 DOI: 10.1111/j.1541-0420.2010.01497.x

Regression Analysis for Recurrent Events Data under Dependent Censoring

Jin-Jian Hsieh,[1,*] A. Adam Ding,[2,] and Weijing Wang[3,***]**

[1]Department of Mathematics, National Chung Cheng University, Chia-Yi, Taiwan, Republic of China
[2]Department of Math, Northeastern University, Boston, Massachusetts 02115, U.S.A.
[3]Institute 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 censorship. 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 $\boldsymbol{Z}$ be the vector of covariates. Throughout the article, we assume that C may depend on $\boldsymbol{Z}$ but C is independent of $N^*(t)$ and D given $\boldsymbol{Z}$. On the other hand, D is correlated with the process $N^*(t)$, despite $\boldsymbol{Z}$. The recurrence process can also be expressed in terms of recurrence times. Define T_k as the time from the origin to the kth recurrent event ($k = 1, 2, \ldots$). Hence $N^*(t) = \sum_{k=1}^{\infty} I(T_k \le t)$. Observed variables can be denoted as $N(t) = N^*(t \wedge D \wedge C)$, $X = D \wedge C$, and $\delta = I(D \le C)$. There are $K = N(X)$ events observed successively at times $T_k = \min\{t\colon N(t) \ge k\}$ 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:

$$
\begin{aligned}
h_1(T_k) &= \boldsymbol{\beta}_0'\boldsymbol{Z} + \varepsilon_k, \\
h_2(D) &= \boldsymbol{\eta}_0'\boldsymbol{Z} + \xi,
\end{aligned} \tag{1}
$$

where $h_1(\,\cdot\,)$ is a known monotone function, $h_2(\,\cdot\,)$ is a monotone function that may be known or unknown, $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\eta}_0)'$ are the true parameters, ε_k and ξ are the error variables. ε_k and ξ may be correlated with each other but both are independent of $\boldsymbol{Z}$. The marginal distribution of ε_k is unspecified. The marginal distribution of ξ 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 ξ following the extreme value distribution.

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

$$
\begin{bmatrix} N^*\{h_1^{-1}(t + \boldsymbol{\beta}_0'\boldsymbol{Z})\} \\ h_2(D) - \boldsymbol{\eta}_0'\boldsymbol{Z} \end{bmatrix} \overset{d}{=} \begin{bmatrix} N_{\varepsilon}^*(t) \\ \xi \end{bmatrix}, \tag{2}
$$

where "$\overset{d}{=}$" means "with the same distribution."

Estimation of $\boldsymbol{\eta}_0$ is straightforward by applying existing methods that assume independent censorship. Estimation of $\boldsymbol{\beta}_0$ is complicated due to dependent censoring. To avoid specifying the dependent relationship, we propose to modify nonparametric statistics originally constructed based on random replicates of $(N_{\varepsilon}^*(t), \xi)'$.

3. The Proposed Inference Methods

Let $\{N_i^*(.), T_{ik}, C_i, D_i, \boldsymbol{Z}_i\}$ be independent realizations of $\{N^*(.), T_k, C, D, \boldsymbol{Z}\}$ for $k = 1, \ldots$ and $i = 1, 2, \ldots, n$. Let $\varepsilon_{ik}(\boldsymbol{\beta}) = h_1(T_{ik}) - \boldsymbol{\beta}'\boldsymbol{Z}_i$ and $\xi_i(\boldsymbol{\eta}) = h_2(D_i) - \boldsymbol{\eta}'\boldsymbol{Z}_i$. It is easy to see that $(\varepsilon_{ik}, \xi_i) \equiv \{\varepsilon_{ik}(\boldsymbol{\beta}_0), \xi_i(\boldsymbol{\eta}_0)\}$ for $i = 1, \ldots, n$ are independent and identical replications of (ε_k, ξ). Define $N_{\varepsilon i}^*(t) = N^*\{h_1^{-1}(t + \boldsymbol{\beta}_0'\boldsymbol{Z}_i)\}$. Accordingly $\{(N_{\varepsilon i}^*(t), \xi_i)\ (i = 1, \ldots, n)\}$ constitute a bivariate random sample with the joint distribution independent of $\boldsymbol{Z}_i$. 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 $N_{\varepsilon i}(t) = N\{h_1^{-1}(t + \boldsymbol{\beta}_0'\boldsymbol{Z}_i)\}$. Note that $N_{\varepsilon i}(t)$ and $N_{\varepsilon j}(t)$ no longer have the same distribution for $\boldsymbol{Z}_i \ne \boldsymbol{Z}_j$ due to the dependence between ε_{ik} and ξ_i.

3.1 *Estimation of* $\boldsymbol{\theta}_0$ *When* $h_2(\,\cdot\,)$ *Is Known*

We suggest estimating $\boldsymbol{\theta}_0 = (\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)'$ by first estimating $\boldsymbol{\eta}_0$ and then $\boldsymbol{\beta}_0$. Estimation of the latter is the challenge. To motivate the proposed idea, we first discuss estimation of $\boldsymbol{\eta}_0$ when $h_2(\,\cdot\,)$ is known. Define $\tilde{\xi}_i(\boldsymbol{\eta}) = h_2(X_i) - \boldsymbol{\eta}'\boldsymbol{Z}_i$. When $\delta_i = 1$, $\tilde{\xi}_i(\boldsymbol{\eta}) = \xi_i(\boldsymbol{\eta})$; whereas when $\delta_i = 0$, $\tilde{\xi}_i(\boldsymbol{\eta}) = \xi_i^C(\boldsymbol{\eta}) = h_2(C_i) - \boldsymbol{\eta}'\boldsymbol{Z}_i$. Two statistics can be applied. The first one is the log-rank type statistic:

$$
\boldsymbol{U}_1^L(\boldsymbol{\eta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\infty} \left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} I\{\tilde{\xi}_j(\boldsymbol{\eta}) \ge t\}\boldsymbol{Z}_j}{\sum_{l=1}^{n} I\{\tilde{\xi}_l(\boldsymbol{\eta}) \ge t\}} \right] dN_{\xi i}(t;\boldsymbol{\eta}), \tag{3}
$$

where $N_{\xi i}(t;\boldsymbol{\eta}) = I(\tilde{\xi}_i(\boldsymbol{\eta}) \le t, \delta_i = 1)$. The other is the Gehan-type statistics:

$$
\boldsymbol{U}_1^G(\boldsymbol{\eta}) = 2\sqrt{n} \sum_{1 \le i < j \le n} \frac{(\boldsymbol{Z}_i - \boldsymbol{Z}_j)\Phi_{ij}(\boldsymbol{\eta})}{n(n-1)}, \tag{4}
$$

where

$$
\Phi_{ij}(\boldsymbol{\eta}) = I\{\tilde{\xi}_i(\boldsymbol{\eta}) \le \tilde{\xi}_j(\boldsymbol{\eta}), \delta_i = 1\} - I\{\tilde{\xi}_j(\boldsymbol{\eta}) \le \tilde{\xi}_i(\boldsymbol{\eta}), \delta_j = 1\}. \tag{5}
$$

Note that $\Phi_{ij}(\boldsymbol{\eta}_0)$ and $\Phi_{ji}(\boldsymbol{\eta}_0)$ have the same distribution for $i \ne j$. Estimators of $\boldsymbol{\eta}$ 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 be further 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 $\boldsymbol{\beta}$ in the presence of dependent censoring.

Under the error scale, ε_{ik} is subject to censoring by $\varepsilon_i^C = h_1[h_2^{-1}\{(\xi_i \wedge \xi_i^C) + \boldsymbol{\eta}_0'\boldsymbol{Z}_i\}] - \boldsymbol{\beta}_0'\boldsymbol{Z}_i$. As a result, $(\varepsilon_{ik}, \varepsilon_i^C)$ no longer have the same distribution for $\boldsymbol{Z}_i \neq \boldsymbol{Z}_j$ due to dependence between ε_{ik} and ξ_i. 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 ε_i^C by a new censoring variable $\tilde{\varepsilon}_i^C(\boldsymbol{\theta}) = H_{\boldsymbol{\theta}}\{\xi_i(\boldsymbol{\eta}) \wedge \xi_i^C(\boldsymbol{\eta})\}$, where $H_{\boldsymbol{\theta}}(t) = \inf_{\text{all } \boldsymbol{z}} h_1\{h_2^{-1}(t + \boldsymbol{\eta}'\boldsymbol{z})\} - \boldsymbol{\beta}'\boldsymbol{z}$. The corresponding estimating function for $\boldsymbol{\beta}$, which also depends on $\boldsymbol{\eta}$, can be written as

$$\boldsymbol{U}_2^L(\boldsymbol{\beta}, \boldsymbol{\eta}) = \sum_{i=1}^n \int_0^\infty \left\{ \boldsymbol{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{\varepsilon}_j^C(\boldsymbol{\theta}) \geq t\}\boldsymbol{Z}_j}{\sum_{j=1}^n I\{\tilde{\varepsilon}_j^C(\boldsymbol{\theta}) \geq t\}} \right\} d\tilde{N}_{\varepsilon i}(t; \boldsymbol{\theta}), \quad (6)$$

where $\tilde{N}_{\varepsilon i}(t; \boldsymbol{\theta}) = \sum_{k=1}^\infty I\{\varepsilon_{ik}(\boldsymbol{\beta}) \leq t \wedge \tilde{\varepsilon}_i^C(\boldsymbol{\theta})\}$. It follows that $(\varepsilon_{ik}, \tilde{\varepsilon}_i^C(\boldsymbol{\theta}_0))$ 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 $h_j(\cdot)$ $(j = 1, 2)$ for semicompeting risks data (i.e., $k = 1$). Here because $\tilde{\varepsilon}_j^C(\boldsymbol{\theta}_0) \leq \varepsilon_j^C$, some observations are artificially censored. Consequently if such extra artificial censoring is heavy, which may occur when $\boldsymbol{Z}$ has a wide range, the resulting estimator of $\boldsymbol{\beta}$ becomes inefficient.

The Gehan statistics requires the homogeneity only between a pair rather than for the whole sample. Define a different censoring variable $\tilde{\varepsilon}_{i(j)}^C(\boldsymbol{\theta}) = H_{\boldsymbol{\theta}}^{ij}\{\xi_i(\boldsymbol{\eta}) \wedge \xi_i^C(\boldsymbol{\eta})\}$, where $H_{\boldsymbol{\theta}}^{ij}(t) = \inf_{\boldsymbol{z}=\boldsymbol{Z}_i, \boldsymbol{Z}_j} h_1\{h_2^{-1}(t + \boldsymbol{\eta}'\boldsymbol{z})\} - \boldsymbol{\beta}'\boldsymbol{z}$. Let $\tilde{\varepsilon}_{i(j)}^k(\boldsymbol{\theta}) = \varepsilon_{ik}(\boldsymbol{\beta}) \wedge \tilde{\varepsilon}_{i(j)}^C(\boldsymbol{\theta})$ and $\tilde{\delta}_{i(j)}^k(\boldsymbol{\theta}) = I\{\varepsilon_{ik}(\boldsymbol{\beta}) \leq \tilde{\varepsilon}_{i(j)}^C(\boldsymbol{\theta})\}$. We may consider the following Gehan-type statistic

$$\boldsymbol{U}_2^G(\boldsymbol{\beta}, \boldsymbol{\eta}) = 2\sqrt{n} \sum_{1 \leq i < j \leq n} \frac{(\boldsymbol{Z}_i - \boldsymbol{Z}_j)\Psi_{ij}(\boldsymbol{\theta})}{n(n-1)}, \quad (7)$$

where

$$\Psi_{ij}(\boldsymbol{\theta}) = \sum_k \left[I\{\tilde{\varepsilon}_{i(j)}^k(\boldsymbol{\theta}) \leq \tilde{\varepsilon}_{j(i)}^k(\boldsymbol{\theta}), \tilde{\delta}_{i(j)}^k(\boldsymbol{\theta}) = 1\} - I\{\tilde{\varepsilon}_{j(i)}^k(\boldsymbol{\theta}) \leq \tilde{\varepsilon}_{i(j)}^k(\boldsymbol{\theta}), \tilde{\delta}_{j(i)}^k(\boldsymbol{\theta}) = 1\} \right]. \quad (8)$$

Note that the kernel $\Psi_{ij}(\boldsymbol{\theta})$ can be viewed as a modification of (5). When $k = 1$, $\boldsymbol{U}_2^G(\boldsymbol{\beta}, \boldsymbol{\eta})$ 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

$$\boldsymbol{U}_2^L(\boldsymbol{\beta}, \boldsymbol{\eta}) = \sum_{i=1}^n \sum_{j=1}^n (\boldsymbol{Z}_i - \boldsymbol{Z}_j) \int_0^\infty \frac{I\{\tilde{\varepsilon}_j^C(\boldsymbol{\theta}) \geq t\}}{\sum_{l=1}^n I\{\tilde{\varepsilon}_l^C(\boldsymbol{\theta}) \geq t\}} d\tilde{N}_{\varepsilon_i}(t; \boldsymbol{\theta}).$$

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

$$\sum_{i=1}^n \sum_{j=1}^n (\boldsymbol{Z}_i - \boldsymbol{Z}_j) \times \int_{t=0}^\infty \left\{ \frac{I\{\tilde{\varepsilon}_{j(i)}^C(\boldsymbol{\theta}) \geq t\}}{I\{\tilde{\varepsilon}_{i(j)}^C(\boldsymbol{\theta}) \geq t\} + I\{\tilde{\varepsilon}_{j(i)}^C(\boldsymbol{\theta}) \geq t\}} \right\} d\tilde{N}_{\varepsilon_{i(j)}}(t; \boldsymbol{\theta}),$$

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

$$\boldsymbol{U}_2^{LG}(\boldsymbol{\beta}, \boldsymbol{\eta}) = 2\sqrt{n} \sum_{1 \leq i < j \leq n} \frac{(\boldsymbol{Z}_i - \boldsymbol{Z}_j)\Omega_{ij}(\boldsymbol{\theta})}{n(n-1)}, \quad (9)$$

where

$$\Omega_{ij}(\boldsymbol{\theta}) = \sum_k \left[I\{\tilde{\varepsilon}_{i(j)}^k(\boldsymbol{\theta}) \leq \tilde{\varepsilon}_{j(i)}^C(\boldsymbol{\theta}), \tilde{\delta}_{i(j)}^k(\boldsymbol{\theta}) = 1\} - I\{\tilde{\varepsilon}_{j(i)}^k(\boldsymbol{\theta}) \leq \tilde{\varepsilon}_{i(j)}^C(\boldsymbol{\theta}), \tilde{\delta}_{j(i)}^k(\boldsymbol{\theta}) = 1\} \right]. \quad (10)$$

Now we compare the two kernels $\Psi_{ij}(\boldsymbol{\theta})$ and $\Omega_{ij}(\boldsymbol{\theta})$ in (8) and (10), respectively. The first kernel $\Psi_{ij}(\boldsymbol{\theta})$ utilizes only order information for each pair. On the other hand, $\Omega_{ij}(\boldsymbol{\theta})$ computes the difference of observed events before the common censoring time, $\tilde{\varepsilon}_{i(j)}^C(\boldsymbol{\theta}) \wedge \tilde{\varepsilon}_{j(i)}^C(\boldsymbol{\theta})$, for (i, j) pair. Hence $\boldsymbol{U}_2^{LG}(\boldsymbol{\beta}, \boldsymbol{\eta})$ uses the extra information of observed censoring times, which is a special feature of recurrence events.

Formally $\boldsymbol{\theta}$ can be estimated by solving

$$\boldsymbol{U}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{U}_1(\boldsymbol{\eta}) \\ \boldsymbol{U}_2(\boldsymbol{\theta}) \end{pmatrix} = 0, \quad (11)$$

where $\boldsymbol{U}_2(\boldsymbol{\theta}) = \boldsymbol{U}_2(\boldsymbol{\beta}, \boldsymbol{\eta})$ can be $\boldsymbol{U}_2^L(\boldsymbol{\theta})$ or the recommended $\boldsymbol{U}_2^G(\boldsymbol{\theta})$ or $\boldsymbol{U}_2^{LG}(\boldsymbol{\theta})$ with either kernel function. A convenient solution can be obtained by first obtaining $\widehat{\boldsymbol{\eta}}$ and then solving $\boldsymbol{U}_2(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}}) = 0$.

3.2 *Estimation of* $\boldsymbol{\theta}_0$ *When* $h_2(\cdot)$ *Is Unknown*

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

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

$$\prod_{X_i \leq t} \left\{ 1 - \frac{\exp(\widehat{\boldsymbol{\eta}}'\boldsymbol{Z}_i)}{\sum_{j=1}^n I(X_j \geq X_i) \exp(\widehat{\boldsymbol{\eta}}'\boldsymbol{Z}_j)} \right\}^{\delta_i \exp(-\widehat{\boldsymbol{\eta}}'\boldsymbol{Z}_i)}.$$

Under the proportional odds model, Murphy, Rossini, and Van Der Waart (1997) proposed the maximum likelihood estimator of $S_{D_0}(t)$. Denote $\widehat{S}_{D_0}(t)$ as a uniformly consistent estimator of $S_{D_0}(t)$. The proposed estimating functions for $\boldsymbol{\theta}$ discussed above can be modified by replacing $h_2(t)$ with $\widehat{h}_2(t) = S_\xi^{-1} \circ \widehat{S}_{D_0}(t)$ and $h_2^{-1}(t)$ by $\widehat{h}_2^{-1}(t) = \widehat{S}_{D_0}^{-1} \circ S_\xi(t)$. Chen, Jin, and Ying (2002) and Zeng and Lin (2007a) proposed methods for estimating $\boldsymbol{\eta}$ 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 (NPMLE) approach of Zeng and Lin (2007a) and usually comparable in efficiency. These estimators of $h_2(t)$ also denoted as $\widehat{h}_2(t)$ can be directly applied. Therefore, we can obtain $\widehat{\boldsymbol{\theta}}$ by solving $\boldsymbol{U}^*(\boldsymbol{\theta}) = 0$, where $\boldsymbol{U}^*(\boldsymbol{\theta})$ is $\boldsymbol{U}(\boldsymbol{\theta})$ with $h_2(t)$ in the expression replaced by $\widehat{h}_2(t)$ and $h_2^{-1}(t)$ replaced by $\widehat{h}_2^{-1}(t)$.

3.3 *Asymptotic Properties of* $\widehat{\boldsymbol{\theta}}$

Now we derive asymptotic properties of $\widehat{\boldsymbol{\theta}}$, which solves $\boldsymbol{U}(\boldsymbol{\theta}) = 0$ with $\boldsymbol{U}_2(\boldsymbol{\theta}) = \boldsymbol{U}_2^G(\boldsymbol{\theta})$ or $\boldsymbol{U}_2^{LG}(\boldsymbol{\theta})$.

Theorem 1:

(1) $E[\boldsymbol{U}(\boldsymbol{\theta}_0)] = 0$ *where* $\boldsymbol{\theta}_0$ *is the true parameter value.*

(2) *Under the regularity conditions listed in Appendix B,* $\widehat{\boldsymbol{\theta}}$ *is a consistent estimator.*

(3)

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_{2p}(0, \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Sigma}_0(\boldsymbol{\Lambda}_0^{-1})'), \tag{12}$$

where $\boldsymbol{\Lambda}_0 = \nabla_{\boldsymbol{\theta}} E[\boldsymbol{U}(\boldsymbol{\theta})]|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ *and* $\boldsymbol{\Sigma}_0$ *can be estimated by* $n^{-1}\sum_i \boldsymbol{J}_i\boldsymbol{J}_i'$, *where* $\boldsymbol{J}_i = \binom{\boldsymbol{J}_i^{(1)}}{\boldsymbol{J}_i^{(2)}}$ *and*

$$\boldsymbol{J}_i^{(1)} = \delta_i \left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^n \{\tilde{\xi}_j(\widehat{\boldsymbol{\eta}}) \geq \tilde{\xi}_i(\widehat{\boldsymbol{\eta}})\}\boldsymbol{Z}_j}{\sum_{j=1}^n \{\tilde{\xi}_j(\widehat{\boldsymbol{\eta}}) \geq \tilde{\xi}_i(\widehat{\boldsymbol{\eta}})\}} \right] - \sum_{l=1}^n \frac{\delta_l I\{\tilde{\xi}_i(\widehat{\boldsymbol{\eta}}) \geq \tilde{\xi}_l(\widehat{\boldsymbol{\eta}})\}}{\sum_{j=1}^n I\{\tilde{\xi}_j(\widehat{\boldsymbol{\eta}}) \geq \tilde{\xi}_l(\widehat{\boldsymbol{\eta}})\}} \times \left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^n \{\tilde{\xi}_j(\widehat{\boldsymbol{\eta}}) \geq \tilde{\xi}_l(\widehat{\boldsymbol{\eta}})\}\boldsymbol{Z}_j}{\sum_{j=1}^n \{\tilde{\xi}_j(\widehat{\boldsymbol{\eta}}) \geq \tilde{\xi}_l(\widehat{\boldsymbol{\eta}})\}} \right],$$

$$\boldsymbol{J}_i^{(2)} = \frac{2}{n-1}\sum_{j=1}^n (\boldsymbol{Z}_i - \boldsymbol{Z}_j)Q_{ij}(\widehat{\boldsymbol{\theta}}), \tag{13}$$

where $Q_{ij} = \Psi_{ij}$ or Ω_{ij}.

Theorem 2: *Assume that* $|\widehat{S}_{D_0}(t) - S_{D_0}(t)| \to 0$ *in probability uniformly on the interval* $[0,\ \tau_0]$ *for all* $\tau_0 > 0$. *Let* $\tilde{\boldsymbol{\eta}}$ *and* $\tilde{\boldsymbol{\beta}}$ *denote the solutions to* $\boldsymbol{U}_1^*(\boldsymbol{\eta}) = 0$ *and* $\boldsymbol{U}_2^*(\boldsymbol{\beta}, \boldsymbol{\eta}) = 0$. *Then* $\tilde{\boldsymbol{\eta}}$ *and* $\tilde{\boldsymbol{\beta}}$ *have the same asymptotic distribution as (12) and the variance formula is (13) with* $h_2(t)$ *replaced by* $\widehat{h}_2(t) = S_\xi^{-1} \circ \widehat{S}_{D_0}(t)$.

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

Estimation of the variance involves evaluating $\boldsymbol{\Lambda}_0 = \nabla_{\boldsymbol{\theta}} E[\boldsymbol{U}(\boldsymbol{\theta})]|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, which is quite complicated. Because the observed estimating function $\boldsymbol{U}(\boldsymbol{\theta})$ is a step function, we cannot directly compute its numerical derivatives to estimate $\boldsymbol{\Lambda}_0$. 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:

$$\boldsymbol{U}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{U}_1(\boldsymbol{\eta}) \\ \boldsymbol{U}_2(\boldsymbol{\theta}) \end{pmatrix} = -n^{-1/2}\sum_{i=1}^n \boldsymbol{J}_i G_i, \tag{14}$$

where $(G_1, G_2, \ldots, G_n)$ are independent standard normal variables. Define $\boldsymbol{\theta}^* = (\boldsymbol{\eta}^*, \boldsymbol{\beta}^*)'$ as the root of equation (14). Applying similar arguments in Lin, Robins, and Wei (1996), the conditional distribution of $\sqrt{n}(\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}})$, given the observed data, is asymptotically the same as the unconditional distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. To approximate the distribution of $\widehat{\boldsymbol{\theta}}$, we can obtain a large number of realizations of $\boldsymbol{\theta}^*$, $(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \ldots, \boldsymbol{\theta}_B^*)$, by repeatedly generating random samples of $(G_1, G_2, \ldots, G_n)$ for solving equation (14) B times while fixing the observed data $\{(N_i(.), X_i, \delta_i, \boldsymbol{Z}_i) : i = 1, 2, \ldots, n\}$. Then we can estimate the SE of $\widehat{\boldsymbol{\theta}}$ from the B resampled estimators by

$$\widehat{SE}(\widehat{\beta}_j) = \sqrt{(B-1)^{-1}\sum_{i=1}^B \left(\beta_{j,i}^* - \overline{\beta}_j^*\right)^2},$$

$$\widehat{SE}(\widehat{\eta}_j) = \sqrt{(B-1)^{-1}\sum_{i=1}^B \left(\eta_{j,i}^* - \overline{\eta}_j^*\right)^2}, \tag{15}$$

with $\overline{\beta}_j^* = B^{-1}\sum_{i=1}^B \beta_{j,i}^*$ and $\overline{\eta}_j^* = B^{-1}\sum_{i=1}^B \eta_{j,i}^*$. The 95% confidence interval (Cov) is calculated as $\widehat{\beta}_j \pm 1.96\widehat{SE}(\widehat{\beta}_j)$ and $\widehat{\eta}_j \pm 1.96\widehat{SE}(\widehat{\eta}_j)$, where $\widehat{\beta}_j$ is the jth component of $\widehat{\boldsymbol{\beta}}$, $\widehat{\eta}_j$ is the jth component of $\widehat{\boldsymbol{\eta}}$, $\beta_{j,i}^*$ is the jth component of $\boldsymbol{\beta}_i^*$, and $\eta_{j,i}^*$ is the jth component of $\boldsymbol{\eta}_i^*$.

4. Model Checking and Selection

Recall that we have defined the two counting processes: $N_{\xi_i}(t;\boldsymbol{\eta}) = \delta_i I\{\tilde{\xi}_i(\boldsymbol{\eta}) \leq t\}$ and $\tilde{N}_{\varepsilon_i}(t;\boldsymbol{\theta}) = \sum_{k=1}^\infty I\{\varepsilon_{ik}(\boldsymbol{\beta}) \leq t \wedge \tilde{\varepsilon}_i^C(\boldsymbol{\theta})\}$. Then define $\widehat{M}_{1i}(t;\boldsymbol{\eta}) = N_{\xi_i}(t;\boldsymbol{\eta}) - \int_0^t I\{\tilde{\xi}_i(\boldsymbol{\eta}) \geq u\}d\widehat{H}_0(u;\boldsymbol{\eta})$ and $\widehat{M}_{2i}(t;\boldsymbol{\theta}) = \tilde{N}_{\varepsilon_i}(t;\boldsymbol{\theta}) - \int_0^t I\{\tilde{\varepsilon}_i^C(\boldsymbol{\theta}) \geq u\}d\widehat{R}_0(u;\boldsymbol{\theta})$, where

$$\widehat{H}_0(t;\boldsymbol{\eta}) = \sum_{i=1}^n \int_0^t \frac{dN_{\xi_i}(u;\boldsymbol{\eta})}{\sum_{j=1}^n I\{\tilde{\xi}_j(\boldsymbol{\eta}) \geq u\}}$$

and

$$\widehat{R}_0(t;\boldsymbol{\theta}) = \sum_{i=1}^{n}\int_0^t \frac{d\tilde{N}_{\varepsilon_i}(u;\boldsymbol{\theta})}{\sum_{j=1}^{n} I\{\tilde{\varepsilon}_j^C(\boldsymbol{\theta}) \geq u\}}.$$

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

$$\boldsymbol{S}_1(t;\boldsymbol{\eta}) = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i\widehat{M}_{1i}(t;\boldsymbol{\eta});$$

$$\boldsymbol{S}_2(t;\boldsymbol{\theta}) = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i\widehat{M}_{2i}(t;\boldsymbol{\theta}).$$

Let $\boldsymbol{S}(u,v;\widehat{\boldsymbol{\theta}}) = \binom{\boldsymbol{S}_1(u;\widehat{\boldsymbol{\eta}})}{\boldsymbol{S}_2(v;\widehat{\boldsymbol{\theta}})}$. By the argument in Appendix 2 of Lin et al. (1996), under the assumed models, $\boldsymbol{S}(u,v;\widehat{\boldsymbol{\theta}}) = \binom{\boldsymbol{S}_1(u;\widehat{\boldsymbol{\eta}})}{\boldsymbol{S}_2(v;\widehat{\boldsymbol{\theta}})}$ converges weakly to a zero-mean Gaussian process whose distribution can be approximated by that of $\widehat{\boldsymbol{S}}(u,v) = \binom{\widehat{\boldsymbol{S}}_1(u)}{\widehat{\boldsymbol{S}}_2(v)}$, where

$$\widehat{\boldsymbol{S}}_1(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t \left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} I\{\tilde{\xi}_j(\widehat{\boldsymbol{\eta}}) \geq s\}\boldsymbol{Z}_j}{\sum_{j=1}^{n} I\{\tilde{\xi}_j(\widehat{\boldsymbol{\eta}}) \geq s\}}\right] \times d\widehat{M}_{1i}(s;\widehat{\boldsymbol{\eta}})G_i + \boldsymbol{S}_1(t;\boldsymbol{\eta}^*) - \boldsymbol{S}_1(t;\widehat{\boldsymbol{\eta}}),$$

$$\widehat{\boldsymbol{S}}_2(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t \left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} I\{\tilde{\varepsilon}_i^C(\widehat{\boldsymbol{\theta}}) \geq s\}\boldsymbol{Z}_j}{\sum_{j=1}^{n} I\{\tilde{\varepsilon}_j^C(\widehat{\boldsymbol{\theta}}) \geq s\}}\right] \times d\widehat{M}_{2i}(s;\widehat{\boldsymbol{\theta}})G_i + \boldsymbol{S}_2(t;\boldsymbol{\theta}^*) - \boldsymbol{S}_2(t;\widehat{\boldsymbol{\theta}}).$$

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

Graphical diagnostics can be conducted by plotting $\boldsymbol{S}(u,v;\widehat{\boldsymbol{\theta}})$ together with a few, say 20 to 30, resampled realizations of $\widehat{\boldsymbol{S}}(u,v)$. Furthermore, a formal testing procedure can be conducted based on the deviation statistics $\sup_t \|\boldsymbol{S}_1(t;\widehat{\boldsymbol{\eta}})\|$ and $\sup_t \|\boldsymbol{S}_2(t;\widehat{\boldsymbol{\theta}})\|$ with the p-values being approximated by the empirical probabilities obtained from resampling the process $\widehat{\boldsymbol{S}}(u,v)$ many times. Specifically in the ith resampling step, we first obtain $\widehat{\boldsymbol{\eta}}_i^*$ and $\widehat{\boldsymbol{\theta}}_i^*$ and then calculate $\widehat{MS}_{1,i} = \sup_t \|\boldsymbol{S}_1(t)\|$ and $\widehat{MS}_{2,i} = \sup_t \|\boldsymbol{S}_2(t)\|$ for $i = 1, 2, \ldots, B$. Consider the p-values defined as $p_1 = B^{-1}\sum_{i=1}^{B} I\{\widehat{MS}_{1,i} \geq \sup_t \|\boldsymbol{S}_1(t;\widehat{\boldsymbol{\eta}})\|\}$ and $p_2 = B^{-1}\sum_{i=1}^{B} I\{\widehat{MS}_{2,i} \geq \sup_t \|\boldsymbol{S}_2(t;\widehat{\boldsymbol{\theta}})\|\}$. Formal testing procedures can be conducted. Specifically the model assumption $h_2(D) = \boldsymbol{\eta}_0'\boldsymbol{Z} + \xi$ or $h_1(T_k) = \boldsymbol{\beta}_0'\boldsymbol{Z} + \varepsilon_k$ 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's.

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 ξ 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 $\sup_t \|\boldsymbol{S}_1(t;\widehat{\boldsymbol{\eta}})\|$ 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 $\widehat{\boldsymbol{\beta}}^G$ in (7) or $\widehat{\boldsymbol{\beta}}^{LG}$ in (9). For each model of T_k's, we calculate the p-value p_2 based on $\sup_t \|\boldsymbol{S}_2(t;\widehat{\boldsymbol{\theta}})\|$ through resampling. The one with the largest value of p_2 is selected as the model for T_k's. Diagnostics plots for all the fitted models can also be reported.

5. Simulation Studies

First, we compare several estimators of β under four simulation settings. The two proposed estimators $\widehat{\beta}^G$ and $\widehat{\beta}^{LG}$ solve the Gehan-type estimating equations $U_2^G(\beta,\widehat{\eta}) = 0$ in (7) or $U_2^{LG}(\beta,\widehat{\eta}) = 0$ in (9). The estimator of Ghosh and Lin (2003) assumes the AFT models and solves $U_2^L(\beta,\widehat{\eta}) = 0$ 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 ν. That is,

$$E[dN(t)\,|\,\nu, Z] = \nu E[dN_0(t)]\exp(-\beta_0 Z), \tag{16}$$

$$E[dI\{D \leq t\}\,|\,\nu, Z] = \nu E[dI\{D_0 \leq t\}]\exp(-\beta_0 Z). \tag{17}$$

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

$$E[dI\{D \leq t\}\,|\,Z] = E[dI\{D_0 \leq t\}]\exp(-\beta_0 Z). \tag{18}$$

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 ν from a Gamma distribution with mean and variance both equal to 1. The latent variable ν is used to create the association between ε_k and ξ through model (16) and (17). Let W and $\exp(\xi)$ follow exponential distributions with hazard rates 5ν and ν, respectively. Set $\exp(\varepsilon_k) = \sum_{j=1}^{k} W_j$, where $W_j > 0$, W_i and W_j are independent for $i \neq j$ but follow the same distribution as W. We set

$h_1(t) = h_2(t) = \log(t)$ in (1) and (2) for the marginal generation of W and $\exp(\xi)$. That is, $T_k = \exp(\beta_0 Z)\sum_{j=1}^{k} W_j$ and $D = \exp(\eta_0 Z)\exp(\xi)$. Hence the recurrent events process and the dependent censoring time both follow AFT models as assumed by Ghosh and Lin (2003), and also follows model (16) and (17) as assumed by Huang and Wang (2004). Note that the Gamma frailty variable ν means that $(D, W)|Z$ follow a Clayton copula. In the second setup (case 2), data are generated similar to the first case, except that W and $\exp(\xi)$ follow Weibull distributions such that $\Pr(W > t) = \exp\{-10\nu(\frac{t}{2})^4\}$ and $\Pr(\exp(\xi) > t) = \exp\{-\nu(\frac{t}{2})^4\}$. Accordingly the recurrent events process and the dependent censoring time still both follow AFT models but no longer satisfy the assumption in (16). For case 3, we let $(D, W)|Z$ follow the Clayton copula with Kendall's tau equal to 0.5, with W following an exponential distribution $\Pr(W > t) = \exp\{-10t\}$ and D following a log-logistic distribution $Pr(D > t\,|\,Z) = \{1 - \frac{1}{1+t^{-8}}\}^{\exp(\eta_0 Z)}$. Then, set $T_k = \exp(\beta_0 Z)\sum_{j=1}^{k} W_j$. Hence, the recurrent events process follow the AFT model, and the dependent censoring time follow the PH model (18). However, the dependent censoring time does not follow the AFT model due to the log-logistic distribution. For case 4, we let $(\exp(\xi), W)$ follow the Clayton copula with Kendall's tau equal to 0.5, W follow an exponential distribution $\Pr(W > t) = \exp\{-10t\}$, and $\exp(\xi)$ follow an exponential distribution $\Pr(\exp(\xi) > t) = \exp\{-t\}$. Then set $h_1(t) = t$ and $h_2(t) = \log(t)$ in (1) and (2) for the marginal generation of W and $\exp(\xi)$. That is, $T_k = \beta_0 Z + \sum_{j=1}^{k} W_j$ follows the location-shift (Loc) model and $D = \exp(\eta_0 Z)\exp(\xi)$ follows the AFT model. The model assumptions in both Ghosh and Lin (2003) and Huang and Wang (2004) are violated in the last two cases. Three different configurations of Z are evaluated in cases $1 \sim 4$, namely, $Z \sim Ber(0.5)$, $Z \sim U(0, 2)$, and $Z \sim$ a truncated $N(0, 1)$ constrained within $[-2, 2]$. For case 4, the third configuration is replaced by Z, which follows a truncated $N(2, 1)$ constrained within $[0, 4]$ to keep the covariate values positive. The independent censoring variable C is generated from $U(0, a)$ distribution, where $a = 5$ or $a = 20$. The sample size is set to be $n = 100$. The parameter values are set to be $(\eta_0, \beta_0) = (1, 0.5)$ or $(0.5, 1)$.

For each simulation run, the four estimators, namely proposed $\widehat{\beta}^G$, $\widehat{\beta}^{LG}$, the Ghosh–Lin estimator, and the Huang–Wang estimator are calculated. Based on $r = 500$ simulation runs, we report the average bias (Bias) $\sum_{i=1}^{500}\widehat{\beta}_i/500 - \beta_0$, the empirical SE $\sqrt{\sum_{i=1}^{500}(\widehat{\beta}_i - \overline{\beta})^2/499}$, where $\overline{\beta} = \sum_{i=1}^{500}\widehat{\beta}_i/500$, the average of the standard error estimator (SEE) $\sum_{i=1}^{500}\widehat{SE}(\widehat{\beta}_i)/500$ with $\widehat{SE}$ 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, $\widehat{\beta}^{LG}$ outperforms the Ghosh–Lin estimator $\widehat{\beta}^L$. This confirms that the pairwise construction does alleviate

Table 1

Finite-sample performances of four estimators under case 1: $T_k - T_{k-1}$ and and D follows AFT–AFT model with exponential marginals

	Z	$Ber(0.5)$				$U(0, 2)$				trun. $N(0, 1)$			
	(η_0,β_0)	(1,0.5)		(0.5,1)		(1,0.5)		(0.5,1)		(1,0.5)		(0.5,1)	
	a	5	20	5	20	5	20	5	20	5	20	5	20
$\widehat{\beta}^G$	Bias	−0.0143	−0.0186	0.0112	0.0064	0.0093	−0.0085	−0.0083	0.0182	−0.0002	0.0001	−0.0063	−0.0058
	SE	0.3014	0.2918	0.3100	0.3051	0.2436	0.2579	0.3100	0.2812	0.1786	0.1828	0.1927	0.1854
	SEE	0.3236	0.3112	0.3126	0.3290	0.2719	0.2568	0.3135	0.2981	0.1947	0.1839	0.1897	0.1802
	Cov	0.942	0.958	0.928	0.974	0.954	0.920	0.921	0.952	0.950	0.934	0.925	0.934
$\widehat{\beta}^{LG}$	Bias	−0.0140	−0.0167	0.0096	0.0052	0.0086	−0.0068	−0.0075	0.0187	0.0003	0.0012	−0.0058	−0.0049
	SE	0.2931	0.2871	0.3029	0.2981	0.2394	0.2539	0.3003	0.2744	0.1735	0.1803	0.1873	0.1827
	SEE	0.3119	0.3037	0.3047	0.3205	0.2637	0.2536	0.3026	0.2913	0.1874	0.1779	0.1858	0.1773
	Cov	0.944	0.962	0.934	0.970	0.960	0.942	0.920	0.956	0.940	0.936	0.932	0.936
$\widehat{\beta}^L$	Bias	−0.0159	−0.0153	0.0107	0.0062	0.0021	−0.0024	−0.0053	0.0270	−0.0092	−0.0111	0.0037	0.0031
	SE	0.3015	0.3188	0.3194	0.3332	0.2602	0.2785	0.3282	0.3089	0.2144	0.2127	0.2154	0.2083
$\widehat{\beta}^{HW}$	Bias	−0.0241	−0.0029	0.0534	0.0292	0.0098	−0.0301	0.0014	0.0349	−0.0189	0.0119	−0.0084	−0.0091
	SE	0.3516	0.3468	0.4029	0.3787	0.2911	0.3103	0.3843	0.3856	0.2420	0.2628	0.2577	0.2390

Note: $\widehat{\beta}^G$: from (7); $\widehat{\beta}^{LG}$: from (9); $\widehat{\beta}^L$: Ghosh–Lin estimator; $\widehat{\beta}^{HW}$: 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 sizes $n = 100$.

Table 2

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

Z		$Ber(0.5)$				$U(0,2)$				trun. $N(0,1)$			
	(η_0, β_0)	(1,0.5)		(0.5,1)		(1,0.5)		(0.5,1)		(1,0.5)		(0.5,1)	
	a	5	20	5	20	5	20	5	20	5	20	5	20
$\widehat{\beta}^G$	Bias	−0.0037	−0.0096	−0.0003	−0.0145	−0.0015	−0.0111	0.0001	−0.0197	−0.0053	−0.0088	0.0011	−0.0164
	SE	0.0562	0.0606	0.0897	0.0929	0.0480	0.0578	0.0816	0.0953	0.0304	0.0607	0.0529	0.0715
	SEE	0.0569	0.0627	0.0940	0.0989	0.0505	0.0656	0.0897	0.1016	0.0466	0.0616	0.0742	0.0838
	Cov	0.946	0.962	0.962	0.948	0.952	0.972	0.940	0.934	0.976	0.972	0.959	0.976
$\widehat{\beta}^{LG}$	Bias	−0.0024	−0.0121	0.0051	−0.0164	−0.0026	−0.0152	0.0023	−0.0223	−0.0083	−0.0161	0.0026	−0.0249
	SE	0.0796	0.0883	0.1178	0.1252	0.0647	0.0828	0.1093	0.1331	0.0388	0.0858	0.0691	0.0995
	SEE	0.0755	0.0851	0.1166	0.1241	0.0639	0.0868	0.1102	0.1283	0.0551	0.0799	0.0842	0.1039
	Cov	0.928	0.944	0.938	0.936	0.932	0.962	0.921	0.926	0.972	0.952	0.949	0.956
$\widehat{\beta}^L$	Bias	0.1453	0.0128	0.2864	0.0655	0.0601	−0.1066	−0.2237	−0.4111	−0.0056	−0.1645	−0.4965	−0.6539
	SE	0.1805	0.1476	0.1601	0.1387	0.1638	0.1257	0.1325	0.1094	0.1469	0.1071	0.0954	0.0791
$\widehat{\beta}^{HW}$	Bias	1.4823	1.2260	2.2971	2.0038	2.0206	1.3089	2.8447	2.5638	1.5678	1.0895	2.3132	1.9047
	SE	0.2901	0.2083	0.2989	0.2442	0.8708	0.5288	0.6766	0.7318	1.0166	0.7852	0.9294	0.9293

Note: $\widehat{\beta}^G$: from (7); $\widehat{\beta}^{LG}$: from (9); $\widehat{\beta}^L$: Ghosh–Lin estimator; $\widehat{\beta}^{HW}$: 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 sizes $n = 100$.

the drawback of artificial censoring. The Ghosh–Lin estimator produces large biases in cases 3 and 4 under which the imposed model assumptions do not hold. For similar reasons, the Huang–Wang estimator produces large biases in cases 2 to 4 when the underlying assumptions are violated. We also see that the average of the SEE for the proposed estimators is close to the empirical SE and the 95% Cov has accurate coverage probability. This confirms the validity of the resampling approach for variance estimation.

The summary statistics of the censoring proportions and the average number of observed recurrent events per subject for the four settings can be found in Web Appendix B. We also compare the percentages of artificial censoring for the two proposed estimators and the Ghosh–Lin estimator. We first need to define a formula for the proportion of artificial censoring with recurrent events. Recall that $\delta_i^k = I\{T_{ik} \leq X_i\}$, $\tilde{\delta}_i^k = I\{\varepsilon_{ik}(\widehat{\beta}) \leq \tilde{\varepsilon}_i^C(\widehat{\theta})\}$, and $\tilde{\delta}_{i(j)}^k = I\{\varepsilon_{ik}(\widehat{\beta}) \leq \tilde{\varepsilon}_{i(j)}^C(\widehat{\theta})\}$. Hence the pairwise adjusted estimators use $\sum_k \sum_i \sum_{j\neq i} \tilde{\delta}_{i(j)}^k$ pairs of events in the estimating equation. The number of artificially censored pairs is the difference between $\sum_k \sum_i \sum_{j\neq i} \tilde{\delta}_{i(j)}^k$ and $\sum_k \sum_i \sum_{j\neq i} \delta_i^k$, which is the number of uncensored event pairs in the original data set. Hence the proportion of artificially censored pairs in the proposed estimator can be defined as:

$$\text{ACP1} = 1 - \frac{\sum_k \sum_i \sum_{j\neq i} \tilde{\delta}_{i(j)}^k}{\sum_k \sum_i \sum_{j\neq i} \delta_i^k} = 1 - \frac{\sum_k \sum_i \sum_{j\neq i} \tilde{\delta}_{i(j)}^k}{(n-1)\sum_k \sum_i \delta_i^k}.$$

For the Ghosh–Lin estimator without pairwise adjustment, the proportion then becomes

$$\text{ACP2} = 1 - \frac{\sum_k \sum_i \sum_{j\neq i} \tilde{\delta}_i^k}{\sum_k \sum_i \sum_{j\neq i} \delta_i^k} = 1 - \frac{\sum_k \sum_i \tilde{\delta}_i^k}{\sum_k \sum_i \delta_i^k}.$$

The two proposed estimators based on pairwise adjustment yield a much lower proportion of artificial censoring than Ghosh–Lin estimator in all simulation settings. This may explain why the proposed estimators have smaller variances than the Ghosh–Lin estimator. See detailed summary results in Web Appendix B.

We also evaluate the performances of $\widehat{\boldsymbol{\beta}}^G$ and $\widehat{\boldsymbol{\beta}}^{LG}$ in the presence of multiple covariates. When there is only one covariate, we calculated the root $\widehat{\beta}$ based on the bisection approach. When there are multiple covariates, we need to find the roots of several step and nondifferentiable functions. We apply the Nelder–Mead algorithm (Nelder and Mead, 1965) to find the value of $\boldsymbol{\beta}$, which minimizes the norms of the estimating functions: $\|\boldsymbol{U}_2^G(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}})\|$ in (7) and $\|\boldsymbol{U}_2^{LG}(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}})\|$ in (9). Here $\|(x_1, x_2, \ldots;, x_k)'\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_k^2}$ denotes the vector norm. The four simulation settings mentioned earlier are still adopted with $\boldsymbol{Z} = (Z_1, Z_2)'$, where $Z_1 \sim ber(0.5)$, $Z_2 \sim U(0, 2)$, $(\eta_{01}, \eta_{02}) = (0.5, 0.5)$, and $(\beta_{01}, \beta_{02}) = (0.5, 0.5)$. The results based on $r = 500$ simulation runs each with sample size $n = 100$ are reported in Web Appendix B. The two proposed estimators still perform well in the presence of multiple covariates.

Table 3
Performances of $\widehat{\beta}^G$ and $\widehat{\beta}^{LG}$ under model misspecification for D

			$\widehat{\beta}^G$				$\widehat{\beta}^{LG}$			
Data and method	Z	$(\eta_0,\ \beta_0)$	Bias	SE	SEE	Cov	Bias	SE	SEE	Cov
D: AFT–AFT	$Ber(0.5)$	(1,0.5)	−0.0151	0.0807	0.0869	0.956	−0.0104	0.0809	0.0851	0.962
(case 2)		(0.5,1)	−0.0086	0.1087	0.1167	0.956	−0.0112	0.1034	0.1113	0.948
M: AFT–PH	$U(0,\ 2)$	(1,0.5)	−0.0047	0.0809	0.0792	0.942	−0.0040	0.0789	0.0755	0.932
		(0.5,1)	−0.0058	0.1197	0.1354	0.956	−0.0056	0.1173	0.1291	0.952
	$tN(0,\ 1)$	((1,0.5)	−0.0205	0.0631	0.0674	0.944	−0.0224	0.0611	0.0639	0.942
		(0.5,1)	−0.0083	0.0616	0.0638	0.950	−0.0069	0.0602	0.0599	0.946
D: AFT–AFT	$Ber(0.5)$	(1,0.5)	−0.0126	0.0881	0.0912	0.962	−0.0010	0.0826	0.0868	0.950
(case 2)		(0.5,1)	0.0365	0.1137	0.1184	0.942	0.0377	0.1123	0.1144	0.940
M: AFT–Loc	$U(0,\ 2)$	(1,0.5)	−0.0173	0.0704	0.0721	0.940	−0.0091	0.0684	0.0672	0.948
		(0.5,1)	0.0584	0.1164	0.1349	0.938	0.0560	0.1132	0.1290	0.944
	$tN(0,\ 1)$	(1,0.5)	−0.0506	0.0606	0.0594	0.858	−0.0378	0.0577	0.0573	0.886
		(0.5,1)	0.0368	0.0604	0.0612	0.934	0.0289	0.0582	0.0573	0.930
D: AFT–PH	$Ber(0.5)$	(1,0.5)	0.0544	0.1379	0.1398	0.928	0.0459	0.1322	0.1319	0.930
(case 3)		(0.5,1)	0.0210	0.1401	0.1509	0.960	0.0161	0.1309	0.1395	0.948
M: AFT–AFT	$U(0,\ 2)$	(1,0.5)	0.0835	0.1254	0.1320	0.900	0.0703	0.1180	0.1204	0.914
		(0.5,1)	0.0318	0.1495	0.1494	0.936	0.0246	0.1374	0.1364	0.930
	$tN(0,\ 1)$	(1,0.5)	0.0366	0.0786	0.0777	0.916	0.0317	0.0759	0.0730	0.930
		(0.5,1)	0.0159	0.0882	0.0863	0.930	0.0128	0.0817	0.0811	0.942

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

Table 3 reports the results for checking the robustness of $\widehat{\beta}^G$ and $\widehat{\beta}^{LG}$ in (7) and (9) with $h_2(t)$ misspecified. True data are generated either from the AFT–AFT model or the AFT–PH model with one covariate. The data are analyzed based on three model alternatives: namely, AFT–PH, AFT–Loc, and AFT–AFT models. We report the bias, SE, SEE, and empirical coverage probability of the 95% confidence intervals for $\widehat{\beta}^G$ and $\widehat{\beta}^{LG}$ with $C \sim U(0, a = 20)$ based on $r = 500$ simulation runs. The results show that the performances of $\widehat{\beta}^G$ and $\widehat{\beta}^{LG}$ are robust when fitting AFT–PH model to AFT–AFT data. They are also somewhat robust most of the time in the other two settings. But for the last two settings (fitting AFT–Loc to AFT–AFT data, and fitting AFT–AFT to AFT–PH data), they do produce larger bias and sometimes result in Covs with coverage probabilities significantly lower than the nominal level of 95%.

Finally, we evaluate the proposed model-checking procedure when data follow the AFT–AFT combination with one covariate based on $r = 400$ replications with $B = 400$ resampling times within each run. The results are summarized in Web Appendix B. The tests based on $\sup_t \|S_1(t;\widehat{\eta})\|$ and $\sup_t \|S_2(t;\widehat{\theta})\|$ are both conservative. That is, the percentages of falsely rejecting the correct AFT assumption (type I error) are less than the nominal level 0.05 in all cases. For the first test, the powers under the PH model assumption are relatively low (15% and 11%) but become higher (65% and 77%) under the Loc model. This is because the AFT model is more similar to the PH model than the Loc model. As we have seen in the robustness study, wrongly fitting AFT–PH model to AFT–AFT data still yields reasonably good performance for $\widehat{\beta}^G$ and $\widehat{\beta}^{LG}$. For the second test, the power of rejecting the Loc assumption for recurrent event times is large, above 90%.

We also evaluate the model-selection procedure. There are six choices of model combinations, namely, AFT–AFT, AFT–PH, AFT–Loc, Loc–AFT, Loc–PH, and Loc–Loc when the true data are generated from AFT–AFT models. We first select among the three models on survival time (2): AFT, PH, and Loc by choosing the one with the highest p-value based on $\sup_t \|S_1(t;\widehat{\eta})\|$; then we select the recurrence events model with the highest p-value based on $\sup_t \|S_2(t;\widehat{\theta})\|$. Table 4 reports the proportion of each model combination being selected based on $r = 400$ simulation runs with $B = 400$ resampling times within each run. We see that the model-selection procedure does choose the correct AFT–AFT model combination most of the time (80% and 86% of the time). The AFT–PH model combination, which still provides good estimates in the robustness study, is selected in another 10% to 11% of the time. Other worse choices are rarely selected.

Table 4
Performance of the proposed model-selection procedure when data follow AFT–AFT models

	Selected proportions					
	AFT–AFT (true)	AFT–PH	AFT–Loc	Loc–AFT	Loc–PH	Loc–Loc
$Z \sim Ber(0.5)$	0.795	0.115	0.088	0	0	0.003
$Z \sim U(0, 2)$	0.863	0.108	0.030	0	0	0

Note: Sample sizes $n = 200$, resampling times $B = 400$, and replications $r = 400$.

Comparing the two proposed estimators, $\widehat{\beta}^{LG}$ is slightly less variable than $\widehat{\beta}^G$ in cases 1 to 3. However under case 4, the standard deviation of $\widehat{\beta}^{LG}$ is much larger than $\widehat{\beta}^G$ in case 4. So no one estimator is always better than the other. We suggest choosing the one with smaller estimated SE. Notice that in the simulations, SEE provides reliable information about the magnitude of SE (when one estimator has smaller SE, it also has smaller SEE in all cases).

Table 5
Summary of ALIVE cohort data

HIV status	Number of subjects	Number of hospitalizations							Number of deaths
		0	1	2	3	4	5	≥6	
Negative	746	294	98	65	68	76	33	112	23
Positive	297	70	40	31	34	35	20	67	47

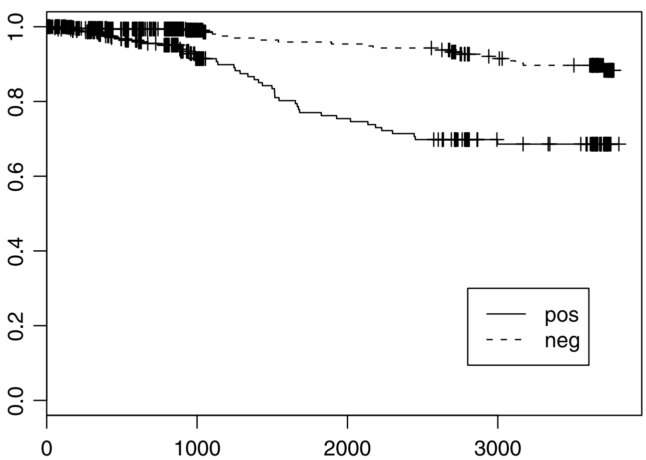

Figure 1. Kaplan–Meier survival curves for the HIV-negative and HIV-positive groups.

6. Data Analysis

The proposed methodology is applied to the AIDS linked to the intravenous experiences (ALIVE) cohort study. Covariates include the information of inpatient admissions, HIV status, and other variables from a group of injection drug users in Baltimore. This study is a prospective cohort study with semiannual follow-up visits to gather clinical, behavioral, and laboratory data. Self-reported repeated hospitalizations were systematically recorded since February 1, 1988. Here we analyze the recent 10 years data, collected from January 1, 1998 to June 30, 2008, which contains 1043 injection drug users. Note that we exclude five subjects whose HIV status changed from negative to positive during the study period. The HIV status is coded as a binary covariate with 1 indicating HIV negative and 0 indicating HIV positive.

Table 5 provides a brief summary of the data including the number of hospitalizations and the number of deaths for users with different HIV status. The Kaplan–Meier estimators of the survival functions are plotted in Figure 1. For example, the 5-year survival probabilities are 0.959 for HIV-negative subjects and 0.762 for HIV-positive subjects. The value of log-rank statistic is 36.1 with p-value equal to 1.9×10^{-9}. Thus, the difference of the two survival curves is significant.

The first step is to choose a suitable regression model for the survival time. Based on 1000 resampling runs, $\sup_t \|S_1(t;\widehat{\eta})\| = 0.0442$ with p-value $= 0.991$ for the AFT model, $\sup_t \|S_1(t;\widehat{\eta})\| = 0.1069$ with p-value $= 0.711$ for the Cox PH model, and $\sup_t \|S_1(t;\widehat{\eta})\| = 0.2263$ with p-value $=$ 0.008 for the Loc model. Thus we fit the AFT model for D and then choose between the AFT and Loc models for T_k. Based on 1000 replications, $\sup_t \|S_2(t;\widehat{\theta})\| = 0.7473$ with p-value $= 0.374$ for AFT model and $\sup_t \|S_2(t;\widehat{\theta})\| = 0.7943$ 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 $S_1(t;\widehat{\eta})$ and $S_2(t;\widehat{\theta})$ with 20 realizations of $\widehat{S}_1(t)$ and $\widehat{S}_2(t)$ 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 $\widehat{\beta}^G$, 1.07 based on $\widehat{\beta}^{LG}$, and 1.05 based on the Ghosh–Lin estimator). Among the three estimators, $\widehat{\beta}^G$ gives the smallest estimated SE. Covs for all three estimators $\widehat{\beta}$ 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 $h_1(\,\cdot\,)$ 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

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting process: A large sample study. *Annals of Statistics* **10,** 1100–1120.

Chang, S. H. (2000). A two-sample comparison for multiple ordered event data. *Biometrics* **56,** 183–189.

Chang, S. H. and Wang, M. C. (1999). Condition regression analysis for recurrence time data. *Journal of the American Statistical Association* **94,** 1221–1230.

Chen, K., Jin, Z., and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89,** 659–668.

Cook, R. J. and Lawless, J. F. (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* **16,** 911–924.

Ding, A. A., Wang, W., Hsieh, J. J., and Shi, G. (2009). Marginal regression analysis for semi-competing risks data under dependent censoring. *Scandinavian Journal of Statistics* **36,** 481–500.

Ghosh, D. and Lin, D. Y. (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* **59,** 877–885.

Huang, C. Y. and Wang, M. C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* **99,** 1153–1165.

Huang, Y. (2000). Two-sample multistate accelerated sojourn times model. *Journal of the American Statistical Association* **95,** 619–627.

Kalbfleisch and Prentice (2002). *The Statistical Analysis of Failure Time Data*, 2nd edition. New York: Wiley.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37,** 158–168.

Lin, D. Y., Robins, J. M., and Wei, L. J. (1996). Comparing two failure time distributions in the presence of dependent censoring. *Biometrika* **83,** 381–393.

Lin, D. Y., Wei, L. J., and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85,** 605–618.

Liu, L., Wolfe, R. A., and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60,** 747–756.

Miloslavsky, M., Keles, S., van der Laan, M. J., and Butler, S. (2004). Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *Journal of the Royal Statistical Society, Series B* **66,** 239–257.

Murphy, S. A., Rossini, A. J., and Van Der Waart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association* **92,** 968–976.

Nelder, J. A. and Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal* **7,** 308–313.

Parzen, M. I., Wei, L. J., and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81,** 341–350.

Peng, L. and Fine, J. P. (2006). Rank estimation of accelerated lifetime models with dependent censoring. *Journal of the American Statistical Association* **101,** 1085–1093.

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

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68,** 373–379.

Schaubel, D. E. and Cai, J. (2004). Regression methods for gap time hazard functions for sequentially ordered multivariate failure time data. *Biometrika* **91,** 291–303.

Wang, M-C., Qin, J., and Chiang, C-T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association* **96,** 1057–1065.

Ye, Y., Kalbfleisch, J., and Schaubel, D. E. (2007). Semiparametric analysis for correlated recurrent and terminal events. *Biometrics* **63,** 78–87.

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

Zeng, D. and Lin, D. Y. (2007a). Maximum likelihood estimation in semiparametric models with censored data (with discussion). *Journal of the Royal Statistical Society, Series B* **69,** 1–30.

Zeng, D. and Lin, D. Y. (2007b). Semiparametric transformation models with random effects for recurrent events. *Journal of the American Statistical Association* **102,** 167–180.

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

$$\boldsymbol{U}_1^L(\boldsymbol{\eta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\infty \frac{\sum_{j=1}^n I\{\tilde{\xi}_j(\boldsymbol{\eta}) \geq t\}(\boldsymbol{Z}_i - \boldsymbol{Z}_j)}{\sum_{l=1}^n I\{\tilde{\xi}_l(\boldsymbol{\eta}) \geq t\}} dN_{\xi i}(t;\boldsymbol{\eta})$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^n(\boldsymbol{Z}_i - \boldsymbol{Z}_j)\int_0^\infty \frac{I\{\tilde{\xi}_j(\boldsymbol{\eta}) \geq t\}}{\sum_{l=1}^n I\{\tilde{\xi}_l(\boldsymbol{\eta}) \geq t\}} dN_{\xi i}(t;\boldsymbol{\eta}).$$

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

$$\boldsymbol{U}_1^G(\boldsymbol{\eta}) = \frac{2\sqrt{n}}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n(\boldsymbol{Z}_i - \boldsymbol{Z}_j)I\{\tilde{\xi}_i(\boldsymbol{\eta}) \leq \tilde{\xi}_j(\boldsymbol{\eta}), \delta_i = 1\}$$

$$= \frac{4\sqrt{n}}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n(\boldsymbol{Z}_i - \boldsymbol{Z}_j)$$

$$\times \int_0^\infty \frac{I\{\tilde{\xi}_j(\boldsymbol{\eta}) \geq t\}}{I\{\tilde{\xi}_i(\boldsymbol{\eta}) \geq t\} + I\{\tilde{\xi}_j(\boldsymbol{\eta}) \geq t\}} dN_{\xi i}(t;\boldsymbol{\eta}).$$

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 $\boldsymbol{\eta}$ in Ying (1993).

C1: The parameter space $\mathcal{P}$ for $\boldsymbol{\beta}$ is compact, and true parameter $\boldsymbol{\beta}_0$ is an interior point of $\mathcal{P}$.

C2: $\boldsymbol{\theta}_0$ is the unique solution to (11).

C3: $|\boldsymbol{Z}|$ is bounded. Conditional on $\boldsymbol{Z}$, the conditional densities of ξ, C, and ε_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: $E[\boldsymbol{U}(\boldsymbol{\theta})]$ is differentiable and the Jacobian matrix is nonsingular at the true parameter value $\boldsymbol{\theta}_0$.

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 ∞.

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.