Published in final edited form as:

Stat Med. 2014 January 15; 33(1): 46–58. doi:10.1002/sim.5914.

Marginal Regression Approach for Additive Hazards Models with Clustered Current Status Data

Pei-Fang Su and Yunchan Chi

Department of Statistics, National Cheng Kung University, Tainan, Taiwan R.O.C.

Yunchan Chi: ycchi@stat.ncku.edu.tw

Abstract

Current status data arise naturally from tumorigenicity experiments, epidemiology studies, biomedicine, econometrics, demographic and sociology studies. Moreover, clustered current status data may occur with animals from the same litter in tumorigenicity experiments or with subjects from the same family in epidemiology studies. Since the only information extracted from current status data is whether the survival times are before or after the monitoring or censoring times, the nonparametric maximum likelihood estimator of survival function converges at a rate of $n^{1/3}$ to a complicated limiting distribution. Hence, semiparametric regression models, such as the additive hazards model, have been extended for independent current status data to derive the test statistics, whose distributions converge at a rate of $n^{1/2}$, for testing the regression parameters. However, a straightforward application of these statistical methods to clustered current status data is not appropriate because intra-cluster correlation needs to be taken into account. Therefore, this paper proposes two estimating functions for estimating the parameters in the additive hazards model by extending the methodologies in Lin *et al.* (1998) and Martinussen and Scheike (2002) for clustered current status data. The comparative results from simulation studies are presented, and application of the proposed estimating functions to one real data set is illustrated.

Keywords

additive hazards model; clustered current status data; counting process; estimating function; marginal regression approach

1. INTRODUCTION

The time to a certain event is often of interest in many research areas, such as biomedicine, reliability, tumorigenicity experiments, demographic studies, sociology studies, and econometrics, among others. For example, the Pakistan Fertility Survey is a demographic survey of breastfeeding in developing countries (Diamond *et al.* [1]) which estimates the distribution of time-to-weaning and studies its dependence on covariates. Because direct queries about age at weaning are subject to serious measurement errors, this survey recorded whether or not the child was weaned at the time of the survey and the current age of the child in months. Consequently, the exact age of weaning can not be directly observed but is known to be above or below a random monitoring time. These kinds of data are known as current status data. The data set from the Pakistan Fertility Survey consists of 3350 children born during the four years prior to the survey and surviving to the survey. Since a woman may give birth to more than one child within four years, the times to weaning are correlated within a family. This is referred to as clustered current status data.

Another example is from the 2001 National Health Interview Survey conducted by the National Health Research Institute in Taiwan. The cataract data set is a subset of this survey and consists of the occurrence of cataracts in each eye before or after examination time (the current age) for elderly people aged 65 to 100 years. The survival times are the age at onset of cataracts for both eyes. As a result, current status data are observed from this survey. Moreover, since both eyes of the same subject share common genetic or environmental effects, one might expect that the times to cataracts for a given subject would be correlated. Thus, clustered current status data were also observed in this study.

Because the only information extracted from current status data is whether the survival times occur before or after the monitoring or censoring times, the nonparametric maximum likelihood estimator of the survival function converges at a rate of $n^{1/3}$ to a complicated limiting distribution (Groeneboon and Wellner [2]). Hence, semiparametric regression models and linear regression models have been extended for independent current status data to derive the test statistics, whose distributions converge at a rate of $n^{1/2}$, for testing the regression parameters. For instance, the proportional hazards model has been considered by Finkelstein [3], while linear regression models were used by Klein and Spady [4] and Rabinowitz *et al.* [5]. The additive hazards model has been studied by Lin *et al.* [6] based on the partial score function derived from the partial likelihood function, which is constructed according to defined counting processes. The advantage of their approach is that both the hazard functions of survival times and monitoring times can be eliminated from the partial likelihood function when the hazard function of monitoring times is independent of covariates. Recently, Martinussen and Scheike [7] pointed out that the approach proposed by Lin *et al.* [6] does not make efficient use of the data, and they proposed an empirical version of the efficient score function as an estimation function for regression parameter. Their approach is more efficient because it reaches the semiparametric information bound, and it does not involve any modeling of the monitoring times.

A straightforward application of these statistical methods to clustered current status data is not appropriate, however. More recently, Wen and Chen [8] extended the gamma-frailty Cox model for clustered current status data, which can account for within-cluster correlation. When the correlation structure within clusters is not of interest, this paper uses a marginal regression approach to constructing estimating functions for estimating the parameters in the additive hazards model for clustered current status data by extending the methodologies in Lin *et al.* [6] and Martinussen and Scheike [7]. Because both estimating approaches for additive hazards models utilize the martingale techniques to derive asymptotic properties of the regression parameter estimators, it seems that these methodologies may be easily extended to multivariate cases without specifying any correlation structure. Because the performance of these two methods has not yet been well addressed for clustered current status data, we conducted a simulation investigation in this article.

The rest of this paper is organized as follows: Section 2 contains three subsections; subsection 2.1 reviews additive hazards models and states the common settings for the two extended procedures. The extension of the partial score function is developed in subsection 2.2, and the extension of the empirical efficient score function is derived in subsection 2.3. In addition, the regularity conditions shared with those two methods for deriving asymptotic distributions are listed in Appendix A. Section 3 discusses the comparative results from simulation studies. The implementation of the proposed methods is illustrated through one example based on real data in Section 4. Finally, concluding remarks are presented in the last section.

2. THE ESTIMATION PROCEDURES WITH CLUSTERED CURRENT STATUS DATA

2.1 The additive hazards models

Let C_{ij} and T_{ij} be monitoring and survival times, respectively, for the j th subject in the i th cluster, $j = 1, 2, \dots, m_i$, $i = 1, 2, \dots, n$, and $X_{ij}(t)$ denote a p -dimensional time-dependent covariate vector for the j th subject in the i th cluster. We assume C_{ij} and T_{ij} are conditional independent, given the covariate vector $X_{ij}(t)$. The additive hazards model of marginal survival times T_{ij} is

$$\lambda(t|X_{ij}(t)) = \lambda_0(t) + \beta' X_{ij}(t)$$

where $\lambda_0(t)$ is an unspecified baseline hazard function of survival times, β is a p -dimensional unknown regression parameter vector, and β' stands for the transpose of β . The only information extracted from current status data is whether the survival times T_{ij} are before or after the random monitoring times (or censoring times) C_{ij} . Thus, the right-censored indicator is defined as $\delta_{ij} = I(C_{ij} - T_{ij})$, where $I(A)$ is an indicator function of event A , taking value 1 if the event A occurs and 0 otherwise. Throughout this article, we assume that the distribution of C_{ij} is absolutely continuous with an unspecified cumulative hazard function $\Lambda_c(t)$ and that the corresponding hazard function is $\lambda_c(t) = d\Lambda_c(t)$ which is independent of the covariate vector $X_{ij}(t)$. The extension of the partial score function and the efficient score function are introduced in the following sections.

2.2 The extension of partial score function

In the case of $j = 1$, all observations independent of each other, Lin *et al.* [6] proposed estimating β based on the partial score function derived from the partial likelihood function, which is constructed based on defined counting processes. For clustered current status data, however, it is difficult to derive full likelihood and partial likelihood functions, since the joint distribution of correlated survival times is often unknown. Therefore, the partial score function is proposed as an estimating function for clustered current status data based on a defined counting process, $N_{1ij}(t) = \delta_{ij} I(C_{ij} - t)$. This counting process jumps with unity when the j th subject of the i th cluster is monitored at time t and found still to be failure-free with intensity process $e^{-\beta' Z_{ij}(t)} e^{-\Lambda_0(t)} d\Lambda_c(t)$, where $Z_{ij}(t) = \int_0^t X_{ij}(s) ds$ is a p -dimensional vector, and $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ is the baseline cumulative hazard function of survival times. Note that the jump of this counting process at time t can also be expressed as $dN_{1ij}(t) = I(C_{ij} = t, T_{ij} > t)$.

Let $c_1 < c_2 < \dots < c_D$ be the distinct ordered observed monitoring times. Let a be a p -dimensional vector with $a^{\otimes 1} = a$, and let $a^{\otimes 0}$ denote a p -dimensional vector of 1's. When the cumulative hazard function for monitoring times, $\Lambda_c(t)$, is independent of covariates, the proposed estimating function analogous to partial score functions for clustered current status

data is $U(\beta) = \sum_{i=1}^n U_i(\beta)$, where $U_i(\beta)$ is a p -dimensional vector and is the contribution from the i th cluster that

$$U_i(\beta) = \sum_{j=1}^{m_i} \sum_{l=1}^D \{Z_{ij}(c_l) - \bar{Z}(\beta, c_l)\} I(C_{ij} = c_l, T_{ij} \geq c_l),$$

where $Z(\beta, t) = S^{(1)}(\beta, t) / S^{(0)}(\beta, t)$ is a p -dimensional vector with

$S^{(r)}(\beta, t) = \sum_{i=1}^n \sum_{j=1}^{m_i} Y_{ij}(t) \exp(-\beta' Z_{ij}(t)) (Z_{ij}(t))^{\otimes r}$, and $Y_{ij}(t) = I(C_{ij} \leq t)$ indicates whether the j th subject in the i th cluster has been monitored by time t . Because $Z_{ij}(t) - Z(\beta, t)$ is a predictable process, as described in Lin *et al.* [6], $U_i(\beta)$ is a martingale and has a mean of zero. Therefore, $U(\beta)$ is an appropriate estimating function for β . Under regularity conditions [9], the unbiased estimating functions for β have roots which are consistent estimators. Hence, a consistent estimate of β , $\hat{\beta}_S$, can be obtained by solving $U(\beta) = 0$. To make an inference about β , the asymptotic distribution of $n^{-1/2}U(\beta)$ needs to be derived first and is given in *Theorem 1* below.

Theorem 1:

Under regularity conditions, the asymptotic distribution of $n^{-1/2}U(\beta)$ is normal with mean vector 0 and a $p \times p$ variance-covariance matrix Σ_S , where Σ_S is the limit of

$n^{-1} \sum_{i=1}^n \text{Var}(U_i(\beta))$ as n goes to infinity. Moreover, the asymptotic distribution of $n^{1/2}(\hat{\beta}_S - \beta)$ is normal with mean vector 0 and a $p \times p$ variance-covariance matrix $\Sigma_{\beta_S} = J^{-1}(\beta) \Sigma_S J^{-1}(\beta)$ where $J(\beta)$ is a $p \times p$ matrix and is the expected value of the second derivative of the partial likelihood function, which ignores the dependence within the cluster, with respect to β .

More details are shown in Appendix A. Note that Σ_S can be estimated by $\hat{\Sigma}_S = \sum_{i=1}^n \hat{\sigma}_i^{\otimes 2}(\beta)$, where $\hat{\sigma}_i^{\otimes 2}(\beta)$ is a moment estimator of the within-cluster variance-covariance matrix for the i th cluster and the explicit form is given in Appendix A. The estimator $\hat{\beta}_S$ can be obtained by solving $U(\beta) = 0$ using the following Newton-Raphson algorithm until convergence:

$$\beta^{(k+1)} = \beta^{(k)} - [\hat{\Sigma}_S^{(k)}]^{-1} U(\beta^{(k)}), k=0, 1, \dots,$$

where $\beta^{(0)}$ is an initial value and $\beta^{(k)}$ is the k th parameter estimate at iteration k . To estimate the variance-covariance matrix Σ_{β_S} of $\hat{\beta}_S$, the robust sandwich estimator

$$\hat{\Sigma}_{\beta_S} = \hat{J}^{-1}(\beta) \left[\sum_{i=1}^n \hat{\sigma}_i^{\otimes 2}(\beta) \right] \hat{J}^{-1}(\beta)$$

is suggested, where $\hat{J}(\beta)$ is a $p \times p$ matrix and is the second derivative of the partial likelihood function, which ignores the dependence within the cluster, with respect to β . The expression of $\hat{J}(\beta)$ is given in Appendix A. Consequently, for testing the null hypothesis,

$H_0 : \beta = \beta_0$ versus $H_a : \beta \neq \beta_0$, the score test $U'(\beta_0) \hat{\Sigma}_S^{-1} U(\beta_0)$ and the Wald test

$(\hat{\beta}_S - \beta_0)' \sum_{\beta_S}^{-1} (\hat{\beta}_S - \beta_0)$ can be used. Under the null hypothesis, these tests have an asymptotic χ_p^2 distribution with p degrees of freedom. In Section 3, the accuracy of the asymptotic approximation will be examined.

2.3 The extension of the efficient score function

The defined counting process $N_{1ij}(t)$ is not fully efficient; therefore, Martinussen and Scheike [7] included another defined counting process, $N_{2ij}(t) = (1 - \delta_{ij})I(C_{ij} = t)$, to extract as much information as possible from the data. This counting process jumps with unity when the j th subject of the i th cluster is monitored at time t and found to have experienced failure. In the case of $j = 1$ (independent current status data), based on both $N_{1ij}(t)$ and $N_{2ij}(t)$, Martinussen and Scheike [7] derived the efficient score function from the full likelihood function, where $\Lambda_c(t)$ appears in the efficient score function and $\Lambda_0(t)$, the cumulative hazard function of survival times, cannot be eliminated. In contrast, $\Lambda_0(t)$ can be eliminated from the partial likelihood function, and hence the approach proposed by Lin *et al.* [6] does not involve any complicated nonparametric estimation of $\Lambda_0(t)$. Consequently, Martinussen and Scheike [7] suggested estimating $\Lambda_0(t)$ by a sieve estimator (e.g. Geman and Hwang [10]). For clustered data, an estimator for $\Lambda_0(t)$ denoted by $\hat{\Lambda}_0(t)$, which is analogous to the sieve estimator of $\Lambda_0(t)$ for independent current status data, is proposed and its expression is given in Appendix B. Note that $\hat{\Lambda}_0(t)$ is a predictable estimator for $\Lambda_0(t)$. The estimated empirical efficient score function, where $\Lambda_0(t)$ is replaced with $\hat{\Lambda}_0(t)$, is proposed as the second estimating function for β for clustered current status data in this article.

Let the jump of the counting process $N_{2ij}(t)$ at time t be denoted by $dN_{2ij}(t) = I(C_{ij} = t, T_{ij} < t)$. Since the estimated empirical efficient score function is based on both $N_{1ij}(t)$ and $N_{2ij}(t)$, it consists of two terms for each cluster which are p -dimensional vectors and are displayed as

$$\tilde{U}_{1i}(\beta, \hat{\Lambda}_0) = \sum_{j=1}^{m_i} \sum_{l=1}^D (Z_{ij}(c_l) - \tilde{Z}(\beta, \hat{\Lambda}_0, c_l)) I(C_{ij} = c_l, T_{ij} \geq c_l)$$

and

$$\tilde{U}_{2i}(\beta, \hat{\Lambda}_0) = \sum_{j=1}^{m_i} \sum_{l=1}^D (Z_{ij}(c_l) - \tilde{Z}(\beta, \hat{\Lambda}_0, c_l)) L_{ij}(\beta, \hat{\Lambda}_0, c_l) I(C_{ij} = c_l, T_{ij} < c_l),$$

where $Z(\beta, \hat{\Lambda}_0, t) = S^{(1)}(\beta, \hat{\Lambda}_0, t) / S^{(0)}(\beta, \hat{\Lambda}_0, t)$ is a p -dimensional vector with the p -

dimensional vector $\tilde{S}^{(r)}(\beta, \hat{\Lambda}_0, t) = \sum_{i=1}^n \sum_{j=1}^{m_i} Y_{ij}(t) L_{ij}(\beta, \hat{\Lambda}_0, t) (Z_{ij}(t))^{\otimes r}$, and $L_{ij}(\beta, \hat{\Lambda}_0, t) = p_{ij}(\beta, \hat{\Lambda}_0, t) / (1 - p_{ij}(\beta, \hat{\Lambda}_0, t))$ with $p_{ij}(\beta, \hat{\Lambda}_0, t) = e^{-\hat{\Lambda}_0(t) - \beta' Z_{ij}(t)}$. Then, the proposed estimated empirical efficient score function is

$$\tilde{U}(\beta, \hat{\Lambda}_0) = \sum_{i=1}^n \tilde{U}_i(\beta, \hat{\Lambda}_0),$$

where $\tilde{U}_i(\beta, \hat{\Lambda}_0) = \tilde{U}_{1i}(\beta, \hat{\Lambda}_0) - \tilde{U}_{2i}(\beta, \hat{\Lambda}_0)$. Following the derivation of Martinussen and Scheike [7], it can be verified that $\tilde{U}(\beta, \hat{\Lambda}_0)$ is a martingale if $\hat{\Lambda}_0(t)$ is a predictable estimator of $\Lambda_0(t)$. Then, $E(\tilde{U}(\beta, \hat{\Lambda}_0)) = 0$. Therefore, $\tilde{U}(\beta, \hat{\Lambda}_0)$ is an appropriate estimating function for β . Hence, a consistent estimate of β , $\hat{\beta}_E$, can be obtained by solving $\tilde{U}(\beta, \hat{\Lambda}_0) = 0$. To consider within-cluster dependence, the martingale technique developed in Ying and Wei [11] is again applied to derive the asymptotic distribution of $n^{-1/2} \tilde{U}(\beta, \hat{\Lambda}_0)$. The result is summarized in *Theorem 2*.

Theorem 2:

Under regularity conditions, the asymptotic distribution of $n^{-1/2} \tilde{U}(\beta, \hat{\Lambda}_0)$ is normal with mean vector 0 and a $p \times p$ variance-covariance matrix Σ_E , which is the limit of

$n^{-1} \sum_{i=1}^n \text{Var}(\tilde{U}_i(\beta, \hat{\Lambda}_0))$ as n goes to infinity. Moreover, the asymptotic distribution of $n^{1/2}(\hat{\beta}_E - \beta)$ is normal with mean vector 0 and a $p \times p$ variance-covariance matrix $\Sigma_{\hat{\beta}_E}$ equals $J^{-1}(\beta) \Sigma_E J^{-1}(\beta)$, where $J(\beta)$ is a $p \times p$ matrix and is the expected value of the second derivative of the full likelihood function derived from independent current status data with respect to β .

More details are given in Appendix B. Similarly, Σ_E can be estimated by a moment

estimator $\hat{\Sigma}_E = \sum_{i=1}^n \hat{\sigma}_i^{\otimes 2}(\beta, \hat{\Lambda}_0)$, where $\hat{\sigma}_i^{\otimes 2}(\beta, \hat{\Lambda}_0)$ is the moment estimator of the within-cluster variance-covariance matrix for the i th cluster and its explicit form is given in Appendix B. In practice the estimate $\hat{\beta}_E$ can be obtained by using the following equation iteratively until convergence:

$$\beta^{(k+1)} = \beta^{(k)} - [\hat{\Sigma}_E^{(k)}]^{-1} \tilde{U}(\beta^{(k)}, \hat{\Lambda}_0),$$

where $\beta^{(0)}$ is an initial value and $\beta^{(k)}$ is the k th parameter estimate at iteration k . To estimate the variance-covariance matrix $\Sigma_{\hat{\beta}_E}$ of $\hat{\beta}_E$, we recommend the use of a robust sandwich variance estimator

$$\hat{\Sigma}_{\hat{\beta}_E} = \hat{J}^{-1}(\beta) \left[\sum_{i=1}^n \hat{\sigma}_i^{\otimes 2}(\beta, \hat{\Lambda}_0) \right] \hat{J}^{-1}(\beta),$$

where $\hat{J}(\beta)$ is a $p \times p$ matrix and is the second derivative of the full likelihood function derived from independent current status data with respect to β . The expression of $\hat{J}(\beta)$ is given in Appendix B.

For testing the hypotheses $H_0 : \beta = \beta_0$ versus $H_a : \beta \neq \beta_0$, the score test and the Wald test

denoted by $\tilde{U}'(\beta_0, \hat{\Lambda}_0) \sum_E^{-1} \tilde{U}(\beta_0, \hat{\Lambda}_0)$ and $(\hat{\beta}_E - \beta_0)' \sum_{\hat{\beta}_E}^{-1} (\hat{\beta}_E - \beta_0)$, have an asymptotic χ_p^2 distribution with p degrees of freedom. The accuracy of these asymptotic distributions will also be examined in Section 3.

3. SIMULATION

To investigate the performance of the two proposed estimating functions, we conducted a two-part simulation study. In the first part, we investigated the performance of the estimators $\hat{\beta}_S$ and $\hat{\beta}_E$ under different scenarios. In the second part, we investigated the accuracy of asymptotic distributions of the score and the Wald test under the null hypothesis. Moreover, the power comparisons based on these tests for testing regression parameters were also presented.

3.1 Estimation results

Because intra-cluster correlation exists in each cluster, the survival times are generated from multivariate distributions. Under additive hazards models, the joint survival function of correlated survival times for the i th cluster can be derived from the Clayton-Oakes model [12, 13],

$$S(t_{i1}, \dots, t_{im_i} | X_{i1}, \dots, X_{im_i}) = \left\{ \sum_{j=1}^{m_i} S_m(t_{ij})^{-1/\theta} - (m_i - 1) \right\}^{-\theta},$$

where the marginal survival function is $S_m(t) = \exp(-\Lambda_0(t) - \beta'Z_{ij}(t))$. The parameter $\theta > 0$ is used to control the strength of correlation among subjects in each cluster. Since Kendall's $\tau = 1/(2\theta + 1)$, a larger θ induces smaller intra-cluster correlation. In simulation, $\theta = 0.2, 0.5$, and 2 are employed, and the corresponding Kendall's τ are $0.7, 0.5$ and 0.2 , respectively. In addition, survival times are generated independently from marginal distribution $S_m(t)$, and this setting is referred to as $\tau = 0$. As suggested by a referee, both constant ($\lambda_0(t) = 0.5$) and non-constant ($\lambda_0(t) = 0.5t$) baseline hazard functions in the additive hazards models are investigated in the simulation study. The simulation procedures proposed by Cai and Shen [14] are modified based on the marginal additive hazards models. Further, the monitoring times are generated from the exponential distribution with hazard rate equal to 0.5 . The regression parameter β is set to 0.5 . The covariate X_{ij} is generated from uniform $(0, \sqrt{12})$ distribution to investigate the accuracy of estimation procedures. The total number of clusters is set to $50, 100$ and 150 . Equal cluster size is considered for each cluster and cluster size is set to 2 and 4 . Moreover, for the efficient score function, the number of intervals for the sieve estimator of $\Lambda_0(t)$ is taken to be 5 . All comparison results are summarized from 2000 simulation runs.

The estimated biases, the sample standard errors (SE) and the average of the standard error estimates (SE_U) for constant and non-constant baseline hazard rates are displayed in Table 1

and Table 2, respectively. The SE for $\hat{\beta}_S$ is defined as $\left[\sum_{k=1}^{2000} (\hat{\beta}_{S(k)} - \bar{\beta}_S)^2 / 2000 \right]^{1/2}$, where $\hat{\beta}_{S(k)}$ is the estimate of β_S obtained in the k th run, and $\bar{\beta}_S$ is the mean of the estimated regression parameters from 2000 simulation runs. Likewise, the SE for $\hat{\beta}_E$ is defined in the same way as for $\hat{\beta}_S$. The SE_U for $\hat{\beta}_S$ is defined as the main diagonal elements of $\sum_{k=1}^{2000} \sum_{\beta_{S(k)}}^{1/2} / 2000$, where $\hat{\Sigma}_{\beta_{S(k)}}$ is the estimated variance-covariance matrix of regression parameter estimator $\hat{\beta}_S$ in the k th run. Similarly, the SE_U for $\hat{\beta}_E$ is defined in the same way as for $\hat{\beta}_S$. In addition, to examine a larger range of coefficients, we set the total number of clusters to 100 with equal cluster size of 2 for each cluster. The bias, SE and SE_U for β varying from 0 to 1 with respect to Kendall's tau correlations of 0 and 0.5 , respectively, are displayed in Figure 1.

In Table 1 and Table 2, the biases, SE and SE_U , as expected, decrease as sample size, the number of clusters or cluster size, increases. On the other hand, the biases, SE and SE_U slightly increase with an increase in intra-cluster correlation. Because the biases from the estimating procedures of Lin *et al.* [6] and Martinussen and Scheike [7] are slightly larger than that from the proposed procedures, the iterative algorithms proposed here to obtain the estimates of β can not effectively account for the intra-cluster correlation. However, the proposed estimators are consistent for β as demonstrated in both Tables 1 and 2. Moreover, the confidence levels of the confidence intervals for β based on the proposed procedures

should hold reasonably, because the asymptotic results works quite well for the Wald test as demonstrated in the next subsection. Comparing β_S with β_E , the biases of β_E are slightly smaller than that of β_S for most of the cases examined here. Whereas the SE and SE_U of β_E are always smaller than that of β_S , because the extended efficient score function extracts more information from the data.

As expected, the biases, SE, and SE_U increase with an increase in β in Figure 1 for both constant and non-constant baseline hazard functions. Even considering various values of β , the biases, SE and SE_U of β_E are also slightly smaller than that of β_S . Therefore, the estimator β_E obtained from the extended efficient score functions $\tilde{U}(\beta, \Lambda_0)$ is more efficient than β_S obtained from the extended partial score function $U(\beta)$.

3.2 Testing Hypothesis

The accuracy of the score test and Wald test derived from extended estimating functions is investigated. The joint survival functions of correlated survival times are also generated from the Clayton-Oakes model with 100 clusters for each scenario described in subsection 3.1. The test statistics considered are the partial score test proposed by Lin *et al.* [6], the efficient score function developed by Martinussen and Scheike [7], the score tests based on $U(\beta)$ and $\tilde{U}(\beta, \Lambda_0)$, and the Wald test, based on the estimator of β_S and β_E . The empirical type I error rates from these tests for testing the null hypothesis $H_0 : \beta = 0$ with nominal level 0.05 are listed in Table 3 for each of these tests.

As seen in Table 3, the type I error rates of the partial score test proposed by Lin *et al.* [6] and the efficient score function developed by Martinussen and Scheike [7] are much higher than the pre-specified significant level of 0.05 for $\tau > 0$. Further, the type I error rates of the two tests inflate severely for larger cluster sizes ($m_i = 4$). Whereas, the tests derived from the proposed estimating functions $U(\beta)$ and $\tilde{U}(\beta, \Lambda_0)$ have reasonable type I error rates in various scenarios. Therefore, the proposed estimating functions for clustered current status data perform very well. Note that in both simulations, nonconvergence occurred in some cases because of the unstable estimates of $e^{-\Lambda_0(t) - \beta'Z_{ij}(t)}$; these cases are excluded in our simulation.

Next, it is interesting to compare the power of the Wald tests with that of the score tests derived from the extended partial score function and the estimated empirical efficient score function for constant and non-constant baseline hazard functions scenarios, respectively. In this comparison, the total number of clusters is set to 100 with equal cluster size of 2 for each cluster. Likewise, the covariate considered is generated from uniform $(0, \sqrt{12})$. Simulation results show that the power of the Wald tests is almost the same as that of the score tests. For the sake of clarity, only the power curves of the Wald tests for β varying from 0 to 0.9 with respect to two Kendall's tau correlations of 0 and 0.5 are displayed in Figure 2. It can be seen that the power increases as β increases and slightly decreases as intra-cluster correlation increases. As expected, the power of the proposed estimated empirical efficient score function is higher than the proposed partial score function, because the proposed estimated empirical efficient score function makes more efficient use of the dataset.

4. AN EXAMPLE

In this section, the proposed estimating methods developed in Section 2 are illustrated through the cataract data set analyzed by Wen and Chen [8] who used the gamma-frailty Cox model to identify the risk factors for time to occurrence of cataract. This data set is a subset of the 2001 National Health Interview Survey Database from the National Health Research Institute in Taiwan and consists of 2023 subjects. The occurrence of cataracts for

the right and left eyes before or after examination time was recorded for elderly people aged 65 to 100 years, creating a bivariate current status data set. Here, two proposed estimating functions based on additive hazards models are used to identify the risk factors based on bivariate current status data. Two indicator covariates, gender and the presence of diabetes, are taken into consideration in the hazards models.

The estimated survival curves based on Turnbull's algorithm [16] are shown in Figure 3 for the combination of gender and diabetes. Because Figure 3 indicates that there is no interaction between these two covariates on survival rates, an additive hazards model with two time-independent covariates is applied. It seems that the age at examination is not affected by the covariates considered for this data set. Therefore, the two proposed estimating functions considered in Section 2 are employed to get parameter estimates. The number of intervals for the sieve estimator of $\Lambda_0(t)$ is taken to be 5. The p-values of the score tests and Wald tests are listed in Table 3. Using the extended partial score estimating function, $U(\beta)$, gender is statistically significant based on the Wald test ($p < 0.0001$), but diabetes is not ($p = 0.1314$). However, applying the extended efficient score function, $\tilde{U}(\beta, \hat{\Lambda}_0)$, both gender and diabetes are significant risk factors based on the Wald test ($p < 0.0001$ and $p = 0.0212$, respectively). The results based on Wald tests are consistent with that based on the score tests. The estimates of $S\hat{\beta}_S$ and $\hat{\beta}_E$ are quite similar, but the sandwich variance estimates of $\hat{\beta}_E$ are smaller than that of $\hat{\beta}_S$. This implies that $\hat{\beta}_E$ is a more efficient estimator. Therefore, based on the results from the extended efficient score function $\tilde{U}(\beta, \hat{\Lambda}_0)$, females have a higher cataract incidence rate than males, while diabetic patients have a higher cataract incidence rate than non-diabetes. These results are consistent with those in Wen and Chen [8].

5. CONCLUDING REMARKS

For clustered current status data, this paper developed two estimating functions for estimating the regression parameters in additive hazards models based on the ideas of partial score function proposed by Lin *et al.* [6] and the efficient score functions derived by Martinussen and Scheike [7], respectively. The procedures proposed here do not need to specify the structure of the correlation within clusters under independent monitoring. The point estimators derived from the proposed estimating functions are consistent as demonstrated by the simulation study. In general, the score test or the Wald test is used to determine the effect of covariates on survival in regression analysis. Therefore, the validity of these tests is crucial to identifying the covariate effect on risk. From the simulation results, the score tests and the Wald tests behaved reasonably well, as they are derived from the extended partial score function and the extended efficient score functions, which account for the intra-correlation. Therefore, these tests are very useful for clustered current status data, since they control type I error rates at the desired level. Moreover, these tests can be used to construct confidence intervals for regression parameters β , and hence the coverage probability is expected to hold at the pre-specified confidence level.

It is worth noting that if the monitoring time depends on the covariates, the cumulative hazard function of monitoring times can not be eliminated from the partial likelihood. Lin *et al.* [6] formulated the relationship between the monitoring times and covariates through the proportional hazards model to avoid any complicated nonparametric estimation of $\Lambda_c(t)$ and proposed a two-stage estimating procedure. Moreover, Martinussen and Scheike [7] employed kernel functions to estimate the hazard function of monitoring time. Therefore, these procedures can also be extended to clustered current status data in the future.

Acknowledgments

The authors wish to thank Margot Bjoring for editorial work on this manuscript. We also thank Dr. Chi-Chung Wen and Dr. Yi-Hau Chen for providing the cataract data set. This research of Yunchan Chi was supported by National Science Council grants 98-2118-M-006-003 -MY2 of Taiwan. This research of Pei-Fang Su was supported by National Cancer Institute (NCI) P30 CA068485, NCI P50 CA090949 and NCI P50 CA098131. We would like to express our sincere thanks and appreciation both to the editor and the reviewers for their helpful comments.

REFERENCES

1. Diamond ID, McDonald JW, Shah IH. Proportional hazards models for current status data: application of the study of differentials in age at weaning in Pakistan. *Demography*. 1986; 24:607–620. [PubMed: 3803657]
2. Groeneboom, P.; Wellner, J. Information Bounds and Nonparametric Maximum Likelihood Estimation. Birkhauser: Basel: 1992.
3. Frinkelstetn DM. A proportional hazards model for interval-censored failure time data. *Biometrics*. 1986; 42:845–854. [PubMed: 3814726]
4. Klein RW, Spady RH. An efficient semiparametric estimator for binary response models. *Econometrica*. 1993; 61:387–421.
5. Rabinowitz D, Tsiatis A, Aragon J. Regression with interval censored data. *Biometrika*. 1995; 82:501–513.
6. Lin DY, Oakes D, Ying Z. Additive hazards regression with current status data. *Biometrika*. 1998; 85:289–298.
7. Martinussen T, Scheike T. Efficient estimation in additive hazards regression with current status data. *Biometrika*. 2002; 89:649–658.
8. Wen CC, Chen YH. Nonparametric maximum likelihood analysis of clustered current status data with the gamma-frailty Cox model. *Computational Statistics and Data Analysis*. 2011; 55:1053–1060.
9. Liang KY, Zeger SL. Inference based on estimating functions in the presence of nuisance parameters. *Statistical Science*. 1995; 10:158–173.
10. Geman S, Hwang CR. Nonparametric maximum likelihood estimation by the method of sieves. *The annals of Statistics*. 1982; 10:401–414.
11. Ying Z, Wei LJ. The Kaplan-Meier estimate for dependent failure time observations. *Journal of Multivariate Analysis*. 1994; 50:17–29.
12. Clayton D, Cuzick J. Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society Series A*. 1985; 148:82–117.
13. Oakes D. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*. 1989; 84:487–493.
14. Cai J, Shen Y. Permutation tests for comparing marginal survival functions with clustered failure time data. *Statistics in Medicine*. 2000; 19:2963–2973. [PubMed: 11042626]
15. Banerjee M, Wellner JA. Score statistics for current status data: comparisons with likelihood ratio and Wald statistics. *The International Journal of Biostatistics*. 2005; 1 Article 3.
16. Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B*. 1976; 38:290–295.
17. Su, P. PhD thesis. Department of Statistics, National Cheng Kung University; 2010. Regression analysis for clustered current status data based on additive hazards models.

APPENDIX A

First, the regularity conditions mentioned in *Theorem 1* and *Theorem 2*, which are shared with the two estimating equations, are given below:

The regularity conditions are:

- i. Let $s_1(\beta, t) = E(S^{(1)}(\beta, t))$, $s_0(\beta, t) = E(S^{(0)}(\beta, t))$ and $z(\beta, t) = s_1(\beta, t)/s_0(\beta, t)$. There exists an open neighborhood $\mathcal{O}$ of β and functions $s_j(\beta, t)$, $j = 0, 1$ defined on $\mathcal{O} \times [0, \tau]$, where τ is the largest observed time.
- ii. The covariate $X_{ij}(t)$ is bounded.
- iii. $\sup_{t \in [0, \tau]} |\bar{Z}(\beta, t) - \bar{z}(\beta, t)|$ converges to zero as $n \rightarrow \infty$.
- iv. $s_j(\beta, t)$ is bounded away from 0 for $t \in [0, \tau]$. This guarantees that the risk set becomes large at each t and guarantees estimability of the baseline cumulative hazard function everywhere in $[0, \tau]$.
- v. The distribution function of the monitoring times has density function g with respect to Lebesgue measure.
- vi. Dependency is assumed to exist only among each cluster and each cluster size is bounded and small relative to n .

Next, the expression of the estimators of variance-covariance matrix of $U(\beta)$ and $\hat{\beta}_S$ is given below. To consider within-cluster dependence, the martingale technique developed in Ying and Wei [11] is applied. Because the data in different clusters are independent, $U_i(\beta)$ and $U_j(\beta)$ are independent for $i \neq j$. The variance-covariance matrix Σ_S can be derived by summing the within-cluster variance-covariance of $U_i(\beta)$; that is, Σ_S is the limit of

$n^{-1} \sum_{i=1}^n \text{Var}(U_i(\beta))$, where $\text{Var}(U_i(\beta))$ can be computed by $E(U_i(\beta)U_i'(\beta))$. Consequently, Σ_S can be estimated by $\sum_{i=1}^n \hat{\sigma}_i^{\otimes 2}(\beta)$ where $\hat{\sigma}_i(\beta)$ is the moment estimator of the within-cluster variance-covariance matrix for the i th cluster and it can be expressed as

$$n^{-1} \sum_{j=1}^{m_i} \sum_{l=1}^D (Z_{ij}(c_l) - \bar{Z}(\beta, c_l)) \left[I(C_{ij}=c_l, T_{ij} \geq c_l) - Y_{ij}(c_l) e^{-\beta' Z_{ij}(c_l)} d\hat{H}_0(c_l) \right],$$

where

$$d\hat{H}_0(c_l) = \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{I(C_{ij}=c_l, T_{ij} \geq c_l)}{\sum_{h=1}^n \sum_{k=1}^{m_h} Y_{hk}(c_l) e^{-\beta' Z_{hk}(c_l)}},$$

which is an estimate of $e^{-\Lambda_0(t)} d\Lambda_c(t)$. More details of the derivation are given in Su [17].

Based on Taylor series expansion, the asymptotic distribution of $n^{1/2}(\hat{\beta}_S - \beta)$ is normal with mean vector 0 and a $p \times p$ variance-covariance matrix Σ_{β_S} , where Σ_{β_S} is given in *Theorem 1*. It can be consistently estimated by the robust sandwich variance estimator

$$\hat{\Sigma}_{\beta_S} = \hat{J}^{-1}(\beta) \hat{\Sigma}_S \hat{J}^{-1}(\beta) = \hat{J}^{-1}(\beta) \left[\sum_{i=1}^n \hat{\sigma}_i^{\otimes 2}(\beta) \right] \hat{J}^{-1}(\beta),$$

where

$$\hat{J}(\beta) = n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{l=1}^D \left\{ \frac{S^{(2)}(\beta, c_l)}{S^{(0)}(\beta, c_l)} - \frac{S^{(1)}(\beta, c_l)^{\otimes 2}}{S^{(0)}(\beta, c_l)} \right\} I(C_{ij}=c_l, T_{ij} \geq c_l),$$

with $a^{\otimes 2} = aa'$.

APPENDIX B

Note that $\text{Var}(\tilde{U}_i(\beta, \hat{\Lambda}_0))$ can be decomposed as

$$E(\tilde{U}_{1i}^{\otimes 2}(\beta, \hat{\Lambda}_0)) + E(\tilde{U}_{2i}^{\otimes 2}(\beta, \hat{\Lambda}_0)) - 2E(\tilde{U}_{1i}(\beta, \hat{\Lambda}_0)\tilde{U}_{2i}'(\beta, \hat{\Lambda}_0)).$$

Next, each term of the above expression can be consistently estimated by a moment estimator. For example, the first term $E(\tilde{U}_{1i}(\beta, \hat{\Lambda}_0)\tilde{U}_{1i}'(\beta, \hat{\Lambda}_0))$ can be consistently estimated by the moment estimator $V_{1i}^{\otimes 2}(\beta, \hat{\Lambda}_0)$, where $V_{1i}(\beta, \hat{\Lambda}_0)$ is

$$n^{-1} \sum_{j=1}^{m_i} \sum_{l=1}^D (Z_{ij}(c_l) - \tilde{Z}(\beta, \hat{\Lambda}_0, c_l)) \left[I(C_{ij}=c_l)I(T_{ij} \geq c_l) - Y_{ij}(c_l)p_{ij}(\beta, \hat{\Lambda}_0, c_l)d\hat{\Lambda}_c(c_l) \right],$$

with the Aalen-Breslow type estimator

$$d\hat{\Lambda}_c(c_l) = \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{I(C_{ij}=c_l)}{\sum_{h=1}^n \sum_{k=1}^{m_h} Y_{hk}(c_l)e^{-\beta'Z_{hk}(c_l)}}.$$

Similarly, the second term $E(\tilde{U}_{2i}(\beta, \hat{\Lambda}_0)\tilde{U}_{2i}'(\beta, \hat{\Lambda}_0))$ can be consistently estimated by the moment estimator $V_{2i}^{\otimes 2}(\beta, \hat{\Lambda}_0)$, where $V_{2i}(\beta, \hat{\Lambda}_0)$ is

$$n^{-1} \sum_{j=1}^{m_i} \sum_{l=1}^D (Z_{ij}(c_l) - \tilde{Z}(\beta, \hat{\Lambda}_0, c_l))L_{ij}(\beta, \hat{\Lambda}_0, c_l) \left[I(C_{ij}=c_l)I(T_{ij} < c_l) - Y_{ij}(c_l)(1 - p_{ij}(\beta, \hat{\Lambda}_0, c_l))d\hat{\Lambda}_c(c_l) \right].$$

As a result, $E(\tilde{U}_{1i}(\beta, \hat{\Lambda}_0)\tilde{U}_{2i}'(\beta, \hat{\Lambda}_0))$ can be consistently estimated by the moment estimator $V_{1i}(\beta, \hat{\Lambda}_0)V_{2i}'(\beta, \hat{\Lambda}_0)$. Therefore, a consistent estimator of Σ_E can be written as

$$\sum_{i=1}^n (V_{1i}^{\otimes 2}(\beta, \hat{\Lambda}_0) + V_{2i}^{\otimes 2}(\beta, \hat{\Lambda}_0) - 2V_{1i}(\beta, \hat{\Lambda}_0)V_{2i}'(\beta, \hat{\Lambda}_0)).$$

Likewise, the moment estimator of $\text{Var}(\tilde{U}_i(\beta, \hat{\Lambda}_0))$ is

$$\hat{\sigma}_i^{\otimes 2}(\beta, \hat{\Lambda}_0) = \sum_{i=1}^n (V_{1i}^{\otimes 2}(\beta, \hat{\Lambda}_0) + V_{2i}^{\otimes 2}(\beta, \hat{\Lambda}_0) - 2V_{1i}(\beta, \hat{\Lambda}_0)V_{2i}'(\beta, \hat{\Lambda}_0)).$$

To estimate the variance-covariance Σ_{β_E} of $\hat{\beta}_E$, we recommend the use of a robust sandwich variance estimator

$$\hat{J}^{-1}(\beta) \left[\sum_{i=1}^n \hat{\sigma}_i^{\otimes 2}(\beta, \hat{\Lambda}_0) \right] \hat{J}^{-1}(\beta)$$

where $\hat{J}(\beta)$ is derived from independent current status data that

$$\begin{aligned} \hat{J}(\beta) = & n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{l=1}^D \frac{\bar{S}_0 \tilde{S}^{(1)} - \bar{S}_1 \tilde{S}^{(0)}}{(\tilde{S}^{(0)})^2} (Z_{ij} \\ & - \frac{\tilde{S}^{(1)}}{\tilde{S}^{(0)}} \left(\frac{p_{ij}}{1 - p_{ij}} I(C_{ij} \right. \\ & = c_l, T_{ij} \geq c_l) - I(C_{ij} \\ & = c_l, T_{ij} < c_l)) \\ & - \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{l=1}^D (Z_{ij} \\ & - \frac{\tilde{S}^{(1)}}{\tilde{S}^{(0)}} Z'_{ij} \frac{p_{ij}}{(1 - p_{ij})^2} I(C_{ij} \\ & = c_l, T_{ij} \geq c_l), \end{aligned}$$

where $S^{(k)} = S^{(k)}(\beta, \hat{\Lambda}_0, c_l)$, for $k = 0, 1$, $p_{ij} = p_{ij}(\beta, \hat{\Lambda}_0, c_l)$, $Z_{ij} = Z_{ij}(c_l)$, and

$$\bar{S}_k = \bar{S}_k(\beta, \hat{\Lambda}_0, c_l) = - \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{\hat{p}_{ij}(\beta, \hat{\Lambda}_0, c_l)}{(1 - \hat{p}_{ij}(\beta, \hat{\Lambda}_0, c_l))^2} Y_{ij}(c_l) Z_{ij}^{\otimes k+1}(c_l) \text{ for } k=0, 1.$$

for $k = 0, 1$.

For clustered data, an estimator for $\Lambda_0(t)$ denoted by $\hat{\Lambda}_0(t)$, which is analogous to the sieve estimator of $\Lambda_0(t)$ for independent current status data, is proposed. Let $a_0 = 0$, $a_1, a_2, \dots$, and a_m be the selected ordered monitoring time points where a_m is the largest observed monitoring time and m is the number of intervals. The formula of the estimator for cumulative hazard function is

$$\hat{\Lambda}_0(a_k) = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} I(C_{ij} \leq a_k)}{\sum_{i=1}^n m_i}, k=1, \dots, m,$$

Note that $\hat{\Lambda}_0(t)$ is a predictable estimator of $\Lambda_0(t)$.

(A) Constant baseline hazard function

(B) Non-constant baseline hazard function

Figure 1.

Bias, SE and SE_U under various values of β for constant baseline hazard and non-constant baseline hazard functions.

Figure 2.
Empirical power of the Wald tests under various values of β for $\alpha = 0.05$.

Figure 3.
Estimated survival functions for cataract data set.

Table 1

Estimated biases, sample standard errors and the average of the standard error estimates based on the constant baseline hazard function $\lambda_0(t) = 0.5$.

Cluster size	τ	n	$\hat{\beta}_S$			$\hat{\beta}_E$		
			bias	SE	SE _U	bias	SE	SE _U
2	0.0	50	0.0462	0.2948	0.2658	0.0331	0.2598	0.2459
		100	0.0266	0.2103	0.1837	0.0233	0.1762	0.1663
		150	0.0165	0.1606	0.1494	0.0129	0.1452	0.1350
	0.2	50	0.0472	0.3053	0.2715	0.0351	0.2701	0.2495
		100	0.0224	0.1979	0.1840	0.0257	0.1844	0.1683
		150	0.0131	0.1595	0.1496	0.0140	0.1450	0.1355
	0.5	50	0.0573	0.3145	0.2753	0.0443	0.2743	0.2538
		100	0.0296	0.2069	0.1876	0.0268	0.1801	0.1716
		150	0.0171	0.1616	0.1514	0.0135	0.1472	0.1380
	0.7	50	0.0667	0.3316	0.2786	0.0444	0.2806	0.2545
		100	0.0341	0.2053	0.1870	0.0208	0.1899	0.1741
		150	0.0285	0.1600	0.1542	0.0141	0.1490	0.1382
4	0.0	50	0.0267	0.1994	0.1807	0.0190	0.1712	0.1664
		100	0.0112	0.1352	0.1289	0.0104	0.1235	0.1165
		150	0.0074	0.1077	0.1049	0.0067	0.0977	0.0951
	0.2	50	0.0248	0.2035	0.1869	0.0279	0.1814	0.1721
		100	0.0144	0.1422	0.1328	0.0149	0.1275	0.1203
		150	0.0119	0.1138	0.1080	0.0094	0.1043	0.0984
	0.5	50	0.0439	0.2205	0.1910	0.0313	0.2006	0.1785
		100	0.0189	0.1451	0.1362	0.0203	0.1352	0.1253
		150	0.0145	0.1197	0.1123	0.0073	0.1079	0.1027
	0.7	50	0.0392	0.2228	0.1940	0.0382	0.2066	0.1804
		100	0.0194	0.1432	0.1377	0.0170	0.1351	0.1257
		150	0.0135	0.1202	0.1130	0.0093	0.1105	0.1035

Table 2

Estimated biases, sample standard errors and the average of the standard error estimates based on the non-constant baseline hazard function $\lambda_0(t) = 0.5t$.

Cluster size	τ	n	$\hat{\beta}_S$			$\hat{\beta}_E$		
			bias	SE	SE _U	bias	SE	SE _U
2	0.0	50	0.0399	0.2583	0.2594	0.0279	0.2432	0.2314
		100	0.0170	0.1862	0.1793	0.0194	0.1658	0.1600
		150	0.0122	0.1504	0.1461	0.0085	0.1325	0.1290
	0.2	50	0.0402	0.2905	0.2645	0.0357	0.2496	0.2342
		100	0.0148	0.1906	0.1812	0.0175	0.1684	0.1587
		150	0.0079	0.1515	0.1478	0.0110	0.1385	0.1303
	0.5	50	0.0466	0.3023	0.2641	0.0310	0.2598	0.2354
		100	0.0203	0.1980	0.1831	0.0259	0.1735	0.1621
		150	0.0093	0.1531	0.1488	0.0079	0.1392	0.1318
	0.7	50	0.0383	0.2985	0.2636	0.0219	0.2631	0.2369
		100	0.0236	0.2016	0.1826	0.0174	0.1684	0.1610
		150	0.0155	0.1484	0.1494	0.0092	0.1350	0.1311
4	0.0	50	0.0127	0.1825	0.1779	0.0222	0.1612	0.1595
		100	0.0101	0.1280	0.1262	0.0042	0.1174	0.1120
		150	0.0038	0.1035	0.1032	0.0026	0.0942	0.0915
	0.2	50	0.0264	0.1963	0.1836	0.0258	0.1777	0.1638
		100	0.0069	0.1324	0.1289	0.0076	0.1220	0.1148
		150	0.0063	0.1108	0.1060	0.0001	0.0996	0.0941
	0.5	50	0.0301	0.2010	0.1872	0.0241	0.1834	0.1688
		100	0.0113	0.1389	0.1329	0.0183	0.1306	0.1190
		150	0.0067	0.1110	0.1084	0.0078	0.1028	0.0972
	0.7	50	0.0316	0.2089	0.1890	0.0246	0.1880	0.1677
		100	0.0139	0.1397	0.1331	0.0157	0.1258	0.1201
		150	0.0088	0.1143	0.1097	0.0101	0.1009	0.0988

Table 3

Type I error rates of various tests under constant baseline hazard function.

Cluster size	τ	Score test					Wald test	
		Partial Score Function	Efficient Score Function	$U(\beta)$	$\hat{D}(\beta, \Delta_0)$	$\hat{\beta}_s$	$\hat{\beta}_E$	
2	0.0	0.048	0.051	0.047	0.049	0.048	0.055	
	0.2	0.070	0.065	0.054	0.050	0.054	0.049	
	0.5	0.076	0.076	0.051	0.044	0.048	0.058	
	0.7	0.087	0.089	0.051	0.047	0.049	0.045	
4	0.0	0.047	0.049	0.049	0.041	0.053	0.044	
	0.2	0.089	0.102	0.052	0.048	0.053	0.041	
	0.5	0.123	0.149	0.053	0.047	0.055	0.048	
	0.7	0.145	0.166	0.054	0.048	0.048	0.042	

Table 4

Parameter estimates for cataract data set.

Method	Covariate	Estimate	Wald test p-value	Score test p-value
$\hat{\beta}_S$	Gender	0.00291 (0.00071)	<0.0001 *	0.0103 *
	Diabetes	0.00145 (0.00096)	0.1314	0.1345
$\hat{\beta}_E$	Gender	0.00271 (0.00046)	<0.0001 *	<0.0001 *
	Diabetes	0.00146 (0.00063)	0.0212 *	0.0433 *

*
p-value is smaller than 0.05.