Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Nonparametr Stat. Author manuscript; available in PMC 2015 April 23.

Published in final edited form as:

J Nonparametr Stat. 2012 ; 24(4): 1041–1050. doi:10.1080/10485252.2012.720256.

Regression analysis of clustered interval-censored failure time data with the additive hazards model

Junlong Li[a,*], **Chunjie Wang**[b], and **Jianguo Sun**[a]

[a]Department of Statistics, University of Missouri, Columbia, MO 65211, USA

[b]Mathematics School and Institute, Jilin University, Changchun 130012, People's Republic of China

Abstract

This paper discusses regression analysis of clustered failure time data, which means that the failure times of interest are clustered into small groups instead of being independent. Clustering occurs in many fields such as medical studies. For the problem, a number of methods have been proposed, but most of them apply only to clustered right-censored data. In reality, the failure time data is often interval-censored. That is, the failure times of interest are known only to lie in certain intervals. We propose an estimating equation-based approach for regression analysis of clustered interval-censored failure time data generated from the additive hazards model. A major advantage of the proposed method is that it does not involve the estimation of any baseline hazard function. Both asymptotic and finite sample properties of the proposed estimates of regression parameters are established and the method is illustrated by the data arising from a lymphatic filariasis study.

Keywords

additive hazards model; clustered data; estimating equation; interval censoring; semi-parametric regression analysis

1. Introduction

This paper discusses regression analysis of clustered failure time data, which occur when the failure times of interest are clustered into small groups or some study subjects are related such as siblings, families, or communities. The subjects from the same cluster or group usually share certain unobserved characteristics and their failure times tend to be correlated as a result. Siblings, for example, share similar genetic and environmental influences. A key feature of the failure time data is censoring and one type of censoring that has been extensively discussed is right censoring. Another complex type is interval censoring that arises when the failure event of interest cannot be observed directly but is known only to have occurred over a time interval. This is common and natural in a clinical trial or longitudinal study in which there is a periodic follow-up. For instance, an individual who is

© American Statistical Association and Taylor & Francis 2012

*Corresponding author. junlong.li@mail.missouri.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

monitored weekly for a response may miss visits for a few weeks and return in a changed response state, thus contributing an interval-censored observation. In the following, we will focus on the regression analysis of such data.

For regression analysis of clustered failure time data, several methods have been proposed when only right censoring is present (Cai and Prentice 1997; Cai, Wei and Wilcox 2000; Zeng, Lin and Lin 2008). For example, Cai et al. (2000) and Zeng et al. (2008). For example, Cai et al. (2000) and Zeng et al. (2008) investigated the fitting of semi-parametric linear transformation models to clustered right-censored data and Rossini and Moore (1999) considered the use of estimating equations and pseudolikelihood for the analysis. For the case of interval-censored data, there also exist some procedures when there is no clustering (Jewell 1994; Wang and Ding 2000; Jewell and van der Laan 2004). For example, Huang (1996) considered the maximum-likelihood approach for Cox model regression of current status data, a special case of interval-censored data, and Sun (2006) gave a comprehensive review of the existing literature on the topic.

The additive hazards model is one of the most commonly used models for regression analysis of failure time data (Lin, Oakes and Ying 1998; Ghosh 2001; Martinussen and Sheike 2002; Wang, Sun and Tong 2010; Cai and Zeng 2011). For instance, Lin et al. (1998) considered the fitting of the model to current status data and developed some estimating equation approaches for the estimation of regression parameters. Wang et al. (2010) discussed the fitting of the model to general interval-censored failure time data and Cai and Zeng (2011) investigated the additive mixed effect model for clustered right-censored failure time data. In this paper, an approach is developed for fitting the additive hazards model to clustered interval-censored failure time data.

The reminder of the paper is organised as follows. We will begin in Section 2 with describing notation and models that will be used throughout the paper. In particular, the failure time of interest is assumed to follow the additive hazards model marginally. In Section 3, an estimating equation-based approach is developed for estimating regression parameters of interest and the asymptotic properties of the proposed estimates are established. The approach can be easily implemented and does not require the estimation of the baseline hazard function. Section 4 presents some results obtained from a simulation study performed to evaluate the proposed estimation procedure and Section 5 illustrates the proposed approach by a set of clustered interval-censored data arising from a lymphatic filariasis (LF) study. Some concluding remarks and discussion are provided in Section 6.

2. Notation and models

Consider a survival study that involves n clusters of subjects with n_i denoting the size of cluster, where $i = 1,\ldots, n$. Let T_{ij} denote the failure time of interest for subject j in cluster i, where $j = 1,\ldots, n_i$, and $Z_{ij}(t)$ is a possibly time-dependent p-dimensional vector of covariates that is associated with the subject and assumed to be completely observed. Throughout this paper, the T_{ij}s are considered to be independent for subjects in different clusters but could be dependent for subjects within the same cluster, and only interval-censored data on the T_{ij}s are observed. Specifically for each T_{ij}, there exist two monitoring times denoted by U_{ij} and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

V_{ij} with $U_{ij} \leq V_{ij}$ such that T_{ij} is known only to be smaller than U_{ij}, between U_{ij} and V_{ij}, or greater than V_{ij}. That is, we have clustered interval-censored data on the T_{ij}s. Furthermore, it will be assumed that T_{ij} is independent of the monitoring times U_{ij} and V_{ij} given covariate process $Z_{ij}(t)$.

As mentioned above, our focus will be on the estimation of the covariate effect on the T_{ij}s. For this, we assume that for each cluster, there exists a latent variable b_i and all T_{ij}s are independent given b_i. Also, the hazard function for T_{ij} is specified by the following additive hazard model:

$$\lambda_{ij}\left(t|Z_{ij}\left(s\right),b_i,s\leq t\right)=\lambda_0\left(t\right)+\beta_0^{\mathrm{T}}Z_{ij}\left(t\right)+b_i \quad (1)$$

given b_i and the covariate process up to time t (Lin et al. 1998; Ghosh 2001). Here, $\lambda_0(t)$ is an unknown baseline hazard function and β_0 denotes the p-dimensional vector of regression parameters.

In practice, the monitoring variables U_{ij} and V_{ij} could depend on covariates, too. To model this and due to the strict order restriction between them, it is natural to regard them as some failure models, too. Here, we consider employing the Cox-type models

$$\lambda_{ij}^{U}\left(t|Z_{ij}\left(s\right),s\leq t\right)=\lambda_1\left(t\right)exp\left\{\gamma_0^{\mathrm{T}}Z_{ij}\left(t\right)\right\} \quad (2)$$

and

$$\lambda_{ij}^{V}\left(t|U_{ij},Z_{ij}\left(s\right),s\leq t\right)=I\left(t>U_{ij}\right)\lambda_2\left(t\right)exp\left\{\gamma_0^{\mathrm{T}}Z_{ij}\left(t\right)\right\} \quad (3)$$

for their hazard functions. In the above, both $\lambda_1(t)$ and $\lambda_2(t)$ are unspecified baseline hazard functions and γ_0 is a p-dimensional vector of regression parameters. Note that model (3) essentially assumes that the gap time between the two monitoring times U_{ij} and V_{ij} follows a Cox-type model conditional on U_{ij}. There are several motivations for considering the models described. First, it is well known that the Cox model is one of the most widely used models partly due to its simplicity and it will be seen that under the models above, an easy estimation procedure can be developed for regression coefficients without the need to estimate the baseline hazard functions. Also, the model assumptions can be easily checked since we have complete data for the monitoring times. Furthermore, the same baseline hazard functions and covariate effects for all subjects were assumed in the models (1)–(3), respectively, for simplicity and the method developed below can be easily generalised to more general situations. Some comments on these will be given below.

For subject j in cluster i, define

$$\begin{aligned}\delta_{1ij}&=I\left(T_{ij}<U_{ij}\right),\\ \delta_{2ij}&=I\left(U_{ij}\leq T_{ij}<V_{ij}\right),\\ \delta_{3ij}&=I\left(T_{ij}\geq V_{ij}\right)=1-\delta_{1ij}-\delta_{2ij},\end{aligned}$$

where $j = 1,\ldots n_i$; $i = 1,\ldots, n$. Then, define the following counting processes:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$N_{ij}^{(1)}(t) = (1-\delta_{1ij})\, I(U_{ij} \le t) = \delta_{ij}^{(1)} I(U_{ij} \le t)$$

and conditional on U_{ij},

$$N_{ij}^{(2)}(t) = \delta_{3ij} I(V_{ij} \le t) = \delta_{ij}^{(2)} I(V_{ij} \le t),$$

where $\delta_{ij}^{(1)} = 1 - \delta_{1ij}$ and $\delta_{ij}^{(2)} = \delta_{3ij}$. Note that the definition of $N_{ij}^{(2)}$ is naturally based on the order restriction between U_{ij} and V_{ij} and indicates that V_{ij} is considered only after U_{ij} has been observed. Based on the models (1)–(3), following the same arguments as those in Wang et al. (2010), one can derive the intensity functions of $N_{ij}^{(1)}$ and $N_{ij}^{(2)}$ as

$$\lambda_{ij}^{(1)}(t|Z_{ij}) = I(U_{ij} \ge t)\, E_b\left(\mathrm{e}^{-b_i t}\right) \mathrm{e}^{\Lambda_0(t)} \lambda_1(t)\, exp\left\{-\beta_0^{\mathrm{T}} Z_{ij}^*(t) + \gamma_0^{\mathrm{T}} Z_{ij}(t)\right\} \quad (4)$$

and

$$\lambda_{ij}^{(2)}(t|Z_{ij}) = I(U_{ij} < t \le V_{ij})\, E_b\left(\mathrm{e}^{-b_i t}\right) \mathrm{e}^{\Lambda_0(t)} \lambda_2(t)\, exp\left\{-\beta_0^{\mathrm{T}} Z_{ij}^*(t) + \gamma_0^{\mathrm{T}} Z_{ij}(t)\right\}, \quad (5)$$

respectively. It can be seen that models (4) and (5) are Cox-type ones similar to models (2) and (3) and model (5) is a conditional one since the starting time point is the observed monitoring time U_{ij}. In the following section, we will establish some estimating equations for the estimation of regression coefficients β_0 and γ_0.

3. Estimation of regression parameters

Now we consider the estimation of regression parameters β_0 and γ_0 and for this, we will employ the estimating equation approach. For l = 0, 1 and 2, define

$$S_{1,\beta}^{(l)}(t;\beta,\gamma) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i} I(t \le U_{ij})\, Z_{ij}^{*(l)}(t)\, exp\left\{-\beta^{\mathrm{T}} Z_{ij}^*(t) + \gamma^{\mathrm{T}} Z_{ij}(t)\right\}$$

and

$$S_{2,\beta}^{(l)}(t;\beta,\gamma) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i} I(U_{ij} < t \le V_{ij})\, Z_{ij}^{*(l)}(t)\, exp\left\{-\beta^{\mathrm{T}} Z_{ij}^*(t) + \gamma^{\mathrm{T}} Z_{ij}(t)\right\},$$

where $Z_{ij}^{*(l)}(t) = \int_0^t Z_{ij}^{(l)}(s)\,\mathrm{d}s$. Here for any vector a, $a^{(0)} = 1$, $a^{(1)} = a$ and $a^{(2)} = aa^T$. Based on the models and (5) and motivated by Zhu, Tong and Sun (2008) and Wang et al. (2010), we propose the following estimating function:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\begin{aligned} U_\beta(\beta,\gamma) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int_0^\infty \left\{ Z_{ij}^*(t) - \frac{S_{1,\beta}^{(1)}(t;\beta,\gamma)}{S_{1,\beta}^{(0)}(t;\beta,\gamma)} \right\} \mathrm{d}N_{ij}^{(1)}(t) \\ &+ \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int_0^\infty \left\{ Z_{ij}^*(t) - \frac{S_{2,\beta}^{(1)}(t;\beta,\gamma)}{S_{2,\beta}^{(0)}(t;\beta,\gamma)} \right\} \mathrm{d}N_{ij}^{(2)}(t) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{k=1}^{2}\delta_{ij}^{(k)} \left\{ Z_{ij}^*\left(W_{ij}^{(k)}\right) - \frac{S_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\beta,\gamma\right)}{S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\gamma\right)} \right\} \end{aligned}$$

for the estimation of β_0 if γ_0 is known, where $W_{ij}^{(1)}=U_{ij}$ and $W_{ij}^{(2)}=V_{ij}$.

The key idea here is to reduce general interval-censored data to current status data (Zhu et al. 2008). In the expression of $U_\beta(\beta, \gamma)$, the first term is the partial likelihood score function under model (4) if one observes only current status data, while the second term is the partial likelihood score function given under model (5) if one considers only current status data given by the observations on the V_{ij}s. Thus, $U_\beta(\beta, \gamma)$ is an unbiased function. Similar estimating function can be developed for the estimation of γ_0. However, complete data are actually available for the monitoring times U_{ij} and V_{ij} and thus it is more efficient to estimate it based on models (2) and (3).

To this end, define $\tilde{N}_{ij}^{(1)}(t)=I(U_{ij}\le t)$ and $\tilde{N}_{ij}^{(2)}(t)=I(V_{ij}\le t)$ given U_{ij}. Also for l 0, 1 and 2, define

$$S_{1,\gamma}^{(l)}(t;\gamma)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i} I(t\le U_{ij})\, exp\left\{\gamma^{\mathrm{T}} Z_{ij}(t)\right\} Z_{ij}^{(l)}(t)$$

and

$$S_{2,\gamma}^{(l)}(t;\gamma)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i} I(U_{ij}<t\le V_{ij})\, exp\left\{\gamma^{\mathrm{T}} Z_{ij}(t)\right\} Z_{ij}^{(l)}(t).$$

Then, an estimating function for γ_0 can be constructed as

$$\begin{aligned} U_\gamma(\gamma) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int_0^\infty \left\{ Z_{ij}(t) - \frac{S_{1,\gamma}^{(1)}(t;\gamma)}{S_{1,\gamma}^{(0)}(t;\gamma)} \right\} \mathrm{d}\tilde{N}_{ij}^{(1)}(t) \\ &+ \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int_0^\infty \left\{ Z_{ij}(t) - \frac{S_{2,\gamma}^{(1)}(t;\gamma)}{S_{2,\gamma}^{(0)}(t;\gamma)} \right\} \mathrm{d}\tilde{N}_{ij}^{(2)}(t) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{k=1}^{2} \left\{ Z_{ij}\left(W_{ij}^{(k)}\right) = \frac{S_{k,\gamma}^{(1)}\left(W_{ij}^{(k)};\gamma\right)}{S_{k,\gamma}^{(0)}\left(W_{ij}^{(k)};\gamma\right)} \right\}. \end{aligned}$$

Define the estimate of γ_0 and $\hat{\gamma}$ to be the solution to $U_\gamma(\gamma) = 0$ and the estimate of β_0 and $\hat{\beta}$ to be the solution to $U_\beta(\beta,\hat{\gamma})=0$. It can be easily shown that $\hat{\gamma}$ is consistent and has an asymptotic normal distribution (Wei, Lin and Weissfeld 1989; Lin 1994). The consistency

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of $\hat{\beta}$ can be similarly proved by noting that $\hat{A}_{\beta}(\beta,\hat{\gamma}) = -n^{-1}\partial U_{\beta}(\beta,\hat{\gamma})/\partial\beta$ converges to a positive matrix at β_0. For the asymptotic distribution of $\hat{\beta}$, one can first show that $n^{-1/2}U_{\beta}(\beta_0,\hat{\gamma})$ converges in distribution to a vector of normal random variables with a zero mean vector and a covariance matrix that can be consistently estimated. It then follows by the Taylor series expansion of $U_{\beta}\left(\hat{\beta},\hat{\gamma}\right)$ around β_0 that the distribution of $n^{1/2}\left(\hat{\beta}-\beta_0\right)$ can be asymptotically approximated by the normal distribution with mean zero and a covariance matrix that can be consistently estimated and provided in the appendix. The proof of the result is also sketched in the appendix.

4. A simulation study

We conducted an extensive simulation study to assess the finite sample performance of the estimation procedure proposed in the previous sections. To generate the failure times of interest in the study, it was assumed that T_{ij} followed the model

$$\lambda_{ij}(t|Z_{ij},b_i) = \lambda_0(t) + \beta_0^{\mathrm{T}} Z_{ij} + b_i$$

with $\lambda_0(t$ = 2. Note that the covariate process was assumed to be time independent for simplicity and generated from the Bernoulli distribution with success probability p = 0.5. Also, the latent variables b_is were assumed to follow a normal distribution with zero mean and variance equal to $\frac{1}{4}$. For the monitoring variables U_{ij}s and V_{ij}s, they were generated based on the models (2) and (3) with $\lambda_1(t)$ = 4 and $\lambda_2(t)$ = 2, respectively. Finally, the cluster size n_i was assumed to follow the uniform distribution $U\{2, 3, 4\}$.

The following tables give the results based on 1000 replications, including the bias of the estimates given by the average of the estimates minus the true value (BIAS), the sample standard deviation (SSD) of the estimates, the average of the estimated standard errors (ESE) and the 95% empirical coverage probability (95%-CP). In particular, Table 1 presents the obtained results on the estimates $\hat{\beta}$ and $\hat{\gamma}$ based on the simulated data with the true values of β_0 being −0.25, 0, 0.25, 0.5 and 1, respectively with γ_0 = −0.25. Two choices of cluster sizes, n = 200 and 400, are considered in this simulation study. One can see that these results suggest that the proposed estimates seem to be unbiased and the SSD is close to the ESE, suggesting that the proposed variance estimate and the normality of the estimates are reasonable. It is interesting to find out that the parameter γ seems to be better estimated than the parameter β. This is reasonable as complete data are available for the former and one only observes incomplete data for the latter.

Tables 2 and 3 give the results obtained for the estimation of the regression parameters β and γ with γ_0 = 0 and 0.25, respectively, and all other set-ups being the same as those in Table 1. They are similar to those given in Table 1 and again suggest that the proposed estimation approach seems to work well for the situations considered.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

5. An application

In this section, we apply the estimation procedure proposed in the previous sections to a set of clustered interval-censored failure time data arising from an LF study (Williamson, Kim, Manatunga and Addiss 2008). The study followed 47 men with LF, a debilitating parasitic disease in which several worms live together in several nests. An effective treatment is expected to kill the worms in all of the nests. The goal of the study was to compare the effect of the co-administration of diethylcarbamazine (DEC) and albendazole (ALB) (new treatment) versus DEC alone (standard treatment) for the treatment of LF. The patients in the study were followed for a year since their treatment and periodically examined by ultrasound to see if the worms were still alive. Thus, for the times to the clearance of the worms in each nest, the variables of interest, only clustered interval-censored data were observed with each patient serving as a cluster and the cluster size being the number of nests of adult filial worms in the body of each patient.

Among 47 patients, 22 received the co-administration of DEC and ALB, while the others were given DEC alone. In total, 78 adult worm nests were detected by ultrasound with the cluster size n_i ranging from 1 to 5. In addition to the treatment indicator, the age of each subject in years was also observed, ranging from 16 to 66. In the analysis below, we define X_{1i} to be 0 if subject i was given the co-administration of DEC and ALB and 1 otherwise and let X_{2i} be the age of the corresponding patient. Note that here we only have cluster-specific covariates.

To apply the proposed method, we assume that the times to the clearance of the worms and the monitoring times can be described by models (1)–(3), respectively. Since the observed data were given in the form $T_{ij} \in [L_{ij}, R_{ij})$, that is, we have a mixture of left-, interval-, and right-censored observations, we need to transfer them to the form expressed by U_{ij} and V_{ij} to implement the proposed estimation procedure. For this and an observed interval $[L_{ij}, R_{ij})$, we set $U_{ij} = R_{ij}$ and V_{ij} to be the largest observation time in the study if $L_{ij} = 0$; if $0 < L_{ij} < R_{ij} < +\infty$, we let $U_{ij} = L_{ij}$ and $V_{ij} = R_{ij}$; if $R_{ij} = +\infty$, we take $V_{ij} = L_{ij}$ and U_{ij} to be the smallest observation time in the study. Correspondingly, the estimating function $U_\gamma(\gamma)$ needs to be adjusted to

$$U_\gamma(\gamma) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{k=1}^{2}\tilde{\delta}_{ij}^{(k)}\left\{Z_{ij}\left(W_{ij}^{k}\right) - \frac{S_{k,\gamma}^{(1)}\left(W_{ij}^{(k)};\gamma\right)}{S_{k,\gamma}^{(0)}\left(W_{ij}^{(k)};\gamma\right)}\right\},$$

where $\tilde{\delta}_{ij}^{(1)} = 1 - \delta_{3ij}$ and $\tilde{\delta}_{ij}^{(2)} = 1 - \delta_{1ij}$. In contrast, the estimating function $U_\beta(\beta, \gamma)$ remains the same. This essentially treats U_{ij} as missing in the right-censored case, and V_{ij} as missing in the left-censored case.

The results obtained by the application of the proposed estimation procedure to the data are presented in Table 4 and it includes the estimated treatment and age effects on the time to the clearance of the worms, the estimated standard deviation (SD), and the p-values for testing the covariate effects equal to zero. They suggest that the two treatments seem to have no significant difference in killing or cleaning the worms and also the clearance of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

worms did not seem to be significantly related to the age of the patient. However, one may be careful about the conclusions due to the small number of subjects.

6. Concluding remarks and discussion

As mentioned before, clustered failure time data occurs in a study if study subjects are related through being clustered into small groups. In this case, one has to take into account the correlation among them to perform a valid analysis. In the previous sections, an estimation procedure was developed for the problem in the presence of interval censoring for the data arising from the additive hazards model and the asymptotic properties of the proposed estimates were established. One major advantage of the presented method is that it does not involve the estimation of the baseline hazard functions.

For the problem considered here, an alternative approach is to employ the full likelihood approach and a main advantage of this method is that one can expect that it could be more efficient than the estimating equation method proposed here. However, the full likelihood approach would be time-consuming and may be infeasible if a nonparametric or semi-parametric approach was adopted since it involves the estimation of the infinite-dimensional functions. Also, it could be difficult to derive the asymptotic properties of the resulting estimates. In contrast, the proposed method can be easily implemented.

The proposed methodology involves modelling gap times between the adjacent monitoring times using the Cox model. An alternative choice is to model the monitoring times using the Cox model marginally. However, such modelling requires stricter conditions due to the order relationship. In contrast, the gap time modelling approach is more flexible. In the presented approach, the latent variables b_is are assumed to follow the normal distribution for simplicity and the methodology still applies for other distributions.

Acknowledgements

The authors are grateful to the editor and the reviewers for their insightful comments on the article. This work was partially supported by NIH grant 5 R01 CA152035 to the third author.

Appendix

In the following, we will sketch the proofs of the asymptotic normality of the estimates $\hat{\gamma}$ and $\hat{\beta}$. First, we will discuss the asymptotic normality of $\hat{\gamma}$. For this and l = 0, 1, 2, let $s_{0,\gamma}^{(l)}$ and $s_{1,\gamma}^{(l)}$ denote the limits of $S_{0,\gamma}^{(l)}$ and $S_{1,\gamma}^{(l)}$, respectively. By following Wang et al. (2010), one can obtain that

$$\hat{\gamma}-\gamma_0=U_{\gamma\gamma}^{-1}\frac{1}{n}\sum_{i=1}^{n}U_i\left(\gamma_0\right)+o_p\left(n^{-1/2}\right), \quad (1)$$

where

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$U_i(\gamma)=\sum_{j=1}^{n_i}\sum_{k=1}^{2}\left\{Z_{ij}\left(W_{ij}^{(k)}-\frac{s_{k,\gamma}^{(1)}\left(W_{ij}^{(k)};\gamma\right)}{s_{k,\gamma}^{(0)}\left(W_{ij}^{(k)};\gamma\right)}\right\}$$

and

$$U_{\gamma\gamma}=E\left\{\sum_{j=1}^{n_i}\sum_{k=1}^{2}\frac{s_{k,\gamma}^{(2)}\left(W_{ij}^{(k)};\gamma_0\right)}{s_{k,\gamma}^{(0)}\left(W_{ij}^{(k)};\gamma_0\right)}-\frac{s_{k,\gamma}^{(1)}\left(W_{ij}^{(k)};\gamma_0\right)s_{k,\gamma}^{(1)}\left(W_{ij}^{(k)};\gamma_0\right)^{\mathrm{T}}}{s_{k,\gamma}^{(0)}\left(W_{ij}^{(k)};\gamma_0\right)^2}\right\}.$$

with $W_{ij}^{(1)}=U_{ij}$ and $W_{ij}^{(2)}=V_{ij}$. Thus, the distribution of $n^{1/2}(\hat{\gamma}-\gamma_0)$ can be approximated by the normal distribution with mean zero and covariance matrix $U_{\gamma\gamma}^{-1}\Sigma_\gamma U_{\gamma\gamma}^{-1}$ which can be consistently estimated by $\hat{U}_{\gamma\gamma}^{-1}\hat{\Sigma}_\gamma\hat{U}_{\gamma\gamma}^{-1}$, In the above,

$$\hat{\Sigma}_\gamma=\frac{1}{n}\sum_{i=1}^{n}U_i(\hat{\gamma})U_i(\hat{\gamma})^{\mathrm{T}},$$

$$U_i(\hat{\gamma})=\sum_{j=1}^{n_i}\sum_{k=1}^{2}\left\{Z_{ij}\left(W_{ij}^{(k)}\right)-\frac{s_{k,\hat{\gamma}}^{(1)}\left(W_{ij}^{(k)};\hat{\gamma}\right)}{s_{k,\hat{\gamma}}^{(0)}\left(W_{ij}^{(k)};\hat{\gamma}\right)}\right\}$$

and

$$\hat{U}_{\gamma\gamma}=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{k=1}^{2}\left\{\frac{S_{k,\gamma}^{(2)}\left(W_{ij}^{(k)};\hat{\gamma}\right)}{S_{k,\gamma}^{(0)}\left(W_{ij}^{(k)};\hat{\gamma}\right)}-\frac{S_{k,\gamma}^{(1)}\left(W_{ij}^{(k)};\hat{\gamma}\right)S_{k,\gamma}^{(1)}\left(W_{ij}^{(k)};\hat{\gamma}\right)^{\mathrm{T}}}{S_{k,\gamma}^{(0)}\left(W_{ij}^{(k)};\hat{\gamma}\right)^2}\right\}.$$

For the asymptotic normality of $\hat{\beta}$, let $s_{k,\beta}^{(l)}$ denote the limit of $S_{k,\beta}^{(l)}$, $l=0,1,2$ and k = 1, 2. By some calculation and Equation (1), we have

$$\begin{aligned}U(\beta,\gamma_0)-U(\beta,\hat{\gamma})&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{k=1}^{2}\delta_{ij}^{(k)}\left\{\frac{S_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\beta,\hat{\gamma}-S_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\beta,\gamma_0\right.\right.}{S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\hat{\gamma}\right)}-\frac{S_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)\left[S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\hat{\gamma}\right)-S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)\right]}{S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\hat{\gamma}\right)S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\gamma_0\right.}\right\}\\&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{k=1}^{2}\delta_{ij}^{(k)}\left\{\frac{S_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\beta,\hat{\gamma}-S_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\beta,\gamma_0\right.\right.}{s_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)}-\frac{s_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)\left[S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\hat{\gamma}\right)-S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)\right]}{s_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)^2}\right\}+o_p\left(n^{-}\right.\end{aligned}$$

Note that here $\delta_{ij}^{(1)}=1-\delta_{1ij}$ and $\delta_{ij}^{(2)}=\delta_{3ij}$. Furthermore, define the following for l = 0, 1,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$A_1^{(l)}(t;\beta,\gamma)=E\left[\sum_{j=1}^{n_i} I\left(t\le U_{ij}\right) Z_{ij}^{*(l)}(t)\, Z_{ij}(t)^{\mathrm{T}} exp\left\{-\beta^{\mathrm{T}} Z_{ij}^{*}(t)+\gamma^{\mathrm{T}} Z_{ij}(t)\right\}\right],$$

$$A_2^{(l)}(t;\beta,\gamma)=E\left[\sum_{j=1}^{n_i} I\left(U_{ij}<t\le V_{ij}\right) Z_{ij}^{*(l)}(t)\, Z_{ij}(t)^{\mathrm{T}} exp\left\{-\beta^{\mathrm{T}} Z_{ij}^{*}(t)+\gamma^{\mathrm{T}} Z_{ij}(t)\right\}\right]$$

and

$$A_3(\beta,\gamma)=-E\left[\sum_{j=1}^{n_i}\sum_{k=1}^{2}\delta_{ij}^{(k)}\left\{\frac{A_k^{(1)}\left(W_{ij}^{(k)};\beta,\gamma\right)}{s_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\gamma\right)}-\frac{s_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\beta,\gamma\right)A_k^{(0)}\left(W_{ij}^{(k)};\beta,\gamma\right)}{s_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\gamma\right)}\right\}\right].$$

By the Taylor series expansion and Equation (1), one obtains that

$$\begin{aligned}U(\beta,\hat{\gamma})-U(\beta,\gamma_0) &= \frac{1}{n}\sum_{i=1}\sum_{j=1}^{n_i}\sum_{k=1}^{2}\delta_{ij}^{(k)}(\hat{\gamma}-\gamma_0)\left\{\frac{A_k^{(1)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)}{s_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)}-\frac{s_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)A_k^{(0)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)}{s_{k,\beta}^{(0)}\left(W_{ij}^{k};\beta,\gamma_0\right)^2}\right\}+o_p\left(n^{-1/2}\right)\\ &= -E\left[\sum_{j=1}^{n_i}\sum_{k=1}^{2}\delta_{ij}^{(k)}(\hat{\gamma}-\gamma_0)\left\{\frac{A_k^{(1)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)}{s_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)}-\frac{s_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)A_k^{(0)}\left(W_{ij}^{(k)};\beta,\gamma_0\right)}{s_{k,\beta}^{(0)}\left(W_{ij}^{k};\beta,\gamma_0\right)^2}\right\}\right]+o_p\left(n^{-1/2}\right).\end{aligned}$$

This yields that

$$U(\beta,\hat{\gamma})=\frac{1}{n}\sum_{i=1}^{n}a_i(\beta,\gamma_0)+o_p\left(n^{-1/2}\right),$$

where

$$a_i(\beta,\gamma)=\sum_{j=1}^{n_i}\sum_{k=1}^{2}\left[\delta_{ij}^{(k)}\left\{Z_{ij}^{*}\left(W_{ij}^{(k)}\right)-\frac{s_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\beta,\gamma\right)}{s_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\beta,\gamma\right)}\right\}+A_3(\beta,\gamma)\,U_{\gamma\gamma}^{-1}\left\{Z_{ij}\left(W_{ij}^{(k)}-\frac{s_{k,\gamma}^{(1)}\left(W_{ij}^{(k)};\gamma\right)}{s_{k,\gamma}^{(0)}\left(W_{ij}^{(k)};\gamma\right)}\right\}\right].$$

This shows that the distribution of $n^{1/2}\left(\hat{\beta}-\beta_0\right)$ can be approximated by the normal distribution with zero mean and the covariance matrix that can be consistently estimated by $\hat{\Gamma}^{-1}\hat{\Sigma}\hat{\Gamma}^{-1}$, where

$$\hat{\Gamma}=-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{k=1}^{2}\left\{\frac{S_{k,\beta}^{(2)}\left(W_{ij}^{(k)};\hat{\beta},\hat{\gamma}\right)}{S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\hat{\beta},\hat{\gamma}\right)}-\frac{S_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\hat{\beta},\hat{\gamma}\right)S_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\hat{\beta},\hat{\gamma}\right)^{\mathrm{T}}}{S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\hat{\beta},\hat{\gamma}\right)^2}\right\},$$

$$\hat{\Sigma}=\frac{1}{n}\sum_{i=1}^{n}a_i\left(\hat{\beta},\hat{\gamma}\right)a_i\left(\hat{\beta},\hat{\gamma}\right)^{\mathrm{T}},$$

$$a_i\left(\hat{\beta},\hat{\gamma}\right)=\sum_{j=1}^{n_i}\sum_{k=1}^{2}\left[\delta_{ij}^{(k)}\left\{Z_{ij}^{*}\left(W_{ij}^{(k)}\right)-\frac{S_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\hat{\beta},\hat{\gamma}\right)}{S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\hat{\beta},\hat{\gamma}\right)}\right\}+A_3\left(\hat{\beta},\hat{\gamma}\right)\hat{U}_{\gamma\gamma}^{-1}\left\{Z_{ij}\left(W_{ij}^{(k)}\right)-\frac{S_{k,\gamma}^{(1)}\left(W_{ij}^{(k)};\hat{\gamma}\right)}{S_{k,\gamma}^{(0)}\left(W_{ij}^{(k)};\hat{\gamma}\right)}\right\}\right]$$

and

$$A_3\left(\hat{\beta},\hat{\gamma}\right)=-\frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{n_i}\sum_{k=1}^{2}\delta_{ij}^{(k)}\left\{\frac{A_k^{(1)}\left(W_{ij}^{(k)};\hat{\beta},\hat{\gamma}\right)}{S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\hat{\beta},\hat{\gamma}\right)}-\frac{S_{k,\beta}^{(1)}\left(W_{ij}^{(k)};\hat{\beta},\hat{\gamma}\right)A_k^{(0)}\left(W_{ij}^{(k)};\hat{\beta},\hat{\gamma}\right)}{S_{k,\beta}^{(0)}\left(W_{ij}^{(k)};\hat{\beta},\hat{\gamma}\right)^2}\right\}\right].$$

References

Cai J, Prentice R. Regression Estimation Using Multivariate Failure Time Data and a Common Baseline Hazard Function Model. Lifetime Data Analysis. 1997; 3:197–213. [PubMed: 9384652]

Cai J, Zeng D. Additive Mixed Effect Model for Clustered Failure Time Data. Biometrics. 2011; 67:1340–1351. [PubMed: 21418052]

Cai T, Wei L, Wilcox M. Semi-parametric Regression Analysis for Clustered Failure Time Data. Biometrika. 2000; 87:867–878.

Ghosh D. Efficiency Considerations in the Additive Hazard Model with Current Status Data. Statistica Neerlandica. 2001; 55:367–376.

Huang J. Estimation for the Cox Model with Interval Censoring. The Annals of Statistics. 1996; 24:540–568.

Jewell NP. Non-parametric Estimation and Doubly-Censored Data: General Ideas and Applications to Aids. Statistics in Medicine. 1994; 13:2081–2095. [PubMed: 7846412]

Jewell NP, van der Laan MJ. Current Status Data: Review, Recent Developments and Open Problems. Advances in Survival Analysis. 2004; 23:625–642.

Lin D. Cox Regression Analysis of Multivariate Failure Time Data: The Marginal Approach. Statistics in Medicine. 1994; 85:2233–2247. [PubMed: 7846422]

Lin D, Oakes D, Ying Z. Additive Hazard Regression with Current Status Data. Biometrika. 1998; 85:289–298.

Martinussen T, Sheike TH. Efficient Estimation in Additive Hazard Regression with Current Status Data. Biometrika. 2002; 89:649–658.

Rossini A, Moore D. Modeling Clustered, Discrete, or Grouped Time Survival Data with Covariates. Biometrics. 1999; 55:813–819. [PubMed: 11315011]

Sun, J. The Statistical Analysis of Interval Censored Failure Time Data. Springer; New York: 2006.

Wang W, Ding AA. On Assessing the Association for Bivariate Current Status Data. Biometrika. 2000; 87:879–893.

Wang L, Sun J, Tong X. Regression Analysis of Case II Interval-Censored Failure Time Data with the Additive Hazard Model. Statistica Sinica. 2010; 20:1709–1723.

Wei L, Lin D, Weissfeld L. Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. Journal of the American Statistical Association. 1989; 84:1065–1073.

Williamson J, Kim H, Manatunga A, Addiss D. Modeling Survival Data with Informative Cluster Size. Statistics in Medicine. 2008; 27:543–555. [PubMed: 17640035]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Zeng D, Lin D, Lin X. Semiparametric Transformation Models with Random Effects for Clustered Failure Time Data. Statistica Sinica. 2008; 18:355–377. [PubMed: 19809573]

Zhu L, Tong X, Sun J. A Transformation Approach for the Analysis of Interval-Censored Failure Time Data. Lifetime Data Analysis. 2008; 14:167–178. [PubMed: 18165933]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Estimation of γ and β with binary covariate and $\gamma_0 = -0.25$.

		$n = 200$				$n = 400$			
TRUE	**EST**	**BIAS**	**SSD**	**ESE**	**95%-CP**	**BIAS**	**SSD**	**ESE**	**95%-CP**
$\gamma_0 = -0.25$	$\hat{\gamma}$	−0.0068	0.1107	0.1077	0.948	−0.0029	0.0772	0.0760	0.949
$\beta_0 = -0.25$	$\hat{\beta}$	−0.0638	0.8182	0.7605	0.937	−0.0299	0.5587	0.5331	0.940
$\gamma_0 = -0.25$	$\hat{\gamma}$	−0.0084	0.1087	0.1059	0.948	0.0058	0.0752	0.0748	0.965
$\beta_0 = 0.00$	$\hat{\beta}$	−0.0694	0.8162	0.7619	0.966	−0.0310	0.5616	0.5352	0.969
$\gamma_0 = -0.25$	$\hat{\gamma}$	0.0112	0.1074	0.1046	0.941	0.0082	0.0745	0.0739	0.952
$\beta_0 = 0.25$	$\hat{\beta}$	−0.0767	0.8226	0.7672	0.965	−0.0317	0.5658	0.5376	0.960
$\gamma_0 = -0.25$	$\hat{\gamma}$	0.0193	0.1063	0.1036	0.944	0.0120	0.0734	0.0732	0.945
$\beta_0 = 0.50$	$\hat{\beta}$	−0.0832	0.8256	0.7728	0.967	−0.0319	0.5640	0.5419	0.945
$\gamma_0 = -0.25$	$\hat{\gamma}$	0.0341	0.1044	0.1021	0.934	0.0259	0.0721	0.0720	0.937
$\beta_0 = 1.00$	$\hat{\beta}$	−0.0962	0.8518	0.7902	0.955	−0.0480	0.5797	0.5533	0.938

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Estimation of γ and β with binary covariate and $\gamma_0 = 0.00$.

		$n = 200$				$n = 400$			
TRUE	**EST**	**BIAS**	**SSD**	**ESE**	**95%-CP**	**BIAS**	**SSD**	**ESE**	**95%-CP**
$\gamma_0 = 0.00$	$\hat{\gamma}$	−0.0051	0.1126	0.1100	0.952	−0.0034	0.0799	0.0777	0.945
$\beta_0 = -0.25$	$\hat{\beta}$	0.0172	0.8776	0.8255	0.938	0.0155	0.6065	0.5795	0.942
$\gamma_0 = 0.00$	$\hat{\gamma}$	0.0019	0.1106	0.1081	0.948	0.0039	0.0785	0.0764	0.945
$\beta_0 = 0.00$	$\hat{\beta}$	0.0315	0.8804	0.8273	0.975	0.0065	0.6158	0.5810	0.968
$\gamma_0 = 0.00$	$\hat{\gamma}$	0.0073	0.1102	0.1065	0.946	0.0122	0.0768	0.0753	0.943
$\beta_0 0 = 0.25$	$\hat{\beta}$	−0.0225	0.8835	0.8303	0.973	−0.0061	0.6106	0.5838	0.963
$\gamma_0 = 0.00$	$\hat{\gamma}$	0.0122	0.1087	0.1053	0.941	0.0143	0.0755	0.0744	0.946
$\beta_0 0 = 0.50$	$\hat{\beta}$	−0.0422	0.8903	0.8389	0.931	−0.0169	0.6110	0.5892	0.941
$\gamma_0 = 0.00$	$\hat{\gamma}$	0.0263	0.1063	0.1035	0.932	0.0212	0.0746	0.0732	0.940
$\beta_0 = 1.00$	$\hat{\beta}$	−0.0748	0.9116	0.8583	0.933	−0.0498	0.6539	0.6036	0.960

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

Estimation of γ and β with binary covariate and $\gamma_0 = 0.25$.

		$n = 200$				$n = 400$			
TRUE	**EST**	**BIAS**	**SSD**	**ESE**	**95%-CP**	**BIAS**	**SSD**	**ESE**	**95%-CP**
$\gamma_0 = 0.25$	$\hat{\gamma}$	−0.0036	0.1177	0.1151	0.941	−0.0009	0.0829	0.0831	0.956
$\beta_0 = -0.25$	$\hat{\beta}$	0.0992	0.9643	0.9112	0.963	0.0328	0.7017	0.6768	0.956
$\gamma_0 = 0.25$	$\hat{\gamma}$	0.0035	0.1154	0.1126	0.941	0.0054	0.0812	0.0762	0.948
$\beta_0 = 0.00$	$\hat{\beta}$	0.0850	0.9610	0.9150	0.940	0.0353	0.6957	0.6747	0.956
$\gamma_0 = 0.25$	$\hat{\gamma}$	0.0122	0.1138	0.1107	0.948	0.0118	0.0802	0.0783	0.945
$\beta_0 = 0.25$	$\hat{\beta}$	0.0731	0.9634	0.9133	0.971	0.0307	0.7038	0.6763	0.959
$\gamma_0 = 0.25$	$\hat{\gamma}$	0.0193	0.1124	0.1093	0.941	0.0186	0.0792	0.0772	0.938
$\beta_0 = 0.50$	$\hat{\beta}$	0.0627	0.9690	0.9216	0.936	0.0262	0.7102	0.6822	0.960
$\gamma_0 = 0.25$	$\hat{\gamma}$	0.0286	0.1112	0.1070	0.926	0.0208	0.0769	0.0757	0.935
$\beta_0 = 1.00$	$\hat{\beta}$	0.0432	0.9721	0.9289	0.972	0.0299	0.7129	0.6964	0.969

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Estimations from the LF study.

Covariate	Estimate	SD	*p*-Value
Treatment (γ_1)	−0.375	0.383	0.327
Age (γ_1)	0.0061	0.020	0.755
Treatment (β_1)	−0.522	0.442	0.237
Age (β_2)	0.0108	0.415	0.794