t; \theta),$$

We propose to modify the denominator of the above expression. Specifically replacing the number at risk calculated based on the whole sample by the number at risk based on a pair, we obtain

$$\sum_{i=1}^{n} \sum_{j=1}^{n} (Z_i - Z_j) \int_0^\infty \left\{ \sum_{i=1}^{n} I\{\varepsilon_{ij}^C(\theta) \geq t\} \right\} d\hat{N}_i(t; \theta),$$

where $\hat{N}_i(t; \theta) = \sum_{k=1}^{\infty} I(\varepsilon_{ik}(\beta) \leq t \wedge \varepsilon_{ij}^C(\theta))$. Multiplying the above estimating function by a factor $4\sqrt{n}/n(n-1)$, we obtain

$$U_2^{G-C}(\beta, \eta) = \sum_{1 \leq i < j \leq n} \frac{(Z_i - Z_j)\Omega_{ij}(\theta)}{n(n-1)},$$

where

$$\Omega_{ij}(\theta) = \sum_k \left\{ I\{\varepsilon_{ik}(\theta) \leq \varepsilon_{ij}(\theta), \delta_{ij}(\theta) = 1\} - I\{\varepsilon_{ik}(\theta) \leq \varepsilon_{ij}(\theta), \delta_{ij}(\theta) = 1\} \right\}.$$

Now we compare the two kernels $\Psi_{ij}(\theta)$ and $\Omega_{ij}(\theta)$ in (8) and (10), respectively. The first kernel $\Psi_{ij}(\theta)$ utilizes only order information for each pair. On the other hand, $\Omega_{ij}(\theta)$ computes the difference of observed events before the common censoring time, $\varepsilon_{ij}^C(\theta) \wedge \varepsilon_{ij}^C(\theta)$, for $(i, j)$ pair. Hence $U_2^C(\beta, \eta)$ uses the extra information of observed censoring times, which is a special feature of recurrent events.

Formally $\theta$ can be estimated by solving

$$U(\theta) = \begin{pmatrix} U_1(\eta) \\ U_2(\theta) \end{pmatrix} = 0,$$

where $U_2(\theta) = U_2(\beta, \eta)$ can be $U_2^{G-C}(\beta, \eta)$ or the recommended $U_2^C(\beta, \eta)$ or $u_2^{L-C}(\beta, \eta)$ with either kernel function. A convenient solution can be obtained by first obtaining $\hat{\eta}$ and then solving $U_2(\beta, \hat{\eta}) = 0$.

3.2 Estimation of $\theta_0$ When $h_2(\cdot)$ Is Unknown

Now we modify the proposed inference methods when $D | Z$ follows a transformation model with $h_2(t)$ unknown but the distribution of $\xi$ completely specified. Define $S_2(t) = Pr(\xi > t)$, which is a known function and $S_{D_0}(t) = Pr(D > t | Z = 0)$, which is the baseline survival function if death is the event of dependent censoring. Because $S_{D_0}(t) = S_0(t) \circ S_{D_1}(t)$, we have $h_2(t) = S_0^{-1} \circ S_{D_1}(t)$ and $h_2^{-1}(t) = S_0^{-1} \circ S_{D_0}(t)$.

Existing methods can be applied to estimate $S_{D_0}(t)$. For the Cox PH model with $t$ being an observed failure time of $D$, the Nelson–Aalen estimator of $S_{D_0}(t)$ is given by

$$\prod_{X_j \leq t} \left\{ 1 - \frac{\exp(\beta X_j Z_j)}{\sum_{j=1}^{n} I(X_j \geq X_i) \exp(\beta X_j Z_j)} \right\}.$$

Under the proportional odds model, Murphy, Rossini, and Van Der Waart (1997) proposed the maximum likelihood
estimator of $S_{D_0}(t)$. Denote $\hat{S}_{D_0}(t)$ as a uniformly consistent estimator of $S_{D_0}(t)$. The proposed estimating functions for $\theta$ discussed above can be modified by replacing $h_2(t)$ with $\hat{h}_2(t)$, where $\hat{h}_2(t) = \hat{S}_{D_0}(t)$ and $h_2^{-1}(t)$ by $\hat{S}_{D_0}^{-1}(t)$. Chen, Jin, and Ying (2002) and Zeng and Lin (2007a) proposed methods for estimating $\eta$ and $h_2(\cdot)$, which are applicable to the whole class of transformation models. The moment-type estimating equations proposed by Chen et al. (2002) are easier to implement than the nonparametric maximum likelihood estimation (NPMLLE) approach of Zeng and Lin (2007a) and usually comparable in efficiency. These estimators of $h_2(t)$ also denoted as $\hat{h}_2(t)$ can be directly applied. Therefore, we can obtain $\hat{\theta}$ by solving $U^*(\theta) = 0$, where $U^*(\theta)$ is $U(\theta)$ with $h_2(t)$ in the expression replaced by $\hat{h}_2(t)$ and $h_2^{-1}(t)$ replaced by $\hat{h}_2^{-1}(t)$.

3.3 Asymptotic Properties of $\hat{\theta}$

Now we derive asymptotic properties of $\hat{\theta}$, which solves $U(\theta) = 0$ with $U_2(\theta) = U_2^G(\theta)$ or $U_2^G(\theta)$.

**Theorem 1.**

1. $E[U(\theta_0)] = 0$ where $\theta_0$ is the true parameter value.
2. Under the regularity conditions listed in Appendix B, $\hat{\theta}$ is a consistent estimator.

$$\sqrt{n}(\hat{\theta} - \theta_0) \overset{d}{\rightarrow} N_\theta(0, \Lambda_0 \Sigma_0 \Lambda_0^{-1}'\Sigma_0 \Lambda_0^{-1})$$

where $\Lambda_0 = V_{\theta}E[U(\theta)]\big|_{\theta = \theta_0}$ and $\Sigma_0$ can be estimated by $n^{-1} \sum_{i=1}^{n} J_i J_i'$, where $J_i = (J^{(1)}_i)$ and

$$J^{(1)}_i = \begin{bmatrix} Z_i - \frac{1}{n} \sum_{j=1}^{n} \{ \xi_i(\hat{\eta}) \geq \xi_i(\hat{\eta}) \} Z_j \\ - \frac{1}{n} \sum_{j=1}^{n} \delta_i I(\xi_j(\hat{\eta}) \geq \hat{\xi}(\hat{\eta})) \end{bmatrix}$$

$$J^{(2)}_i = \begin{bmatrix} Z_i - \frac{1}{n} \sum_{j=1}^{n} \{ \xi_i(\hat{\eta}) \geq \hat{\xi}(\hat{\eta}) \} Z_j \\ - \frac{1}{n} \sum_{j=1}^{n} \delta_i I(\xi_j(\hat{\eta}) \geq \hat{\xi}(\hat{\eta})) \end{bmatrix}$$

$$Q_\theta = \Psi_\theta = \Omega_\theta$$

**Theorem 2.** Assume that $|\hat{S}_{D_0}(t) - S_{D_0}(t)| \to 0$ in probability uniformly on the interval $[0, \tau_0]$ for all $\tau_0 > 0$. Let $\hat{\eta}$ and $\beta$ denote the solutions to $U_1^*(\eta) = 0$ and $U_2^*(\theta, \eta) = 0$. Then $\hat{\eta}$ and $\hat{\beta}$ have the same asymptotic distribution as (12) and the variance formula is (13) with $h_2(t)$ replaced by $\hat{h}_2(t) = S_{D_0}^{-1}(t) \circ \hat{S}_{D_0}(t)$.

The proofs of Theorems 1 and 2 are presented in Web Appendix A.

Estimation of the variance involves evaluating $\Lambda_0 = \nabla_\theta E[U(\theta)]\big|_{\theta = \theta_0}$, which is quite complicated. Because the observed estimating function $U(\theta)$ is a step function, we cannot directly compute its numerical derivatives to estimate $\Lambda_0$. As suggested by Kalbfleisch and Prentice (2002, p. 238), we apply the resampling technique originally developed by Parzen, Wei, and Ying (1994) for variance estimation and constructing confidence intervals. Specifically given the observed data, define the equation:

$$U(\theta) = \begin{bmatrix} U_1(\eta) \\ U_2(\theta) \end{bmatrix} = -n^{-1/2} \sum_{i=1}^{n} J_i G_i,$$

where $(G_1, G_2, \ldots, G_n)$ are independent standard normal variables. Define $\theta^* = (\eta^*, \beta^*)'$ as the root of equation (14). Applying similar arguments in Lin, Robins, and Wei (1996), the conditional distribution of $\sqrt{n}(\theta^* - \theta_0)$, given the observed data, is asymptotically the same as the unconditional distribution of $\sqrt{n}(\theta - \theta_0)$. To approximate the distribution of $\hat{\theta}$, we can obtain a large number of realizations of $\theta^*, (\theta^*_1, \theta^*_2, \ldots, \theta^*_B)$, by repeatedly generating random samples of $(G_1, G_2, \ldots, G_n)$ for solving equation (14) $B$ times while fixing the observed data $\{(N_i, X_i, \delta_i, Z_i) : i = 1, 2, \ldots, n\}$. Then we can estimate the SE of $\hat{\theta}$ from the $B$ resampled estimators by

$$\hat{SE}(\beta_j) = \sqrt{(B-1)^{-1} \sum_{i=1}^{B} \beta_{j,i}^* - \bar{\beta}_{j}}^2,$$

$$\hat{SE}(\eta_j) = \sqrt{(B-1)^{-1} \sum_{i=1}^{B} \eta_{j,i}^* - \bar{\eta}_{j}}^2,$$

with $\bar{\beta}_{j} = B^{-1} \sum_{i=1}^{B} \beta_{j,i}^*$ and $\bar{\eta}_{j} = B^{-1} \sum_{i=1}^{B} \eta_{j,i}^*$. The 95% confidence interval (Cov) is calculated as $\beta_j \pm 1.96 \hat{SE}(\beta_j)$ and $\eta_j \pm 1.96 \hat{SE}(\eta_j)$, where $\beta_j$ is the jth component of $\beta$, $\eta_j$ is the jth component of $\eta$, $\beta_{j,i}^*$ is the jth component of $\beta_i^*$, and $\eta_{j,i}^*$ is the jth component of $\eta_i^*$.

4. Model Checking and Selection

Recall that we have defined the two counting processes: $N_i(t; \eta) = \delta_i I(\xi_i(\eta) \leq t)$ and $\tilde{N}_i(t; \theta) = \sum_{k=1}^{\tau_0} I(\xi_{ik}(\beta) \leq t \wedge \xi_{ik}^C(\Theta))$. Then define $\tilde{M}_1(t; \eta) = N_i(t; \eta) - \int_0^t I(\xi_i(\eta) \geq u) d\tilde{R}_0(u; \eta)$ and $\tilde{M}_2(t; \theta) = \tilde{N}_i(t; \theta) - \int_0^t I(\xi_i(\eta) \geq u) dR_0(u; \theta)$, where

$$\tilde{R}_0(t; \eta) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(u; \eta)}{\sum_{j=1}^{n} I(\xi_j(\eta) \geq u)}$$
Regression Analysis for Recurrent Events Data

and

\[ \tilde{R}_0(t; \theta) = \sum_{i=1}^{n} \int_0^t \frac{d\tilde{N}_{i}(u; \theta)}{\sum_{j=1}^{i} \mathbb{I} \{ \tilde{\xi}_{ij}^C(\theta) \geq u \}}. \]

To verify the two marginal regression assumptions in (2), consider the score processes

\[ S_1(t; \eta) = n^{-1/2} \sum_{i=1}^{n} Z_i \tilde{M}_{i1}(t; \eta); \]
\[ S_2(t; \theta) = n^{-1/2} \sum_{i=1}^{n} Z_i \tilde{M}_{i2}(t; \theta). \]

Let \( S(u, v; \hat{\theta}) = (S_{1(u, \hat{\eta}), S_{2(u, \hat{\theta})}}) \). By the argument in Appendix 2 of Lin et al. (1996), under the assumed models, \( S(u, v; \hat{\theta}) = S_{1(u, \hat{\eta})}/S_{2(u, \hat{\theta})} \) converges weakly to a zero-mean Gaussian process whose distribution can be approximated by that of \( \hat{S}(u, v) = (\hat{S}_{1(u)}/\hat{S}_{2(v)}) \), where

\[ \hat{S}_1(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^t \left[ \sum_{j=1}^{n} \mathbb{I} \{ \tilde{\xi}_j(\hat{\eta}) \geq s \} Z_j \right] \]
\[ \times d\tilde{M}_{i1}(s; \hat{\eta})G_i + S_{1(t; \hat{\eta})} = S_{1(t; \hat{\eta})}, \]
\[ \hat{S}_2(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^t \left[ \sum_{j=1}^{n} \mathbb{I} \{ \tilde{\xi}_j^C(\hat{\theta}) \geq s \} Z_j \right] \]
\[ \times d\tilde{M}_{i2}(s; \hat{\theta})G_i + S_{2(t; \hat{\theta})} = S_{2(t; \hat{\theta})}. \]

To approximate the null distribution of \( S(u, v; \hat{\theta}) \), we generate a large number of realizations of \( \hat{S}(u, v) \) by repeatedly generating standard normal random samples \( (G_1, G_2, \ldots, G_n) \).

Graphical diagnostics can be conducted by plotting \( S(u, v; \hat{\theta}) \) together with a few, say 20 to 30, resampled realizations of \( S(u, v) \). Furthermore, a formal testing procedure can be conducted based on the deviation statistics \( \sup \| S_1(t; \hat{\eta}) \| \) and \( \sup \| S_2(t; \hat{\theta}) \| \) with the p-values being approximated by the empirical probabilities obtained from resampling the process \( S(u, v) \) many times. Specifically, in the ith resampling step, we first obtain \( \hat{\eta}_i \) and \( \hat{\theta}_i \), and then calculate \( \hat{M}S_{1i} = \sup \| S_1(t; \hat{\eta}) \| \) and \( \hat{M}S_{2i} = \sup \| S_2(t; \hat{\theta}) \| \) for \( i = 1, 2, \ldots, B \). Consider the p-values defined as \( p_1 = B^{-1} \sum_{i=1}^{B} \mathbb{I} \{ \hat{M}S_{1i} \geq \sup \| S_1(t; \hat{\eta}) \| \} \) and \( p_2 = B^{-1} \sum_{i=1}^{B} \mathbb{I} \{ \hat{M}S_{2i} \geq \sup \| S_2(t; \hat{\theta}) \| \} \). Formal testing procedures can be conducted. Specifically the model assumption \( h_2(D) = \eta_0^C Z + \zeta \) or \( h_1(T_k) = \beta_1^T Z + \epsilon_k \) is rejected when \( p_1 \) or \( p_2 \) is smaller than a prespecified level. Respectively, in practice, more than one model may be selected. The one which gives the largest p-values can be chosen as the best fitted model. Specifically we first select the model for \( D \), which gives the largest \( p_1 \) value and then choose the model with the largest \( p_2 \) value for fitting \( T_k \).

In summary, we recommend the following procedure for analyzing the recurrent events data with dependent censoring. First, we consider models for the dependent censoring times \( D \) in model (1): either linear regression models (\( h_2(t) \) being specified) fitted by log-rank statistics (3) or Gehan-type statistics (4), or the linear transformation models (\( h_2(t) \) unknown and the distribution of \( \xi \) being specified) fitted by standard methods in the literatures cited in Section 3.2. For each fitted model on \( D \), we calculate the p-value \( p_1 \) by applying the model-checking procedure based on \( \sup \| S_1(t; \hat{\eta}) \| \) through resampling. The selected model of \( D \) is the one which gives the largest value of \( p_1 \). Then we consider linear regression models for the recurrent events \( T_k \)’s in model (1). We fit candidate models by \( \hat{\beta}^G \) in (7) or \( \hat{\beta}^{L^G} \) in (9). For each model of \( T_k \)’s, we calculate the p-value \( p_2 \) based on \( \sup \| S_2(t; \hat{\theta}) \| \) through resampling. The one with the largest value of \( p_2 \) is selected as the model for \( T_k \). Diagnostics plots for all the fitted models can also be reported.

5. Simulation Studies

First, we compare several estimators of \( \beta \) under four simulation settings. The two proposed estimators \( \hat{\beta}^G \) and \( \hat{\beta}^{L^G} \) solve the Gehan-type estimating equations \( U^G_2(\beta, \eta) = 0 \) in (7) or \( U^{L^G}_2(\beta, \eta) = 0 \) in (9). The estimator of Ghosh and Lin (2003) assumes the AFT models and solves \( U^G_2(\beta, \eta) = 0 \) in (6) with \( h_1(t) = h_2(t) = \log(t) \). Alternatively Huang and Wang (2004) assume the proportional intensity (PI) model for recurrent events and the Cox PH model for the dependent censoring event conditional on a common latent frailty variable \( \nu \). That is,

\[ E[dN(t) \mid \nu, Z] = \nu E[dN(t)] \exp(-\beta_0 Z), \]

\[ E[dI \{ D \leq t \} \mid \nu, Z] = \nu E[dI \{ D_0 \leq t \}] \exp(-\beta_0 Z). \]

The PI and PH models are conditional on both \( \nu \) and \( Z \), but, conditional on \( Z \), the PH and PI assumptions may not hold. Notice that when \( h_2(t) \) is unknown and \( \xi \) follows the extreme value distribution, our model (1) for \( D \) is the PH model conditional on \( Z \) only with

\[ E[dI \{ D \leq t \} \mid Z] = E[dI \{ D_0 \leq t \}] \exp(-\beta_0 Z). \]

We first evaluate the situation (case 1) that all four different estimators are valid. This happens when \( D_0 \) follows the exponential distribution so that \( D \) follows both the PH and AFT models. In addition the PI model (16) for \( N(t) \) concurs with the AFT model (1) for \( T_k \)'s when the marginal distributions for \( T_k \) are exponential distributions. We first generate a latent random variable \( \nu \) from a Gamma distribution with mean and variance both equal to 1. The latent variable \( \nu \) is used to create the association between \( \epsilon_k \) and \( \xi \) through model (16) and (17). Let \( W \) and \( \exp(\xi) \) follow exponential distributions with hazard rates \( 5\nu \) and \( \nu \), respectively. Set \( \exp(\xi_j) = \sum_{j=1}^{\infty} W_j \), where \( W_j > 0 \), and \( W_j \) are independent for \( i \neq j \) but follow the same distribution as \( W \). We set
For each simulation run, the four estimators, namely proposed $\hat{β}^C$, $\hat{β}^L^G$, the Ghosh–Lin estimator, and the Huang–Wang estimator are calculated. Based on $r = 500$ simulation runs, we report the average bias (Bias) $\sum_{i=1}^{500} (\hat{β}_i - β)/500$, the empirical SE $\sqrt{\sum_{i=1}^{500} (\hat{β}_i - β)^2}/499$, where $\overline{β} = \sum_{i=1}^{500} \hat{β}_i/500$, the average of the standard error estimator (SEE) $\sum_{i=1}^{500} \overline{SE}(\hat{β}_i)/500$ with $\overline{SE}$ defined in (15), and the coverage probability of nominal 95% Cov and calculated from $B = 50$ resampling datasets. Tables 1 and 2 summarize the results of case 1 and case 4, respectively. The results of case 2 and case 3 can be found in Web Appendix B.

The two proposed estimators perform well and are more efficient than the two competitors in all the cases. For the first case, all four estimators have small biases as their model assumptions are all satisfied. The two proposed estimators have smaller SE than the two competitors in this case. In particular, $\hat{β}^G$ outperforms the Ghosh–Lin estimator $\hat{β}^G$. This confirms that the pairwise construction does alleviate
Table 2

Finite-sample performances of four estimators under case 4: Ti and D follows Loc–AFT model

<table>
<thead>
<tr>
<th>Z</th>
<th>( \hat{\beta}_{LG} )</th>
<th>( \hat{\beta}_{G} )</th>
<th>( \hat{\beta}_{HW} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1697</td>
<td>0.1907</td>
<td>0.2083</td>
</tr>
<tr>
<td>2</td>
<td>0.1917</td>
<td>0.2068</td>
<td>0.2249</td>
</tr>
<tr>
<td>3</td>
<td>0.2097</td>
<td>0.2268</td>
<td>0.2459</td>
</tr>
<tr>
<td>4</td>
<td>0.2277</td>
<td>0.2468</td>
<td>0.2659</td>
</tr>
</tbody>
</table>

Note: \( \hat{\beta}_{LG} \) from (7): \( \hat{\beta}_{G} \): Ghosh–Lin estimator. \( \hat{\beta}_{HW} \): Huang–Wang estimator. The averaged bias (Bias), standard error (SE), average of the standard error estimate (SEE), and coverage probability of nominal 95% confidence interval (Cov) are calculated based on 500 replications each with sample size \( n = 100 \).
Table 3

Performances of $\hat{\beta}_G$ and $\hat{\beta}_{LG}$ under model misspecification for $D$

<table>
<thead>
<tr>
<th>Data and method</th>
<th>$Z$</th>
<th>$(\eta_0, \beta_0)$</th>
<th>$\hat{\beta}_G$</th>
<th>Bias</th>
<th>SE</th>
<th>SEE</th>
<th>Cov</th>
<th>$\hat{\beta}_{LG}$</th>
<th>Bias</th>
<th>SE</th>
<th>SEE</th>
<th>Cov</th>
</tr>
</thead>
<tbody>
<tr>
<td>D: AFT–AFT</td>
<td>$Ber(0.5)$</td>
<td>$(1.0, 0.5)$</td>
<td>0.0151</td>
<td>0.0807</td>
<td>0.0869</td>
<td>0.956</td>
<td>$-$0.0104</td>
<td>0.0809</td>
<td>0.0851</td>
<td>0.962</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(case 2)</td>
<td></td>
<td>$(0.5, 1)$</td>
<td>$-$0.0086</td>
<td>0.1087</td>
<td>0.1167</td>
<td>0.956</td>
<td>0.00134</td>
<td>0.1131</td>
<td>0.1113</td>
<td>0.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M: AFT–PH</td>
<td>$U(0, 2)$</td>
<td>$(0.5, 1)$</td>
<td>$-$0.0058</td>
<td>0.1197</td>
<td>0.1354</td>
<td>0.956</td>
<td>$-$0.0056</td>
<td>0.1173</td>
<td>0.1291</td>
<td>0.952</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$tN(0, 1)$</td>
<td>$(1.0, 0.5)$</td>
<td>$-$0.0205</td>
<td>0.0631</td>
<td>0.0674</td>
<td>0.944</td>
<td>0.0224</td>
<td>0.0611</td>
<td>0.0639</td>
<td>0.942</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(case 2)</td>
<td></td>
<td>$(0.5, 1)$</td>
<td>$-$0.0083</td>
<td>0.0616</td>
<td>0.0638</td>
<td>0.950</td>
<td>$-$0.0069</td>
<td>0.0602</td>
<td>0.0599</td>
<td>0.946</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D: AFT–AFT</td>
<td>$Ber(0.5)$</td>
<td>$(1.0, 0.5)$</td>
<td>0.0126</td>
<td>0.0881</td>
<td>0.0912</td>
<td>0.962</td>
<td>$-$0.0010</td>
<td>0.0826</td>
<td>0.0868</td>
<td>0.950</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(case 2)</td>
<td></td>
<td>$(0.5, 1)$</td>
<td>0.0365</td>
<td>0.1137</td>
<td>0.1184</td>
<td>0.942</td>
<td>0.0377</td>
<td>0.1123</td>
<td>0.1144</td>
<td>0.940</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M: AFT–Loc</td>
<td>$U(0, 2)$</td>
<td>$(0.5, 1)$</td>
<td>$-$0.0173</td>
<td>0.0704</td>
<td>0.0721</td>
<td>0.940</td>
<td>$-$0.0091</td>
<td>0.0684</td>
<td>0.0672</td>
<td>0.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$tN(0, 1)$</td>
<td>$(1.0, 0.5)$</td>
<td>$-$0.0506</td>
<td>0.0606</td>
<td>0.0594</td>
<td>0.858</td>
<td>$-$0.0378</td>
<td>0.0577</td>
<td>0.0573</td>
<td>0.886</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(case 2)</td>
<td></td>
<td>$(0.5, 1)$</td>
<td>0.0368</td>
<td>0.0604</td>
<td>0.0612</td>
<td>0.934</td>
<td>0.0289</td>
<td>0.0582</td>
<td>0.0573</td>
<td>0.930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D: AFT–PH</td>
<td>$Ber(0.5)$</td>
<td>$(1.0, 0.5)$</td>
<td>0.0544</td>
<td>0.1379</td>
<td>0.1398</td>
<td>0.928</td>
<td>0.0459</td>
<td>0.1322</td>
<td>0.1319</td>
<td>0.930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(case 3)</td>
<td></td>
<td>$(0.5, 1)$</td>
<td>0.0210</td>
<td>0.1401</td>
<td>0.1509</td>
<td>0.960</td>
<td>0.0161</td>
<td>0.1309</td>
<td>0.1395</td>
<td>0.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M: AFT–AFT</td>
<td>$U(0, 2)$</td>
<td>$(0.5, 1)$</td>
<td>0.0835</td>
<td>0.1254</td>
<td>0.1320</td>
<td>0.900</td>
<td>0.0703</td>
<td>0.1180</td>
<td>0.1204</td>
<td>0.914</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$tN(0, 1)$</td>
<td>$(1.0, 0.5)$</td>
<td>0.0318</td>
<td>0.1495</td>
<td>0.1494</td>
<td>0.936</td>
<td>0.0246</td>
<td>0.1374</td>
<td>0.1364</td>
<td>0.930</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(case 3)</td>
<td></td>
<td>$(0.5, 1)$</td>
<td>0.0366</td>
<td>0.0786</td>
<td>0.0777</td>
<td>0.916</td>
<td>0.0317</td>
<td>0.0759</td>
<td>0.0730</td>
<td>0.930</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.5, 1)$</td>
<td>0.0159</td>
<td>0.0852</td>
<td>0.0863</td>
<td>0.930</td>
<td>0.0128</td>
<td>0.0817</td>
<td>0.0811</td>
<td>0.942</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: “D” means the data generation model and “M” means the fitted model. Sample sizes $n = 100$ and replications $r = 500$.

Table 4

Performance of the proposed model-selection procedure when data follow AFT–AFT models

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(true)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Z \sim Ber(0.5)$</td>
<td>0.795</td>
<td>0.115</td>
<td>0.088</td>
<td>0</td>
<td>0.003</td>
</tr>
<tr>
<td>$Z \sim U(0, 2)$</td>
<td>0.863</td>
<td>0.108</td>
<td>0.030</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: Sample sizes $n = 200$, resampling times $B = 400$, and replications $r = 400$.
Based on 1000 replications, sup\(|S_1(t; \tilde{\theta})| = 0.7473\) with p-value = 0.374 for AFT model and sup\(|S_2(t; \tilde{\theta})| = 0.7943\) with p-value = 0.339 for Loc model. Hence we fit the AFT model for recurrent event times. In Web Appendix C, we present graphical model-checking plots for \(S_1(t; \tilde{\eta})\) and \(S_2(t; \tilde{\eta})\) with 20 realizations of \(\hat{S}_1(t)\) and \(\hat{S}_2(t)\) based on the AFT–AFT model combination. The figures also show that the AFT–AFT assumption is suitable.

In Web Appendix C, we also provide the fitted results of three estimators, namely, the proposed estimators based on (7) and (9) and the Ghosh–Lin estimator (2003). On average, the survival time for a HIV-negative subject is almost three times of that for a HIV-positive subject and the difference is significant. However, on average, the time to each hospitalization for a HIV-negative subject is about the same as the time for a HIV-positive subject (i.e., 1.003 times based on \(\hat{\beta}_G\), 1.07 based on \(\hat{\beta}^{G,C}\), and 1.05 based on the Ghosh–Lin estimator). Among the three estimators, \(\hat{\beta}_G\) gives the smallest estimated SE. Covs for all three estimators \(\hat{\beta}\) contain zero, indicating that HIV status has no significant effect on the times to repeated hospitalizations. In summary, our analysis shows that the two groups with different HIV status do not differ in the time to repeated hospitalizations but are significantly different in survival time. Furthermore, for the artificial censoring proportion, ACP1 is 0.127 for our proposed method and ACP2 is 0.441 for the Ghosh–Lin estimator, which indicates that our proposed method indeed reduces the artificial censoring proportion in analysis of the ALIVE data.

7. Concluding Remarks
The proposed estimating functions are originally derived from nonparametric statistics so that distributional assumptions can be avoided. To handle dependent censoring, the technique of artificial censoring is applied to maintain the homogeneity for (hypothetical) observations used in the computation. In particular, we propose to apply artificial censoring to two Gehan-type statistics constructed based on pairwise comparison that can utilize more data. The two proposals differ in their kernel functions. One kernel function is a direct extension from the Gehan statistics suitable for semicompeting risks data to recurrence events data. The other type of kernel function uses extra time information in pairwise comparison. The simulations indicate that neither of the two estimators dominate each other. The simulation analysis suggests using the one with smaller estimated SE.

For practical applications, the proposed approach permits flexible model combination for the recurrent event times and the survival time without specifying the form of dependence. We also provide concrete guidelines for selecting the best fitted model combination and a more efficient estimator based on the data at hand. Extension of the work to allow for \(h_1\) (-) being unknown (i.e., transformation models for \(T_k\)) will be our future work.

8. Supplementary Materials
The Web Appendices referenced in Sections 3.3, 5, and 6 are available under the Paper Information link at the Biometrics website http://www.biometrics.tibs.org.
ACKNOWLEDGEMENTS

This article was financially supported by the National Science Council of Taiwan (NSC97-2118-M-194-003-MY2). The authors are grateful to Dr Shruti Mehta and Jacquie Astemborski for providing the ALIVE cohort study data, which was supported by National Institute on Drug Abuse (DA12568 and DA04334).

REFERENCES


Received August 2009. Revised July 2010. Accepted July 2010.

APPENDIX

Appendix A: Relationship between Log Rank and Gehan Statistics

The log-rank statistics in (3) can be reexpressed as

\[
U^*_1(\eta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\infty \left( \sum_{j=1}^n I(\xi_i(t) \geq t) \right) dN_{ij}(t; \eta)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^n (Z_i - Z_j) \int_0^\infty \frac{I(\xi_i(t) \geq t)}{\sum_{l=1}^n I(\xi_l(t) \geq t)} dN_{ij}(t; \eta).
\]

The Gehan-type statistics in (4) can be written as

\[
U^*_G(\eta) = \frac{2\sqrt{n}}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n (Z_i - Z_j) I(\xi_i(t) \leq \xi_j(t), \delta_i = 1)
\]

\[
= \frac{4\sqrt{n}}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n (Z_i - Z_j)
\]

\[
\times \int_0^\infty \frac{I(\xi_i(t) \geq t)}{I(\xi_i(t) \geq t) + I(\xi_j(t) \geq t)} dN_{ij}(t; \eta).
\]

The main difference between the two statistics is just in the denominator: whether to sum over all observations (i.e., the log rank) or sum over only pairs (i.e., the Gehan).

Appendix B: Regularity Conditions of Theorem 1

We assume the following regularity conditions:

C0: The regularity conditions for \( \eta \) in Ying (1993).

C1: The parameter space \( \mathcal{P} \) for \( \beta \) is compact, and true parameter \( \beta_0 \) is an interior point of \( \mathcal{P} \).
C2: \( \theta_0 \) is the unique solution to (11).

C3: \(|Z|\) is bounded. Conditional on \(Z\), the conditional densities of \(\xi, C, \) and \(\varepsilon_k\) for \(k = 1, 2, \ldots\), and the conditional second moment of \(K = N^*(X)\) are all uniformly bounded. We denote a constant \(K_0\) for the uniform bound.

C4: \(E[U(\theta)]\) is differentiable and the Jacobian matrix is nonsingular at the true parameter value \(\theta_0\).

C5: Both \(\lim_{t \to 0} h'_1(t)/h'_2(t)\) and \(\lim_{t \to \infty} h'_1(t)/h'_2(t)\) exist with the limits allowed to be \(\infty\).

Compared to the regularity conditions in Peng and Fine (2006), the conditions C1 and C2 are the same; the condition C3 includes an additional bound for \(E(K^2)\) to address the generalization to recurrent events. The condition C5 is added to address the general marginal model (1) including models other than AFT.

Under these conditions, the theorems can be proved similar to those in Peng and Fine (2006) although the proof is a bit more technically involved with the generalized model. The detailed proof is in Web Appendix A.