Lifetime Data Analysis, 11, 389–404, 2005

Marginal Regression of Multivariate Event Times Based on Linear Transformation Models

WENBIN LU lu@stat.ncsu.edu
Department of Statistics, North Carolina State University, 2501 Founder's Drive, Raleigh, NC 27695, USA

Received August 30, 2004; Accepted March 2, 2005

Abstract. Multivariate event time data are common in medical studies and have received much attention recently. In such data, each study subject may potentially experience several types of events or recurrences of the same type of event, or event times may be clustered. Marginal distributions are specified for the multivariate event times in multiple events and clustered events data, and for the gap times in recurrent events data, using the semiparametric linear transformation models while leaving the dependence structures for related events unspecified. We propose several estimating equations for simultaneous estimation of the regression parameters and the transformation function. It is shown that the resulting regression estimators are asymptotically normal, with variance–covariance matrix that has a closed form and can be consistently estimated by the usual plug-in method. Simulation studies show that the proposed approach is appropriate for practical use. An application to the well-known bladder cancer tumor recurrences data is also given to illustrate the methodology.

Keywords: cluster event times, estimating equations, informative cluster size, linear transformation models, multiple events, recurrent events

1. Introduction

Multivariate event time data are commonly encountered in many medical studies because each study subject can potentially experience multiple events or failures or the events times may be grouped or clustered, which leads to dependencies within the same cluster. As usual, we refer to the former situation as multiple events data and the latter as clustered events data. A very special case of multiple events data is recurrent events data, in which subjects often experience repeated occurrences of the same type of event. Since the dependence structure among the event times for the same subject or cluster is often complicated, much of the focus has been on modeling the marginal distributions of event times, e.g. Wei et al. (1989). Lin (1994) provided a review of the marginal approach based on Cox-regression models (Cox, 1972). More recent research was conducted by Pepe and Cai (1993), Lawless et al. (1997), Prentice and Hsu (1997), Spiekerman and Lin (1998) and Lin et al. (2000) among others.

As noted by many authors, the proportional hazards model may not be appropriate for modeling survival times in some medical studies. For example, if the hazard functions for the two treatment groups converge to the same limit, the proportional odds model is more preferable than the proportional hazards model to fit such data. See Pettitt (1982, 1984), Bennett (1983), Dabrowska and Doksum (1988) and Murphy et al. (1997).

More generally, a class of linear transformation models (Clayton and Cuzick, 1985; Bickel et al., 1993; Cheng et al., 1995; Fine et al., 1998) may be used for survival times. The semiparametric linear transformation model is specified by

$$H(T) = -\beta' Z + \epsilon, \tag{1}$$

where H is an unknown monotone increasing function, β a p-dimensional regression parameter vector and ϵ the error term with a known continuous distribution that is independent of censoring variable C and covariate vector Z. If ϵ is chosen to follow the extreme value distribution, then (1) becomes the proportional hazards model. On the other hand, if ϵ follows the logistic distribution, then it becomes the proportional odds model.

For univariate failure time data, a number of methods have been proposed for analysis of the linear transformation models. Specifically, Cheng et al. (1995) proposed a general estimating equation approach, which was further developed by Cheng et al. (1997) and Fine et al. (1998). A key assumption in their methods is that the censoring times are independent of covariates, which is often too restrictive. More recently, Chen et al. (2002) made use of a martingale integral representation in constructing estimating equations, which does not require the common censoring distribution assumption.

In this article, we study the marginal linear transformation models for multivariate event time data. Our approach is motivated by the recent work of Chen et al. (2002). Several estimating equations are proposed for simultaneous estimation of the regression parameters and the transformation function. The resultant estimators are proven to be consistent and asymptotically normal. Furthermore, the asymptotic variance–covariance matrix has a closed form and can be consistently estimated by the usual plug-in method.

The extension of Chen et al.'s (2002) approach to multiple events data is straightforward, like the usual WLW method for the marginal proportional hazards model (Wei et al., 1989), and is sketched in Section 2. Section 3 presents the models and the corresponding inference procedures for clustered events data when the cluster size is both noninformative and informative. Models and methods for recurrent events data are discussed in Section 4. Section 5 is devoted to numerical studies. Some concluding remarks are given in Section 6. All the proof are relegated to the Appendix.

2. Multiple Events Data

Suppose each study subject can potentially experience K types of events or failures, where K is a fixed integer. Let T_{ki} be the time to the kth event of the ith subject, where $i=1,\ldots,n$; $k=1,\ldots,K$. In addition, let C_{ki} be the corresponding censoring time, and $Z_{ki} \equiv (Z_{1ki},\ldots, Z_{pki})'$ be the corresponding p-dimensional vector of covariates. We assume that $(T_{1i},\ldots, T_{Ki})$ is independent of $(C_{1i},\ldots, C_{Ki})$ given $(Z_{1i},\ldots, Z_{Ki})$. The observed data consist of $(\tilde{T}_{ki}, \delta_{ki}, Z_{ki})(k = 1, \ldots, K; i = 1, \ldots, n)$, where $\tilde{T}_{ki} = T_{ki} \wedge C_{ki}$ and $\delta_{ki} = I(T_{ki} \leq C_{ki})$. Here and thereafter, $a \wedge b = \min(a, b)$, and $I(\cdot)$ is the indicator function.

The marginal distributions of K types of event times are formulated with the linear transformation models. That is

$$H_k(T_{ki}) = -\beta_k' Z_{ki} + \epsilon_{ki}, \qquad i = 1, \ldots, n; \quad k = 1, \ldots, K, \tag{2}$$

where β_k is a p-dimensional vector of unknown regression parameters, H_k is an unknown monotone increasing function, and ϵ_{ki} is the error term with a known continuous distribution that is independent of censoring variable C_{ki} and covariate vector Z_{ki}. In addition, $(\epsilon_{1i},\ldots, \epsilon_{Ki})$ $(i=1,\ldots, n)$ are independent random vectors with a common joint distribution. For example, if $K=2$, the joint distribution of $(\epsilon_{1i},\epsilon_{2i})$ $(i=1,\ldots, n)$ can be specified by the Gumbel (1960)'s bivariate distribution, i.e. $F(x_1, x_2) = F_1(x_1)F_2(x_2)[1 + \theta\{1 - F_1(x_1)\}\{1 - F_2(x_2)\}]$, where $-1 \leq \theta \leq 1$, and $F_1(\cdot)$ and $F_2(\cdot)$ are the two known marginal distributions for ϵ_{1i} and ϵ_{2i}, respectively.

In model (2), the event times of same type follows the usual linear transformation model. However, the event times from different types may have different distributions. For example, the event times from the first type may come from the proportional hazards model while the event times from the second type may come from the proportional odds model. Obviously, Chen et al.'s (2002) method can be applied to event times from a single type. In addition, using the method of Wei et al. (1989), it can be also generalized to carry out simultaneous inference on the regression parameter $\beta = (\beta_1', \ldots, \beta_K')'$. Since the derivation is straightforward, the details are omitted here.

3. Clustered Events Data

3.1. *The Cluster Size is Noninformative*

Suppose that a random sample of n clusters is chosen and there are K_i $(i=1,\ldots,n)$ members in the ith cluster. Let T_{ik} and C_{ik} be the event time and censoring time for the kth member of the ith cluster, and let Z_{ik} denote the corresponding p-dimensional vector of covariates. We assume that $T_i \equiv (T_{i1}, \ldots, T_{iK_i})'$ and $C_i \equiv (C_{i1}, \ldots, C_{iK_i})'$ are independent conditional on $Z_i \equiv (Z_{i1}', \ldots, Z_{iK_i}')'$. And the cluster size K_i is assumed to be independent of T_i and C_i, i.e. noninformative, and be small relative to n. Then the observed data consist of $(\tilde{T}_{ik}, \delta_{ik}, Z_{ik})(k = 1, \ldots, K_i; i = 1, \ldots, n)$, where $\tilde{T}_{ik} = T_{ik} \wedge C_{ik}$ and $\delta_{ik} = I(T_{ik} \leq C_{ik})$.

We specify the marginal distributions of T_{ik} with the linear transformation models. That is

$$H(T_{ik}) = -\beta' Z_{ik} + \epsilon_{ik}, \qquad k = 1, \ldots, K_i; \quad i = 1, \ldots, n, \tag{3}$$

where β is a p-dimensional vector of unknown regression parameters, H is an unknown monotone increasing function, and ϵ_{ik} is the error term with a known continuous distribution that is independent of censoring variable C_{ik} and covariate vector Z_{ik}. In addition, (ϵ_{i1},..., ϵ_{iK_i}) (i = 1,..., n) are independent random vectors. For each i, the error terms ϵ_{i1},..., ϵ_{iK_i} are potentially correlated, but assumed to be exchangeable with a common specified marginal distribution. And for any i and j, and $K \leq K_i \wedge K_j$, the vectors (ϵ_{i1},..., ϵ_{iK}) and (ϵ_{j1},..., ϵ_{jK}) have the same distribution. Let M denote the common cumulative hazard function for ϵ_{ik} (k = 1,..., K_i; i = 1,..., n), i.e. $P(\epsilon_{ik} > t) = \exp\{-\Lambda(t)\}$.

Define the usual counting process $N_{ik}(t) = \delta_{ik} I(\tilde{T}_{ik} \leq t)$ and at-risk process $Y_{ik}(t) = I(\tilde{T}_{ik} \geq t)$. And let

$$M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(s) d\Lambda\{H_0(s) + \beta_0' Z_{ik}\}, \qquad k = 1, \ldots, K_i; \quad i = 1, \ldots, n, \tag{4}$$

where (β_0, H_0) are the true values of (β,H). It is easy to show that $M_{ik}(t)$ is a mean zero process. Therefore, we proposed the following two estimating equations for H and β, respectively

$$\sum_{i=1}^{n} \sum_{k=1}^{K_i} [dN_{ik}(t) - Y_{ik}(t) d\Lambda\{H(t) + \beta' Z_{ik}\}] = 0, \qquad t \geq 0, \quad H(0) = -\infty, \tag{5}$$

$$\sum_{i=1}^{n} \sum_{k=1}^{K_i} \int_0^{\tau} Z_{ik} [dN_{ik}(t) - Y_{ik}(t) d\Lambda\{H(t) + \beta' Z_{ik}\}] = 0. \tag{6}$$

where τ denote the duration of the study. And we assume that $P(\tilde{T}_{ik} > \tau | Z_{ik}) > \epsilon$ for any fixed value of Z_{ik}, where ϵ is a positive constant. For any fixed β, the solution to (5) is unique, denoted by $\hat{H}(\cdot; \beta)$. Here like Chen et al. (2002), (5) can be solved in two different ways to obtain $\hat{H}(\cdot; \beta)$. The first method is based on the numerical differences of transformation function H at the observed failure times; while the second method directly gives a closed solution by replacing $d\Lambda\{H(t) + \beta' Z_{ik}\}$ in (5) with $\lambda\{H(t-) + \beta' Z_{ik}\} dH(t)$. The two different methods give equivalent solutions in theory. However, from our experience in numerical studies, the first method is more stable than the second one for obtaining $\hat{H}$ in finite sample. Therefore, in this paper, only the first method was applied to solve (5). Then the estimator $\hat{\beta}$ of β can be obtained by solving the following estimating equation

$$\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_0^{\tau} Z_{ik}[dN_{ik}(t) - Y_{ik}(t)d\Lambda\{\hat{H}(t;\beta) + \beta' Z_{ik}\}] = 0. \tag{7}$$

We first introduce some notations and then the consistency and asymptotic normality of $\hat{\beta}$ will be established in the following theorem. For any t, $s \in (0, \tau]$, define

$$B_2(t) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\lambda\{H_0(t) + \beta_0' Z_{ik}\}Y_{ik}(t),$$

$$b_1(t) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\dot{\lambda}\{H_0(t) + \beta_0' Z_{ik}\}Y_{ik}(t),$$

$$B(t,s) = \exp\left\{\int_s^t \frac{b_1(u)}{B_2(u)}dH_0(u)\right\},$$

$$\mu_Z(t) = \frac{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K_i} Z_{ik}\lambda\{H_0(\tilde{T}_{ik}) + \beta_0' Z_{ik}\}Y_{ik}(t)B(t,\tilde{T}_{ik})}{B_2(t)},$$

and

$$A = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_0^{\tau}\{Z_{ik} - \mu_Z(t)\}Z_{ik}'\dot{\lambda}\{H_0(t) + \beta_0' Z_{ik}\}Y_{ik}(t)dH_0(t), \tag{8}$$

$$\Sigma = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[\sum_{k=1}^{K_i}\int_0^{\tau}\{Z_{ik} - \mu_Z(t)\}dM_{ik}(t)\sum_{k'=1}^{K_i}\int_0^{\tau}\{Z_{ik'} - \mu_Z(s)\}'dM_{ik'}(s)\right], \tag{9}$$

Assume that A and T are finite and nonsingular.

THEOREM 1 *Under suitable regularity conditions, we have that*

$$n^{\frac{1}{2}}(\hat{\beta} - \beta_0) \to N\{0, A^{-1}\Sigma(A^{-1})'\} \tag{10}$$

in distribution, as $n \to \infty$. Moreover, A *and* T *can be consistently estimated by*

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_0^{\tau}\{Z_{ik} - \bar{Z}(t)\}Z_{ik}'\dot{\lambda}\{\hat{H}(t) + \hat{\beta}'Z_{ik}\}Y_{ik}(t)d\hat{H}(t),$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{k=1}^{K_i}\int_0^{\tau}\{Z_{ik} - \bar{Z}(t)\}d\hat{M}_{ik}(t)\sum_{k'=1}^{K_i}\int_0^{\tau}\{Z_{ik'} - \bar{Z}(s)\}'d\hat{M}_{ik'}(s)\right],$$

respectively, where

$$\bar{Z}(t) = \frac{\sum_{i=1}^{n}\sum_{k=1}^{K_i} Z_{ik}\lambda\{\hat{H}(\tilde{T}_{ik}) + \hat{\beta}'Z_{ik}\}Y_{ik}(t)\hat{B}(t,\tilde{T}_{ik})}{\sum_{i=1}^{n}\sum_{k=1}^{K_i}\lambda\{\hat{H}(t) + \hat{\beta}'Z_{ik}\}Y_{ik}(t)},$$

$$\hat{B}(t,s) = \exp\left(\int_s^t \frac{\sum_{i=1}^{n}\sum_{k=1}^{K_i}\dot{\lambda}\{\hat{H}(u) + \hat{\beta}'Z_{ik}\}Y_{ik}(u)}{\sum_{i=1}^{n}\sum_{k=1}^{K_i}\lambda\{\hat{H}(u) + \hat{\beta}'Z_{ik}\}Y_{ik}(u)}d\hat{H}(u)\right),$$

$$\hat{M}_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(s)d\Lambda\{\hat{H}(s) + \hat{\beta}'Z_{ik}\}.$$

for s, $t \in [0,\tau]$.

Similarly, a natural estimator for H_0 is $\hat{H}(\cdot) = \hat{H}(\cdot;\hat{\beta})$, which can be shown to be consistent and converge weakly to a Gaussian process.

3.2. The Cluster Size is Informative

When cluster size is informative, the proposed estimating equations (5) and (6) are no longer valid. Different methods have been proposed for the marginal analysis of clustered data when cluster size is informative (e.g. the within-cluster resampling approach of Hoffman, et al. (2001) and the cluster weighted generalized estimating equations approach of Williamson, et al. (2003)). Here like Williamson, et al. (2003), we propose several weighted estimating equations based on (5) and (6) for the analysis of clustered events data. To be specific, use the following two estimating equations for H and β, respectively

$$\sum_{i=1}^{n}\frac{1}{K_i}\sum_{k=1}^{K_i}[dN_{ik}(t) - Y_{ik}(t)d\Lambda\{H(t) + \beta'Z_i\}] = 0, \quad t \geq 0, \quad H(0) = -\infty, \quad (11)$$

$$\sum_{i=1}^{n}\frac{1}{K_i}\sum_{k=1}^{K_i}\int_0^{\tau} Z_i[dN_{ik}(t) - Y_{ik}(t)d\Lambda\{H(t) + \beta'Z_i\}] = 0. \quad (12)$$

The large sample results for the resulting estimator $\hat{\beta}$ will be established in Theorem 2 below after introducing some notations. For any t, $s \in (0, \tau]$, define

$$B_2(t) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{K_i}\sum_{k=1}^{K_i} \lambda\{H_0(t) + \beta_0' Z_i\} Y_{ik}(t),$$

$$b_1(t) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{K_i}\sum_{k=1}^{K_i} \dot{\lambda}\{H_0(t) + \beta_0' Z_i\} Y_{ik}(t),$$

$$B(t,s) = \exp\left\{\int_s^t \frac{b_1(u)}{B_2(u)} dH_0(u)\right\},$$

$$\mu_Z(t) = \frac{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K_i}\sum_{k=1}^{K_i} Z_i\lambda\{H_0(\tilde{T}_{ik})+\beta_0' Z_i\} Y_{ik}(t) B(t,\tilde{T}_{ik})}{B_2(t)},$$

and

$$A = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{K_i}\sum_{k=1}^{K_i} \int_0^\tau \{Z_i - \mu_Z(t)\} Z_i' \dot{\lambda}\{H_0(t) + \beta_0' Z_i\} Y_{ik}(t) dH_0(t), \tag{13}$$

$$\Sigma = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \left[\frac{1}{K_i}\sum_{k=1}^{K_i} \int_0^\tau \{Z_i - \mu_Z(t)\} dM_{ik}(t) \frac{1}{K_i}\sum_{k'=1}^{K_i} \int_0^\tau \{Z_i - \mu_Z(s)\}' dM_{ik'}(s)\right], \tag{14}$$

Assume that A and T are finite and nonsingular.

THEOREM 2 *Under suitable regularity conditions, we have that*

$$n^{\frac{1}{2}}(\hat{\beta} - \beta_0) \to N\{0, A^{-1}\Sigma(A^{-1})'\} \tag{15}$$

in distribution, as $n \to \infty$. Moreover, A and T can be consistently estimated by

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{K_i}\sum_{k=1}^{K_i} \int_0^\tau \{Z_i - \bar{Z}(t)\} Z_i' \dot{\lambda}\{\hat{H}(t) + \hat{\beta}' Z_i\} Y_{ik}(t) d\hat{H}(t),$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \left[\frac{1}{K_i}\sum_{k=1}^{K_i} \int_0^\tau \{Z_i - \bar{Z}(t)\} d\hat{M}_{ik}(t) \frac{1}{K_i}\sum_{k'=1}^{K_i} \int_0^\tau \{Z_i - \bar{Z}(s)\}' d\hat{M}_{ik'}(s)\right],$$

respectively, where

$$\bar{Z}(t) = \frac{\sum_{i=1}^{n} \frac{1}{K_i} \sum_{k=1}^{K_i} Z_i \lambda\{\hat{H}(\tilde{T}_{ik}) + \hat{\beta}^i\} Y_{ik}(t) \hat{B}(t, \tilde{T}_{ik})}{\sum_{i=1}^{n} \frac{1}{K_i} \sum_{k=1}^{K_i} \lambda\{\hat{H}(t) + \hat{\beta}' Z_i\} Y_{ik}(t)},$$

$$\hat{B}(t,s) = \exp\left(\int_s^t \frac{\sum_{i=1}^{n} \frac{1}{K_i} \sum_{k=1}^{K_i} \dot{\lambda}\{\hat{H}(u) + \hat{\beta}' Z_i\} Y_{ik}(u)}{\sum_{i=1}^{n} \frac{1}{K_i} \sum_{k=1}^{K_i} \lambda\{\hat{H}(u) + \hat{\beta}' Z_i\} Y_{ik}(u)} d\hat{H}(u) \right),$$

$$\hat{M}_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(s) d\Lambda\{\hat{H}(s) + \hat{\beta}' Z_i\}.$$

for s, $t \in [0,\tau]$.

4. Recurrent Events Data

Let T^*_{ik} be the gap time between the $(k-1)$th and kth events on the ith subject, where $i=1,\ldots, n$ and $k=1,2,\ldots$. And let C_i be the corresponding follow-up or censoring time and Z_i the p-dimensional vector of covariates. Assume that C_i is independent of T^*_{ik} $(k=1,2,\ldots)$ conditional on Z_i. For $i=1,\ldots,n$, let M_i be the number of observed events for subject i, i.e. $\sum_{j=1}^{M_i} T^*_{ij} \le C_i$ and $\sum_{j=1}^{M_i+1} T^*_{ij} > C_i$, where $\sum_1^0 \equiv 0$. Then the observed data consist of $(T^*_{ij} : j = 1, \ldots, M_i; C_i; Z_i)$, which are n iid copies of $(T^*_j : j = 1, \ldots, M; C; Z)$.

We formulate the marginal distribution of gap time T^*_{ik} with the linear transformation models. That is

$$H(T^*_{ik}) = -\beta' Z_i + \epsilon_{ik}, \qquad k = 1, 2, \ldots; \quad i = 1, \ldots, n; \tag{16}$$

where β is a p-dimensional vector of unknown regression parameters, H is an unknown monotone increasing function, and ϵ_{ik} is the error term with a known continuous distribution that is independent of censoring variable C_i and covariate vector Z_i. In addition, $(\epsilon_{i1},\epsilon_{i2},\ldots)$ $(i=1,\ldots, n)$ are n iid random vectors. For each i and any $k \neq j$, the error terms ϵ_{ik} and ϵ_{ij} are potentially correlated, but assumed to be exchangeable with a common specified marginal distribution. Let M denote the common cumulative hazard function for ϵ_{ik} $(k=1,2,\ldots;\ i=1,\ldots,\ n)$, i.e. $P(\epsilon_{ik} > t) = \exp\{-\Lambda(t)\}$. If $\mathrm{M}(t)=\exp(t)$, then (16) specifies the proportional hazards model for the gap times, which was the model studied by Huang and Chen (2003).

Like Wang and Chang (1999) and Huang and Chen (2003), we establish a connection between a subset of the observed gap times and clustered events data. A key observation of Wang and Chang (1999) is that, for individual i and given C_i, M_i, $T^+_{i(M_i)} \equiv C_i - \sum_{j=1}^{M_i} T^*_{ij}$, the observed complete gap times, T^*_{ij}, $j=1,\ldots,M_i$ are identically distributed, which suggests that we can treat a subset of observed gap times as clustered events data with informative cluster size. Specifically, define $\delta_{ik} = I(M_i \geq 1)$, $K_i = \max(M_i, 1)$, and

$$T_{ik} = \{ T^*_{ik} \quad \text{if } \delta_{ik} = 1,_i \quad \text{if } \delta_{ik} = 0.$$

where $k = 1,\ldots,K_i$ and $i = 1,\ldots,n$. Then the subset consist of $(T_{ik}, \delta_{ik}, Z_i)$ $(k = 1,\ldots,K_i;$ $i = 1,\ldots,n)$.

Since the first gap time is subject to independent censoring, the estimating equations proposed by Chen et al. (2002) can be applied to the time-to-the-first-event data, i.e. $(T_{i1}, \delta_{i1}, Z_i)$, $i = 1,\ldots,n$. Specifically, the estimating equations for H and β are given by

$$\sum_{i=1}^{n}[dN_{i1}(t) - Y_{i1}(t)d\Lambda\{H(t) + \beta' Z_i\}] = 0, \qquad t \geq 0, \quad H(0) = -\infty, \tag{17}$$

and

$$\sum_{i=1}^{n}\int_0^{\tau} Z_i[dN_{i1}(t) - Y_{i1}(t)d\Lambda\{H(t) + \beta' Z_i\}] = 0. \tag{18}$$

respectively, where $N_{ik}(t)$ and $Y_{ik}(t)$ $(k = 1,\ldots,K_i;\ i = 1,\ldots,n)$ are the counting and at-risk processes defined as before. Let $\hat{\beta}_{(1)}$ be the estimator for β obtained from (17) and (18). Then $\hat{\beta}_{(1)}$ is known to be consistent and asymptotically normal. However, it may lose much efficiency since only the first event times are used.

As noted by Huang and Chen (2003), the first event time may be replaced by a random choice from the same cluster. Thus, weighted estimating equations, like (11) and (12), among different gap times within the same cluster may yield more efficient estimation. It is easy to show that the limits of $1/n$ multiplying the left-hand side of (11) and (12) are the same as those of $1/n$ multiplying the left-hand side of (17) and (18) as n goes to infinity. Since the consistency and asymptotic normality of $\hat{\beta}_{(1)}$ are determined by the above two limits, the solutions of (11) and (12) also give the valid estimator for β. In addition, the asymptotic properties of the resulting estimators can still be summarized in Theorem 2, since the two equations used here are the same as those used for clustered events data with informative cluster size.

5. Numerical Results

5.1. Simulations

We carried out a series of simulation studies to evaluate the small-sample performance of the methods developed in Sections 3 and 4. For each study, the sample size $n = 100$ and the hazard function of error term ϵ is chosen as the form $\lambda(t, r) = \exp(t)/\{1 + r\exp(t)\}$, with $r = 0,1$ (Dabrowska and Doksum, 1988; Chen et al.2002). Note that the proportional hazards and proportional odds model correspond to $r = 0$ and $r = 1$, respectively. For $r = 0$, we choose $H_0(t) = \log(t)$, while for $r = 1$, $H_0(t) = \log\{\exp(t) - 1\}$.

For clustered events data, suppose that there were two event times within each cluster. And the corresponding two error terms were generated from Gumbel (1960)'s bivariate distribution: $F(x_1, x_2) = F_1(x_1)F_2(x_2)[1 + \theta\{1 - F_1(x_1)\}\{1 - F_2(x_2)\}]$, where $-1 \leq \theta \leq 1$ and $F_1(x) = F_2(x) = 1 - \exp\{-\Lambda(x, r)\}$. Then the correlation between ϵ_1 and ϵ_2 is just $\theta/4$. Let $Z_i = (Z_{i1}, Z_{i2})'$ be the covariate vector, where Z_{i1} were uniformly distributed on $(-1, 1)$ and Z_{i2} were Bernoulli with 0.5 success probability. The regression coefficient was chosen as $\beta = (0.5, -1.0)$. The censoring times were generated from the uniform $(0, c)$ distribution with desired level of censoring.

For recurrent events data, the covariates were generated in the same manner as above. And the regression coefficient was still chosen as $\beta = (0.5, -1.0)$. The two successive gap times were generated from the same Gumbel's bivariate distribution considered above. The follow-up times were generated from uniform (0, 7) distribution. For $r = 0$, it yielded an average of approximately 2.35, 2.50 and 2.76 events per subject for $\theta = -1$, 0 and 1, respectively; while for $r = 1$, it yielded an average of approximately 2.65, 2.79 and 3.04 events per subject for $\theta = -1$, 0 and 1, respectively.

Tables 1 and 2 summarize the results on the estimation of regression coefficients. Each entry in the table was based on 500 simulated datasets. The simulation results show that the proposed methods perform well in small samples. The parameter estimators are essentially unbiased and the means of the estimated standard error are quite close to the empirical standard errors of the parameter estimators. The 95% confidence intervals also have reasonable coverage rate.

5.2. *Application to the Bladder Cancer Data*

We also apply our estimation procedure to the well-known bladder cancer data originally reported by Byar (1980). The dataset was obtained from a randomized clinical trial assessing the effect of treatment thiotepa on the recurrence of bladder tumors. There were 38 patients in the thiotepa group with a total of 45 observed recurrence times, and 48 placebo patients with a total 87 observed recurrences. Here the gap times between any successive recurrences are modeled with the marginal linear transformation models (16) with three covariates: treatment indicator (1 for thiotepa and 0 for placebo), number of initial tumors and size of the largest initial tumor. The extreme value and logistic distributions ($r = 0$ or 1, respectively) are chosen for the error term in (16).

Table 3 displays the results of our analysis. We find that the results for the proportional hazards model ($r = 0$) are similar to those reported in Huang and Chen (2003), and are also consistent with those for the proportional odds model ($r = 1$). In both cases, the number of initial tumor appears to be significant factor while the size of initial tumor does not appear to be influential.

Table 1. Simulation results for clustered events data[a].

			$\beta_1 = 0.5$				$\beta_2 = -1.0$			
Model	θ	Censoring	Bias	*SD*	*SE*	*CP*	Bias	*SD*	*SE*	*CP*
$r = 0$	0	25%	0.014	0.149	0.145	94.8	−0.012	0.179	0.174	95.0
		50%	0.018	0.183	0.177	93.0	−0.014	0.212	0.212	95.4
	1	25%	0.012	0.148	0.145	95.0	−0.011	0.178	0.175	95.4
		50%	0.008	0.181	0.177	93.8	−0.010	0.212	0.212	96.0
$r = 1$	0	25%	0.020	0.255	0.246	93.6	−0.015	0.303	0.291	93.4
		50%	0.019	0.269	0.256	93.6	−0.014	0.296	0.302	96.2
	1	25%	0.011	0.254	0.246	92.8	−0.011	0.301	0.291	95.2
		50%	0.008	0.257	0.257	94.0	−0.017	0.294	0.301	96.8

[a]*SD*, sample standard deviation; *SE*, mean of estimated standard error; *CP*, empirical coverage probability of 95% confidence interval for β.

Table 2. Simulation results for recurrent events data[a].

		$\beta_1 = 0.5$				$\beta_2 = -1.0$			
Model	θ	Bias	*SD*	*SE*	*CP*	Bias	*SD*	*SE*	*CP*
$r = 0$	−1	0.010	0.172	0.170	94.8	−0.022	0.219	0.210	95.2
	0	0.006	0.190	0.174	91.6	−0.019	0.218	0.214	94.8
	1	0.002	0.200	0.181	92.0	0.012	0.242	0.221	93.0
$r = 1$	−1	−0.002	0.280	0.261	93.0	−0.013	0.334	0.313	92.2
	0	−0.013	0.316	0.304	93.6	−0.000	0.379	0.367	94.2
	1	0.017	0.351	0.315	91.0	0.001	0.405	0.377	92.6

[a]*SD*, sample standard deviation; *SE*, mean of estimated standard error; *CP*, empirical coverage probability of 95% confidence interval for β.

Table 3. Marginal regression analysis of Bladder cancer data.

Error distribution	Covariate	Parameter estimate	Estimated standard error	95% confidence interval
$r = 0$	Treatment	−0.566	0.306	(−1.166, 0.034)
	Initial number	0.221	0.070	(0.084, 0.358)
	Initial size	0.049	0.088	(−0.123, 0.221)
$r = 1$	Treatment	−0.768	0.430	(−1.611, 0.075)
	Initial number	0.329	0.103	(0.127, 0.531)
	Initial size	0.088	0.127	(−0.161, 0.337)

6. Discussion and Further Work

Although Cox-regression model is widely used for the analysis of survival data because of its simplicity, it is desirable to consider a more general class of semiparametric regression models, such as the linear transformation models, for several reasons. First, the proportional hazards model may not be appropriate for modeling some survival data as we mentioned before. Secondly, the linear transformation models generalize the Box–Cox transformation model with Gaussian error, whereas the hazard function modeled by the Cox-regression has no practical interpretation when the censored response variable is not survival time.

In this article, we formulate the marginal distribution of multivariate event times or gap times with the linear transformation model. And motivated by the recent work of Chen et al. (2002), the proposed estimating equation approach has the advantage of dealing with a large class of semiparametric regression models for multivariate event time data in a unified way. It does not require the covariate independent censoring assumption. The estimating equations are relatively simple to implement and allow a rigorous development of asymptotic normality with an explicit formula for the variance–covariance matrix, which can be consistently estimated by the usual plug-in method.

For theoretical and computational simplicity, we only considered independence working assumptions in the construction of estimating equations. More efficient estimator can be derived by taking into account the correlation structure. In addition, the approach proposed in this paper is limited to time-independent covariates only. A primary reason for the limitation is that the transformation formulation only handles time-independent covariates. These topics certainly warrant future research.

7. Appendix

n this appendix, we prove the theorems established in Sections 3 and 4. We follow the main steps of the Appendix of Chen et al. (2002), which deals with the linear transformation model for the univariate failure time data. To avoid delicate technical issues associated with smoothness and tail fluctuation, we assume that related functions are sufficiently smooth and make similar tail restrictions as in Chen et al. (2002).

Proof of Theorem 1 Let $\hat{H}_0(t) = \hat{H}(t; \beta_0)$. We first show that $\hat{H}_0$ is consistent. Since $\hat{H}_0$ is monotone, it suffices to show that its limiting function is unique. Suppose that $\tilde{H}$ is a limiting function. By (5) and the law of large numbers, it must satisfy

$$\tilde{N}(t) = \int_0^t m(s; \beta_0, \tilde{H}) d\tilde{H}(s).$$

where $\tilde{N}(t) = \lim_{n\to\infty}(1/n)\sum_{i=1}^n \sum_{k=1}^{K_i} N_{ik}(t)$ and

$$m(t;\beta_0,\tilde{H}) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K_i} Y_{ik}(t)\lambda\{\tilde{H}(t)+\beta_0{}'Z_{ik}\}.$$

This implies that $\tilde{H}$ is differentiable and must satisfy

$$\frac{d\tilde{H}(t)}{dt} = \frac{d\tilde{N}(t)}{dt}\Big/ m(t;\beta_0,\tilde{H}), \tag{A.1}$$

which is a smooth function of t and $\tilde{H}(t)$. Since (A.1) is a Cauchy problem, its solution exists and is unique under local smoothness assumptions (Reinhard, 1987, Theorem 3.4.1). For t in a compact subset of the interior of the support of $\tilde{T}$, we can show that the derivative of $\hat{H}(t;\beta)$ with respect to β is bounded in a neighborhood of β_0. Therefore, $\hat{H}(t;\beta_n)$ converges to $H_0(t)$ provided β_n converges to β_0. In particular, $\hat{H}(t;\hat{\beta})$ consistently estimates $H_0(t)$ provided $\hat{\beta}$ is a consistent estimator. Furthermore, let $U(\beta)$ be the left-hand side of equation (7). And let $\dot{u}(\beta)$ denote the derivative of $U(\beta)$ with respect to β. By repeatedly applying the uniform law of large numbers (Pollard, 1990), we can show that, for β in a neighborhood of β_0, $U(\beta)$ and $\dot{u}(\beta)$ converge uniformly to $u(\beta)$ and $\dot{u}(\beta)$, where $\dot{u}(\beta)$ is the derivative of $u(\beta)$.

Suppose that $\dot{u}(\beta_0)$ is nonsingular. Then there exist lower and upper bounds r and R, which are bounded away from 0 and ∞ in probability, such that, for β_1 and β_2 in a neighborhood of β_0, $r\|\beta_1-\beta_2\| \le \|U(\beta_1)-U(\beta_2)\| \le R\|\beta_1-\beta_2\|$. Since $U(\hat{\beta})=0$ and $U(\beta_0)\to u(\beta_0)=0$ as $n\to\infty$, it follows that $\hat{\beta}$ converges to β_0 in probability.

Let $a>0$ and b be fixed finite numbers. Define

$$\lambda^*\{H_0(t)\} = B(t,a), \quad \Lambda^*(t) = \int_b^t \lambda^*(s)ds.$$

where $B(t,s)$ is defined before Theorem 1. Mimicking Steps A2 and A3 in the Appendix of Chen et al. (2002), we have

$$\Lambda^*\{\hat{H}_0(t)\} - \Lambda^*\{H_0(t)\} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)}dM_{ik}(s) + o_p(n^{-1/2}), \tag{A.2}$$

and

$$\frac{\partial}{\partial\beta}\hat{H}(t;\beta)|_{\beta=\beta_0} = -\int_0^t B(s,t)\frac{m_z(s)}{B_2(s)}dH_0(s) + o_p(1), \tag{A.3}$$

where

$$m_z(t) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K_i} Y_{ik}(t)Z_{ik}\dot{\lambda}\{H_0(t)+\beta_0{}'Z_{ik}\}.$$

Then it follows the law of large numbers that

$$
\begin{aligned}
&\frac{1}{n}\frac{\partial}{\partial\beta}U(\beta)|_{\beta=\beta_0}\\
&=-\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K_i}Z_{ik}\lambda\{\hat{H}(\tilde{T}_{ik};\beta_0)+\beta_0'Z_{ik}\}\left\{Z_{ik}+\frac{\partial}{\partial\beta}\hat{H}(t;\beta)|_{\beta=\beta_0}\right\}'+o_p(1)\\
&=-\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K_i}Z_{ik}\lambda\{\hat{H}(\tilde{T}_{ik};\beta_0)+\beta_0'Z_{ik}\}\left(Z_{ik}-\int_0^{\tilde{T}_{ik}}B(s,\tilde{T}_{ik})\frac{m_z(s)}{B_2(s)}dH_0(s)\right)'+o_p(1)\\
&=-\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_0^{\tau}\{Z_{ik}-\mu_Z(t)\}Z_{ik}'\dot{\lambda}\{H_0(t)+\beta_0'Z_{ik}\}Y_{ik}(t)dH_0(t)+o_p(1)\\
&=-A+o_p(1)
\end{aligned}
$$

where A is defined in (8).

And following Step A4 in the Appendix of Chen et al. (2002), we have

$$
\begin{aligned}
&U(\beta_0)\\
&=\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_0^{\tau}Z_{ik}dM_{ik}(t)-\sum_{i=1}^{n}\sum_{k=1}^{K_i}Z_{ik}[\Lambda\{\hat{H}_0(\tilde{T}_{ik})+\beta_0'Z_{ik}\}-\Lambda\{H_0(\tilde{T}_{ik})+\beta_0'Z_{ik}\}]\\
&=\sum_{i=1}^{n}\sum_{k=1}^{K_i}\left(\int_0^{\tau}Z_{ik}dM_{ik}(t)-\frac{Z_{ik}\lambda\{H_0(\tilde{T}_{ik})+\beta_0'Z_{ik}\}}{\lambda^*\{H_0(\tilde{T}_{ik})\}}[\Lambda^*\{\hat{H}_0(\tilde{T}_{ik})\}-\Lambda^*\{H_0(\tilde{T}_{ik})\}]\right)\\
&\quad+o_p(n^{1/2})\\
&=\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_0^{ik}dM_{ik}(t)\\
&\quad-\sum_{i=1}^{n}\sum_{k=1}^{K_i}\frac{Z_{ik}\lambda\{H_0(\tilde{T}_{ik})+\beta_0'Z_{ik}\}}{\lambda^*\{H_0(\tilde{T}_{ik})\}}\frac{1}{n}\sum_{i'=1}^{n}\sum_{k'=1}^{K_i}\int_0^{\tilde{T}_{ik}}\frac{\lambda^*\{H_0(t)\}}{B_2(t)}dM_{i'}(t)+o_p(n^{1/2})\\
&=\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_0^{\tau}\{Z_{ik}-\mu_Z(t)\}dM_{ik}(t)+o_p(n^{1/2})
\end{aligned}
$$

It then follows that $n^{-1/2}U(\beta_0)\to N(0,\Sigma)$, where T is defined in (9). Therefore, Theorem 1 holds by the Taylor expansion and some empirical process approximation techniques.

Proof of Theorem 2 The proof of Theorem 2 is much similar as that given in A.1. And it is omitted here.

References

S. Bennett, "Analysis of survival data by the proportional odds model," *Statistics in Medicine* vol. 2 pp. 273–277, 1983.

P. J. Bickel, C.A. J. Klaassen, Y.Ritov and J.A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press: Baltimore, 1993.

D. P. Byar, "The veterans administration study of chempoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa ," in Bladder Tumors and Other Topics in Urological Oncology, (M. Pavone-Macaluso, P. H. Smith and F. Edsmyn, eds.) Plenum: New York, pp. 363–370, 1980.

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

S. C. Cheng, L.J. Wei and Z.Ying, "Analysis of transformation models with censored data," *Biometrika* vol. 82 pp. 835–845, 1995.

S. C. Cheng, L.J. Wei and Z.Ying, "Prediction of survival probabilities with semi-parametric transformation models," *Journal of the American Statistical Association* vol. 92 pp. 227–235, 1997.

D. Clayton and J.Cuzick, "Multivariate generalizations of the proportional hazards model (with Discussion)," *Journal of the Royal Statistical Society Series A* vol. 148 pp. 82–117, 1985.

D. R. Cox, "Regression models and life tables (with Discussion)," *Journal of the Royal Statistical Society Series B* vol. 34 pp. 187–220, 1972.

D. M. Dabrowska and K.A. Doksum, "Estimation and testing in the two-sample generalized odds rate model," *Journal of the American Statistical Association* vol. 83 pp. 744–749, 1988.

J. Fine, Z.Ying and L.J. Wei, "On the linear transformation model for censored data," *Biometrika* vol. 85 pp. 980–986, 1998.

E. J. Gumbel, "Bivariate exponential distribution," *Journal of the American Statistical Association* vol. 55 pp. 698–707, 1960.

E. B. Hoffman, P.K. Sen and C.R. Weinberg, "Within-cluster resampling," *Biometrika* vol. 88 pp. 1121–1134, 2001.

Y. Huang and Y.Q. Chen, "Marginal regression of gaps between recurrent events," *Lifetime Data Analysis* vol. 9 pp. 293–303, 2003.

J. F. Lawless, C.Nadeau and R.J. Cook, "Analysis of mean and rate functions for recurrent events ," in Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis, (D. Y. Lin and T. R. Fleming, eds.) Springer-Verlag: New York, pp. 37–49, 1997.

D. Y. Lin, "Cox regression analysis of multivariate failure time data: the marginal approach," *Statistics in Medicine* vol. 13 pp. 2233–2247, 1994.

D. Y. Lin, L.J. Wei, I.Yang and Z.Ying, "Robust inferences for the Andersen–Gill counting process model," *Journal of the Royal Statistical Society Series B* vol. 62 pp. 711–730, 2000.

S. A. Murphy, A.J. Rossini and A.W. van der Vaart, "Maximum likelihood estimation in the proportional odds model," *Journal of the American Statistical Association* vol. 92 pp. 968–976, 1997.

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

A. N. Pettitt, "Inference for the linear model using a likelihood based on ranks," *Journal of the Royal Statistical Society Series B* vol. 44 pp. 234–243, 1982.

A. N. Pettitt, "Proportional odds model for survival data and estimates using ranks," *Applied Statistics* vol. 33 pp. 169–175, 1984.

D. Pollard, *Empirical Processes: Theory and Applications, NSF-CBMS Regional Conference Series in Probability and Statistics*, 2IMS: Hayward, 1990.

R. L. Prentice and L.Hsu, "Regression on hazards ratios and cross ratios in multivariate time analysis," *Biometrika* vol. 84 pp. 349–363, 1997.

H. Reinhard, *Differential Equations: Foundations and Applications*, Macmillan: New York, 1987.

C. F. Spiekerman and D.Y. Lin, "Marginal regression models for multivariate failure time data," *Journal of the American Statistical Association* vol. 93 pp. 1164–1175, 1998.

M. C. Wang and S.H. Chang, "Nonparametric estimation of a recurrent survival function," *Journal of the American Statistical Association* vol. 94 pp. 146–153, 1999.

L. J. Wei, D.Y. Lin and L.Weissfeld, "Regression analysis of multivariate incomplete failure time data by modelling marginal distributions," *Journal of the American Statistical Association* vol. 84 pp. 1065–1073, 1989.

J. M. Williamson, S. Datta and G. A. Satten, "Marginal analysis of clustered data when cluster size is informative," *Biometrics* vol. 59 pp. 36–42, 2003.