Semiparametric Multivariate Accelerated Failure Time Model
with Generalized Estimating Equations

Sy Han Chiou, Junghi Kim, and Jun Yan

University of Connecticut, University of Minnesota
and University of Connecticut Health Center

Abstract: The semiparametric accelerated failure time model is not as widely used as the Cox relative
risk model mainly due to computational difficulties. Recent developments in least squares estimation
and induced smoothing estimating equations provide promising tools to make the accelerate failure time
models more attractive in practice. For semiparametric multivariate accelerated failure time models,
we propose a generalized estimating equation approach to account for the multivariate dependence
through working correlation structures. The marginal error distributions can be either identical as in
sequential event settings or different as in parallel event settings. Some regression coefficients can be
shared across margins as needed. The initial estimator is a rank-based estimator with Gehan’s weight,
but obtained from an induced smoothing approach with computation ease. The resulting estimator
is consistent and asymptotically normal, with a variance estimated through a multiplier resampling
method. In a simulation study, our estimator was up to three times as efficient as the initial estimator,
especially with stronger multivariate dependence and heavier censoring percentage. Two real examples
demonstrate the utility of the proposed method.

Key words and phrases: efficiency; induced smoothing; least squares; multivariate survival.

1 Introduction

Multivariate failure times are frequently encountered in biomedical research where failure times are
clustered. For example, a diabetic retinopathy study assessed the efficacy of a laser treatment on
decelerating vision loss, measured by time to blindness in the left eye and in the right eye from the
same patient with diabetes (Diabetic Retinopathy Study Research Group, 1976); a colon cancer
study evaluated the treatment effects on prolonging the time to tumor recurrence and time to death
(Lin, 1994). The failure times within the same cluster are associated. Even though the primary
interest most often lies in the marginal effects of covariates on the failure times, accounting for
the within-cluster dependence may lead to more efficient regression coefficient estimators. For non-
censored multivariate data, the generalized estimating equations (GEE) approach (Liang and Zeger,
1986) has become an important piece in statisticians’ toolbox for marginal regression. For censored
multivariate failure times, the marginal accelerated failure time (AFT) model is a counterpart of the
marginal model. This paper aims to develop a GEE approach to make inferences for multivariate
AFT models, taking advantage of recent developments on AFT models with least squares and
induced smoothing.

A semiparametric AFT model is a linear model for the logarithm of the failure times with
error distribution unspecified. A nice interpretation is that the effect of a covariate is to multiply
the predicted failure time by some constant. It provides an attractive alternative to the popular
relative risk model (Cox, 1972). Three main classes of estimator exist for univariate AFT models.
The Buckley–James (BJ) estimator extends the least squares principle to accommodate censor-
ing through an expectation–maximization (EM) algorithm which iterates between imputing the
censored failure times and least squares estimation (Buckley and James, 1979). Despite the nice asymptotic properties (Lai and Ying, 1991; Ritov, 1990), the BJ estimator may be hard to get as the EM algorithm may not converge. Further, the limiting covariance matrix is difficult to estimate because it involves the unknown hazard function of the error term. The second class is the rank-based estimator motivated by inverting the weighted log-rank test (Prentice, 1978). Its asymptotic properties have been rigorously studied by Tsiatis (1990) and Ying (1993). Due to lack of efficient and reliable computing algorithm, the rank-based estimator has not been widely used in practice until recently, with numerical strategies for drawing inference developed by Huang (2002) and Strawderman (2005). The third class is obtained by minimizing an inverse probability of censoring weighed (IPCW) loss function (Robins and Rotnitzky, 1992). The IPCW estimator is easy to compute, consistent and asymptotically normal (Stute, 1993, 1996; Zhou, 1992), but it requires correct specification of the conditional censoring distribution and overlapping of the supports of the censoring time and the failure time.

More recent works have led to a promising perspective on bringing AFT models into routine data analysis practice. For rank-based inference, Jin et al. (2003) proposed a linear programming approach, exploiting the fact that the weighted rank estimating equation is the gradient of an objective function which can be readily solved by linear programming. Variances of the estimators are obtained from a resampling method. A computationally more efficient approach for rank-based inference with Gehan’s weight (Gehan, 1965) is the induced smoothing procedure of Brown and Wang (2007). This approach is an application of the general induced smoothing method of Brown and Wang (2005), where the discontinuous estimating equations are replaced with a smoothed version, whose solutions are asymptotically equivalent to those of the former. The smoothed estimating equations are differentiable, which facilitates rapid numerical solution and sandwich variance estimator. Jin et al. (2006a) suggested an iterative least-squared procedure that starts from a consistent and asymptotically normal initial estimator such as the one obtained from the rank-based method of Jin et al. (2003). The resulting estimator is consistent and asymptotically normal, with variance estimated from a multiplier resampling approach.

For multivariate AFT models, Jin et al. (2006b) developed rank-based estimating equations that are solved via linear programming for marginal regression parameters. Johnson and Strawderman (2009) extended the induced smoothing approach for a rank-based estimator with Gehan’s weight to the case of clustered failure times and showed that the smoothed estimates perform as well as those from the best competing methods at a fraction of the computational cost. Jin et al. (2006a) considered their least squares method with marginal models for multivariate failure times. All these approaches used independent working model and left the within-cluster dependence structure unspecified. Li and Yin (2009) developed a generalized method of moments approach for rank-based estimator using the quadratic inference function approach (Qu et al., 2000) to incorporate within-cluster dependence. Wang and Fu (2011) incorporated within-cluster ranks for the Gehan type estimator with the aid of induced smoothing. To the best of our knowledge, little work has been done to extend the GEE approach to the setting of multivariate AFT models except a technical report (Hornsteiner and Hamerle, 1996), where the BJ estimator was combined with GEE. Nevertheless, having no access to recent advances on AFT models, they did not solve the convergence problems, and their asymptotic variance estimator formula could not be easily computed because it depends on the derivatives of imputed failure times with respect to regression parameters, which might explain their overestimation of the variance.

We propose an iterative GEE procedure to account for multivariate dependence through a working covariance or weight matrix. This method has the same spirit as GEE in that misspecification of the working covariance matrix does not affect the consistency of the parameter estimator in the
marginal AFT models; when the working covariance is close to the unknown truth, the estimator has higher efficiency than that from working independence as used in Jin et al. (2006a). Our initial estimator is the computationally efficient, rank-based estimator from Johnson and Strawderman (2009), whose consistency and asymptotic normality is inherited by the resulting GEE estimator. We develop methods for cases where all marginal distributions are identical and for cases where at least two margins are different. Regression coefficients can be the same or partially the same across margins as needed.

The rest of the article is organized as follows. The semiparametric multivariate accelerated failure time model and the notation are introduced in Section 2. In Section 3, we propose an iterative GEE procedure to update a consistent and asymptotically normal initial estimator and present asymptotic properties of our estimator. A large scale simulation study is reported in Section 4 to assess the properties of the proposed estimator. The proposed methods are illustrated with the two aforementioned real applications in Section 5. In particular, some new findings are reported in analyzing the diabetic retinopathy study. A discussion concludes in Section 6. The sketch of proofs are relegated to the appendix.

2 Multivariate Accelerated Failure Time Model

There are two types of multivariate failure times depending on whether the multiple events are parallel or sequential. The difference between the two types is that the dimension is fixed for parallel data while random for sequential data. In a regression model, we generally have different covariates and different coefficients at each margin for parallel data. For sequential data, however, some or all covariates and covariate coefficients may be the same across margins. In general, it is desirable to allow some of the regression coefficients to be shared across margins as needed. We develop the methodology for parallel data for notational simplicity but comment when appropriate on how to adapt to sequential data.

Consider a random sample formed by $n$ clusters. For parallel data, all clusters are of size $K$ while for sequential data, cluster $i$ may have size $K_i$. For ease of notation, assume at the moment that the cluster sizes are all equal to $K$. For $i = 1, \ldots, n$ and $k = 1, \ldots, K$, let $T_{ik}$ and $C_{ik}$ be, respectively, the log-transformed failure time and censoring time for margin $k$ in cluster $i$. Let $Y_{ik} = \min(T_{ik}, C_{ik})$ and $\Delta_{ik} = I(T_{ik} < C_{ik})$. We stack $Y_{ik}, T_{ik}, C_{ik}$, and $\Delta_{ik}, k = 1, \ldots, K$, to form $K \times 1$ vector $Y_i, T_i, C_i$, and $\Delta_i$, respectively. Let $X_i = (X_{i1}, \ldots, X_{iK})^T$ be a $K \times p$ covariate matrix, with the $k$th row denoted by $X_{ik}$. The observed data are independent and identically distributed copies of $\{Y, \Delta, X\} : \{(Y_i, \Delta_i, X_i) : i = 1, \ldots, n\}$. We assume that $T_i$ and $C_i$ are conditionally independent given $X_i$.

Our multivariate accelerated failure time model is

$$T_i = X_i\beta + \epsilon_i,$$

where $\beta$ is a $p \times 1$ vector of regression coefficients, and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iK})^T$ is a random error vector with an unspecified multivariate distribution. This formulation accommodates margin-specific regression coefficients, in which case, $\beta$ is a stack of all marginal coefficients, and $X_i$ is a block diagonal matrix. The error vectors $\epsilon_i$’s, $i = 1, \ldots, n$, are independent and identically distributed. For parallel data, the $K$ marginal distributions can be all different, while for sequential data, the number of unique marginal distributions may be smaller or even one as in a recurrent event setting.
With right censoring, Buckley and James (1979) replaced each response $T_{ik}$ with its conditional expectation $\hat{Y}_{ik}(\beta) = E_\beta(T_{ik}|Y_{ik}, \Delta_{ik}, X_{ik})$, where the expectation is evaluated at regression coefficients $\beta$. Let $\hat{Y}_i(\beta) = (\hat{Y}_{i1}(\beta), \ldots, \hat{Y}_{iK}(\beta))^\top$. Jin et al. (2006a) defined

$$U_n(\beta, b) = \sum_{i=1}^n (X_i - \bar{X})^\top \left( \hat{Y}_i(b) - X_i\hat{\beta} \right) = 0,$$

where $\bar{X} = \sum_{i=1}^n X_i/n$, and $b$ is an initial estimator of $\beta$. The solution for $U_n(\beta, b)$ is the Buckley-James estimator. The advantage for fixing the initial value $b$ is to avoid solving for $U_n(\beta, \beta)$ which is neither continuous nor monotone in $\beta$. Let the $L_n(b)$ be the solution for $U_n(\beta, b) = 0$ given $b$. Then $L_n(b)$ has a closed-form,

$$L_n(b) = \left[ \sum_{i=1}^n (X_i - \bar{X})^\top (X_i - \bar{X}) \right]^{-1} \left[ \sum_{i=1}^n (X_i - \bar{X})^\top \left( \hat{Y}_i(b) - \bar{Y}(b) \right) \right],$$

where $\bar{Y}(b) = \sum_{i=1}^n \hat{Y}_i(b)/n$. Equation (3) leads to an iterative algorithm: $\beta_n^{(m)} = L_n(\beta_n^{(m-1)})$, $m \geq 1$. If the initial estimator $b$ is consistent and asymptotically normal, $\beta_n^{(m)}$ is consistent and asymptotically normal for every $m$.

Although this estimator is consistent, its efficiency might be low because it completely ignores the within-cluster dependence. We next propose to accommodate dependence using the GEE approach, which covers the estimator of Jin et al. (2006a) as a special case with working independence.

### 3 Inference with GEE

For a given initial estimator $b$ of $\beta$, we propose an updated estimator by solving the GEE

$$U_n(\beta, b, \alpha) = \sum_{i=1}^n (X_i - \bar{X})^\top \Omega_i^{-1}(\alpha(b)) \left( \hat{Y}_i(b) - X_i\hat{\beta} \right) = 0,$$

where $\bar{X} = \sum_{i=1}^n X_i/n$, and $\Omega_i^{-1}(\alpha(b))$ is a $K \times K$ nonsingular working weight matrix which may involve additional working parameters $\alpha$, which may depend on $b$. For given $\alpha$ and $b$, the solution of the GEEs (4) has a closed-form

$$L_n(b, \alpha) = \left[ \sum_{i=1}^n (X_i - \bar{X})^\top \Omega_i^{-1}(\alpha(b)) (X_i - \bar{X}) \right]^{-1} \left[ \sum_{i=1}^n (X_i - \bar{X})^\top \Omega_i^{-1}(\alpha(b)) \left( \hat{Y}_i(b) - \bar{Y}(b) \right) \right].$$

This process can be carried out iteratively, summarized as follows.

1. Obtain an initial estimate $\hat{\beta}_n^{(0)} = b_n$ of $\beta$ and initialize with $m = 1$.
2. Obtain an estimate $\hat{\alpha}_n$ of $\alpha$ given $\hat{\beta}_n^{(m-1)}$, $\hat{\alpha}_n(\hat{\beta}_n^{(m-1)})$.
3. Update with $\hat{\beta}_n^{(m)} = L_n(\hat{\beta}_n^{(m-1)}, \hat{\alpha}_n)$.
4. Increase $m$ by one and repeat 2 and 3 until convergence.
As in Jin et al. (2006a), a consistent and asymptotically normal estimator is important for avoiding convergence problems. We propose to use the rank-based estimator with Gehan’s weight from the induced smoothing approach of Johnson and Strawderman (2009). This estimator has the same asymptotic property as the non-smoothed version in Jin et al. (2003), but can be obtained with computation ease; its finite sample performance was also reported to be as well as the best competing methods (Johnson and Strawderman, 2009).

The GEEs are most efficient when $\Omega_i$ is chosen to be the covariance matrix of $\tilde{Y}_i(b)$. When $\Omega_i$’s are the identity matrix (working independence with all marginal variances the same), our estimator reduces to the least squares estimator of Jin et al. (2006a). The working covariance matrix $\Omega_i$’s are the same when all clusters have the same size $K$; they only vary with $i$ when the cluster sizes are not equal.

For convenience, we assume from now on that $E(\epsilon_{ik}) = 0$, $i = 1, \ldots, n$, $k = 1, \ldots, K$. This can be achieved by incorporating appropriate columns of ones in $X_i$, and, hence, adding intercepts in $\beta$. Our construction of working covariance involves filling element $\Omega_{kl}$, for $k, l \in \{1, \cdots, K\}$, of the working covariance matrix $\Omega$. To allow arbitrary number of unique marginal distributions, let $m_k \in \{1, \ldots, \kappa\}$ be the index of the $k$th margin among the $\kappa$ unique marginal distributions. The conditional expectation $\tilde{Y}_{ik}(b)$ is computed as

$$\tilde{Y}_{ik}(b) = \Delta_{ik}Y_{ik} + (1 - \Delta_{ik}) \left[ \frac{\int_0^\infty u \, d\tilde{F}_{k,b}(u)}{1 - \tilde{F}_{k,b}(e_{ik}(b))} + X_{ik}^\top b \right],$$

where $e_{ik}(b) = Y_{ik} - X_{ik}^\top b$ is the right-censored error evaluated at $b$, and $\tilde{F}_{k,b}$ is the pooled Kaplan–Meier estimator of the distribution function $F_{k,b}$ from the transformed data $\{e_{ir}(b), \Delta_{ir}; m_r = m_k\}$, which share the same margin $m_k$. Specifically, $\tilde{F}_{k,b}$ is

$$\tilde{F}_{k,b}(t) = 1 - \prod_{1 \leq i \leq n, 1 \leq r \leq K; m_r = m_k, e_{ir} < t} \left( 1 - \frac{\Delta_{ir}}{\sum_{j=1}^n \sum_{1 \leq t \leq K; m_t = m_k} I(e_{jr}(b) \geq e_{ir}(b))} \right).$$

To fill the diagonal elements $\Omega_{kk}$, $1 \leq k \leq K$, evaluate the conditional second moment of $\epsilon_{ik}(b)$ given the observed data:

$$\hat{\epsilon}_{ik}(b) = \Delta_{ik}e_{ik}^2(b) + (1 - \Delta_{ik}) \frac{\int_0^\infty u^2 \, d\tilde{F}_{k,b}(u)}{1 - \tilde{F}_{k,b}(e_{ik}(b))}, \quad i = 1, \ldots, n, \quad k = 1, \ldots, K. \quad (6)$$

For a given $b$, we fill $\Omega_{kk}$ by an unbiased estimator of $\text{Var}(\epsilon_{ik}(b))$

$$\hat{\Omega}_{kk}(b) = \frac{\sum_{1 \leq i \leq n, 1 \leq r \leq K; m_r = m_k} \hat{\epsilon}_{ik}(b)}{n \sum_{1 \leq r \leq K} I(m_r = m_k)}. \quad (7)$$

To fill the off-diagonal elements $\Omega_{kl}$, $k \neq l$, define

$$\hat{e}_{ik}(b) = \hat{Y}_{ik}(b) - X_{ik}^\top b, \quad i = 1, \ldots, n, \quad k = 1, \ldots, K, \quad (8)$$

the conditional expectation of $\epsilon_{ik}(b)$ given the observed data. Only when $\Delta_{ik} = 1$ is $\hat{e}_{ik}(b)$ equal to $e_{ik}(b)$. For a given $b$, we fill $\Omega_{kl}$, $k \neq l$, by

$$\hat{\Omega}_{kl}(b) = \frac{1}{n} \sum_{i=1}^n \hat{e}_{ik}(b)\hat{e}_{il}(b). \quad (9)$$
Because the construction of \( \hat{e}_{ik}(b) \) does not involve the dependence between pair \((k, l)\) in cluster \(i\), \( \hat{e}_{ik}(b)\hat{e}_{il}(b) \) does not have expectation \( \text{Cov}(\hat{e}_{ik}(b), \hat{e}_{il}(b)) \) unless \( \Delta_{ik} = \Delta_{il} = 1 \). Nevertheless, \( \hat{\Omega}_{kl}(b) \) is still usable for its simplicity in constructing working covariance.

Parsimonious working covariance structures such as exchangeable (EX) or autoregressive with order 1 (AR1) can be imposed. Parameters \( \alpha \) in the working covariance can be estimated with method of moment estimator \( \hat{\alpha}_n \) based on \( \hat{\Omega} \) as in the non-censored case (Liang and Zeger, 1986).

When there is no censoring, the working covariance matrix \( \hat{\Omega} \) converges to the true covariance matrix. This is no longer true when censoring is present. Nevertheless, \( \hat{\Omega} \), and consequently, \( \hat{\alpha}_n \), still converges to some limit which helps to improve the efficiency of the GEE estimation.

Extension to unequal cluster sizes as in a recurrent event setting is straightforward. In this case, it is reasonable to assume identical marginal error distributions, hence, identical marginal variances. The working covariance matrix \( \hat{\Omega}_i \) with dimension \( K_i \times K_i \) can be constructed with an given estimator \( \hat{\alpha}_n \) for \( \alpha \) for a specified working covariance structure.

Under certain regularity conditions, the proposed estimator is consistent to the true regression coefficients \( \beta_0 \) and asymptotically normal. The asymptotic results are summarized in the following theorems, whose proofs are sketched in the Appendix.

**Theorem 1.** Under conditions A1–A9 in the Appendix, \( \hat{\beta}_n^{(m)} \) is a consistent estimator of the true parameter \( \beta_0 \) for each \( m \geq 1 \).

**Theorem 2.** Under conditions A1–A9 in the Appendix, \( n^{1/2}(\hat{\beta}_n^{(m)} - \beta_0) \) converges in distribution to multivariate normal with mean zero for each \( m \geq 1 \).

The resampling approach developed by Jin et al. (2006a) is adapted to estimate the covariance matrix of \( \hat{\beta}_n^{(m)} \). Let \( Z_i, i = 1, \cdots, n \), be independent and identically distributed positive random variables, independent of the observed data, with \( E(Z_i) = \text{Var}(Z_i) = 1 \). Define

\[
\hat{Y}_{ik}^*(b) = \Delta_{ik}Y_{ik} + (1 - \Delta_{ik}) \left[ \int_{e_{ik}(b)}^{\infty} u d\hat{F}_{k,b}^*(u) \right] / \left[ 1 - \hat{F}_{k,b}^*\{e_{ik}(b)\} \right]
\]

where

\[
\hat{F}_{k,b}^*(t) = 1 - \prod_{1 \leq i \leq n, 1 \leq r \leq K; m_r = m_k, e_{ir} < t} \left( 1 - \sum_{j=1}^{n} \sum_{1 \leq l \leq K; m_l = m_k} Z_j I(e_{jl}(b) \geq e_{ir}(b)) \right).
\]

Then the multiplier resampling version of equation (5) has the following form,

\[
L_n^*(b, \alpha) = \left[ \sum_{i=1}^{n} Z_i(X_i - \bar{X})\hat{\Omega}_i^{-1}(\alpha(b))(X_i - \bar{X}) \right]^{-1} \left[ \sum_{i=1}^{n} Z_i(X_i - \bar{X})\hat{\Omega}_i^{-1}(\alpha(b)) \left( \hat{Y}_{i}^*(b) - \bar{Y}^*(b) \right) \right],
\]

where \( \alpha(b) \) is an estimator of working correlation parameter given regression coefficients evaluated at \( b \) and \( \bar{Y}^*(b) = \sum_{i=1}^{n} \hat{Y}_{i}^*(b) / n \).

For a realization of \((Z_1, \ldots, Z_n)\) and an initial estimator \( \hat{\beta}_n^{(0)} \), a bootstrap estimator of \( \beta \) is obtained from iteration \( \hat{\beta}_n^{(m)*} = L_n^*(\hat{\beta}_n^{(m-1)*}) \). The covariance matrix of \( \hat{\beta}_n^{(m)} \) can be estimated from the sample covariance matrix of a bootstrap sample of \( \hat{\beta}_n^{(m)*} \). The consistency of this variance estimator can be proved following arguments similar to those in Jin et al. (2006a).
4 Simulation Study

We conducted two simulation studies to assess the performance of proposed estimators and compared its efficiency with the initial estimators from Johnson and Strawderman (2009). The first study had a clustered failure time setting with identical regression coefficients across margins and identical marginal error distributions. The cluster sizes were fixed at three. For cluster $i$, the multivariate failure time $T_i = (T_{i1}, T_{i2}, T_{i3})$ was generated from

$$\log T_{ik} = 2 + X_{1ik} + X_{2ik} + \epsilon_{ik},$$

where $X_{1ik}$ was Bernoulli with rate 0.5, $X_{2ik}$ was $N(0, 0.5^2)$, and $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3})$ was a trivariate random vector specified by identical marginal error distributions and a copula for the dependence structure. Three marginal error distributions were considered: standard normal, standard logistic, and standard Gumbel, abbreviated by N, L, and G, respectively; the tail of the three distributions gets heavier from N to L to G. The dependence structure was specified by a Clayton copula with three levels of dependence measured by Kendall’s tau: 0, 0.3, and 0.6. Censoring times were independently generated from uniform distributions over $(0, c)$, where $c$ was selected for each margin to achieve three levels of censoring percentage: 0%, 25%, and 50%. We considered random samples of size $n = 200$ clusters. Rank-based estimator with Gehan’s weight from the induced smoothing approach of Johnson and Strawderman (2009), denoted by JS, was used as the initial estimator for GEE estimators. Two working covariance structures, EX and AR1, were used for the proposed iterative GEE procedure. The covariance matrix of the estimator was obtained from the resampling approach with 200 bootstrap size in Section 3. For each configuration, we did 1000 replicates.

The results are summarized in Table 1. To save space, only results for nonzero Kendall’s tau were reported. All estimators appear to be virtually unbiased. The empirical variation of the estimates and the estimated variation based on the resampling procedure agree closely for all estimators. For a given censoring percentage, as the dependence level increases, the variance of the JS estimator changes little, but the variance of the GEE estimators with both working covariance structures decreases. Further, the variance from the EX structure is in general smaller than that from the AR1 structure, which is expected because the true covariance structure is exchangeable in this simulation setting. For a fixed dependence level, the effect of censoring percentage on the variances of the estimator depends on the marginal error distributions. The variance increases clearly as the censoring gets heavier when the errors are normally distributed, but this pattern is not observed with Gumbel or logistic marginal error distributions. The relative efficiency of the proposed GEE estimator relative to the rank-based JS estimator is up to 3.5 in the table (with logistic margin and Kendall’s tau 0.6 for $\beta_2$).

The second simulation setting had multiple event data with different regression coefficients and different marginal error distributions. The cluster sizes were still fixed at three. For cluster $i$, the multivariate failure times were generated from

$$\log T_{ik} = \beta_{0k} + \beta_{1k}X_{1ik} + \beta_{2k}X_{2ik} + \epsilon_{ik},$$

where $(\beta_{0k}, \beta_{1k}, \beta_{2k})$, $k = 1, 2, 3$, was the regression coefficient vector for margin $k$, and $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3})$ was a trivariate random vector specified by three marginal distributions and a copula for dependence. The marginal distributions of $\epsilon_i$ were standard normal, standard logistic, and standard Gumbel, respectively, for the first, second and third margin; their copula was Clayton with three dependence levels measured by Kendall’s tau: 0, 0.3, and 0.6. The regression coefficients $(\beta_{0k}, \beta_{1k}, \beta_{2k})$ were set to be $(-1, 1, -1)$, $(1, -1, 1)$, and $(1, 1, 1)$, respectively for $k = 1, 2, 3$. 


| Marg | τ | Cens | β | Bias | Empirical SE | Estimated SE | RE |
|------|---|------|----|------|--------------|--------------|----|
| N    | 0.3 | 0% | β₁ | -0.002 | 0.087 | 0.084 | 1.492 |
|      |     |     | β₂ | 0.001 | 0.083 | 0.084 | 1.349 |
|      | 25% |     | β₁ | -0.008 | 0.091 | 0.089 | 1.543 |
|      |     |     | β₂ | -0.003 | 0.093 | 0.090 | 1.550 |
|      | 50% |     | β₁ | -0.006 | 0.101 | 0.099 | 1.467 |
|      |     |     | β₂ | -0.004 | 0.102 | 0.102 | 1.484 |
| L    | 0.3 | 0% | β₁ | 0.002 | 0.082 | 0.083 | 3.130 |
|      |     |     | β₂ | 0.005 | 0.082 | 0.084 | 3.316 |
|      | 25% |     | β₁ | -0.007 | 0.092 | 0.088 | 3.322 |
|      |     |     | β₂ | -0.003 | 0.090 | 0.090 | 2.931 |
|      | 50% |     | β₁ | -0.003 | 0.101 | 0.100 | 2.567 |
|      |     |     | β₂ | 0.000 | 0.102 | 0.102 | 2.142 |
| G    | 0.3 | 0% | β₁ | 0.006 | 0.138 | 0.142 | 2.966 |
|      |     |     | β₂ | 0.001 | 0.145 | 0.142 | 3.020 |
|      | 25% |     | β₁ | -0.020 | 0.140 | 0.145 | 3.245 |
|      |     |     | β₂ | -0.013 | 0.153 | 0.147 | 3.181 |
|      | 50% |     | β₁ | -0.011 | 0.164 | 0.162 | 3.494 |
|      |     |     | β₂ | -0.008 | 0.166 | 0.166 | 3.439 |
| 0.6  | 0% |     | β₁ | 0.000 | 0.145 | 0.141 | 2.966 |
|      |     |     | β₂ | 0.001 | 0.145 | 0.142 | 3.020 |
|      | 25% |     | β₁ | -0.011 | 0.149 | 0.146 | 3.245 |
|      |     |     | β₂ | -0.014 | 0.150 | 0.146 | 3.494 |
|      | 50% |     | β₁ | -0.009 | 0.164 | 0.162 | 3.439 |
|      |     |     | β₂ | -0.006 | 0.161 | 0.165 | 3.036 |

Table 1: Summary of simulation results with identical regression coefficients and identical marginal error distributions based on 1000 replications. Empirical SE is the standard deviation of the parameter estimates; Estimated SE is the mean of the standard error of the estimator; RE is the empirical relative efficiencies in relative to the JS estimator.
Other settings such as the covariates, censoring time, sample size, initial estimator, bootstrap sample size for variance estimation, replication size were all the same as in the first simulation setting. In addition to the JS estimator, GEE estimators with two working covariance structures were considered: EX and unstructured (UN).

The results are summarized in Tables 2. Similar to the first simulation study, all estimators are virtually unbiased, and their variance estimators are generally close to the empirical variances of the replicates. The variance of the GEE estimators decreases as the dependence gets stronger at any level of censoring percentage. Holding the dependence level, as the censoring percentage increases, the variance increases at the normal margin, but the pattern is different for the other two margins. The variance has little changes at the logistic margin. At the Gumbel margin, it remains its level as the censoring percentage increases from 0 to 25%, but increases notably as the censoring percentage increases from 25% to 50%. There is almost no difference between the two working covariance structures, both leading to about the same relative efficiency compared to the rank-based JS estimator. The relative efficiency of both GEE estimators almost double as Kendall’s tau is increased from 0.3 to 0.6.

5 Application

The diabetic retinopathy study (DRS) was started in 1971 (Diabetic Retinopathy Study Research Group, 1976) with the aim to investigate the efficacy of laser photocoagulation in delaying the onset of severe vision loss. Diabetic retinopathy is the most common and serious eye complication of diabetes, which may lead to poor vision or even blindness. A subset of the DRS data for patients with “high-risk” diabetic retinopathy, categorized by risk group 6 or higher, has been analyzed by many authors (e.g., Huster et al., 1989; Lee and Wei, 1993; Liang et al., 1993; Spiekerman and Lin, 1996). Each of the 197 patients in this subset had one eye randomized to laser treatment and the other eye received no treatment. The outcomes of interest were the actual times from initiation of treatment to the time when visual acuity dropped below 5/200 at two visits in a row (defined as “blindness”). The scientific interest was the effectiveness of the laser treatment and the influence of other risk factors. In addition to the treatment indicator, three covariates are available: age at diagnosis of diabetes, type of diabetes (1 = adult, 0 = juvenile), and risk group (6 to 12, rescaled to 0.5 to 1.0). Since the interaction between treatment and diabetes type was found to be significant in Spiekerman and Lin (1996), we also include this interaction in the model.

We first fit a bivariate AFT model with identical error margins and identical regression coefficients for both left and right eyes. The second AFT model we fit was the opposite, with different error margins and different regression coefficients for left and right eyes. For each model, we report GEE estimators with working independence and working exchangeable covariance structures, in addition to the rank-based JS estimator in Table 3. The GEE estimator with exchangeable working structure from the first model suggests that the treatment was significant in delaying the onset of vision loss; it had a significant higher effect for adult than for juvenile, and patients in higher risk groups tended to lose vision sooner. Note that the treatment effect was not significant if working independence were used in the GEE estimator. The second model offered a possibility to check whether the marginal error distributions and regression coefficients should indeed be identical as assumed in the first model. Figure 1 shows the the Kaplan–Meier survival curves of the censored residuals for the left margin and right margin respectively, overlaid with the pooled estimate from the first model. All three curves appear to be mingled together tightly. A naive log-rank test to compare the two margins, ignoring that the regression coefficients were not known but estimated,
Table 2: Summary of simulation results with different regression coefficients and different marginal error distributions based on 1000 replications. Empirical SE is the standard deviation of the parameter estimates; Estimated SE is the mean of the standard error of the estimator; RE is the empirical relative efficiencies in relative to the JS estimator.

| τ  | Cen | β  | EST JS | Empirical SE JS | Estimated SE JS | RE | EST EX | Empirical SE EX | Estimated SE EX | RE | EST UN | Empirical SE UN | Estimated SE UN |
|----|-----|----|--------|------------------|-----------------|----|--------|------------------|-----------------|----|--------|------------------|-----------------|
| 0.3| 0%  | β11| 0.000  | 0.143  | 0.146  | 1.370  | 0.003  | 0.122  | 0.140  | 1.340  | 0.003  | 0.123  | 1.351  |
|    |     | β21| −0.000 | 0.151  | 0.146  | 1.400  | −0.003 | 1.016  | 0.166  | 1.012  | −0.005 | 0.162  | 1.012  |
|    |     | β12| 0.002  | 0.164  | 0.166  | 0.990  | 0.004  | 1.004  | 0.166  | 1.000  | 0.005  | 0.166  | 1.000  |
| 25%| β11| 0.000  | 0.154  | 0.156  | 1.290  | 0.004  | 0.156  | 0.158  | 1.310  | 0.005  | 0.158  | 1.310  |
|    | β21| −0.005 | 0.160  | 0.162  | 1.340  | −0.007 | 1.340  | 0.162  | 1.340  | −0.009 | 0.162  | 1.340  |
|    | β12| −0.006 | 0.161  | 0.163  | 1.170  | −0.010 | 1.170  | 0.161  | 1.170  | −0.014 | 0.161  | 1.170  |
| 50%| β11| 0.000  | 0.170  | 0.172  | 1.470  | 0.000  | 0.172  | 0.172  | 1.470  | 0.000  | 0.172  | 1.470  |
|    | β21| −0.009 | 0.166  | 0.168  | 1.260  | −0.013 | 1.260  | 0.166  | 1.260  | −0.017 | 0.166  | 1.260  |
|    | β12| 0.002  | 0.168  | 0.170  | 1.170  | 0.004  | 1.170  | 0.168  | 1.170  | 0.006  | 1.170  | 1.170  |
| 0.6| β11| 0.000  | 0.149  | 0.151  | 2.810  | 0.000  | 0.151  | 0.151  | 2.810  | 0.000  | 0.151  | 2.810  |
|    | β21| −0.015 | 0.140  | 0.142  | 2.700  | −0.019 | 0.142  | 0.142  | 2.700  | −0.023 | 0.142  | 2.700  |
|    | β12| −0.010 | 0.167  | 0.169  | 1.750  | −0.014 | 1.750  | 0.167  | 1.750  | −0.018 | 0.167  | 1.750  |
| 25%| β11| 0.000  | 0.155  | 0.157  | 2.810  | 0.000  | 0.157  | 0.157  | 2.810  | 0.000  | 0.157  | 2.810  |
|    | β21| −0.007 | 0.155  | 0.157  | 1.750  | −0.011 | 1.750  | 0.155  | 1.750  | −0.015 | 1.750  | 1.750  |
|    | β12| 0.000  | 0.166  | 0.168  | 2.030  | 0.000  | 0.168  | 0.168  | 2.030  | 0.000  | 0.168  | 2.030  |
| 50%| β11| 0.000  | 0.174  | 0.176  | 3.020  | 0.000  | 0.176  | 0.176  | 3.020  | 0.000  | 0.176  | 3.020  |
|    | β21| −0.009 | 0.192  | 0.194  | 2.470  | −0.013 | 2.470  | 0.192  | 2.470  | −0.017 | 2.470  | 2.470  |
|    | β12| 0.017  | 0.176  | 0.178  | 1.920  | 0.021  | 1.920  | 0.176  | 1.920  | 0.025  | 1.920  | 1.920  |
|    | β13| −0.000 | 0.307  | 0.310  | 2.440  | −0.004 | 2.440  | 0.307  | 2.440  | −0.008 | 2.440  | 2.440  |
|    | β23| 0.036  | 0.322  | 0.325  | 2.471  | 0.040  | 2.471  | 0.322  | 2.471  | 0.043  | 2.471  | 2.471  |
Table 3: Results of analyzing Diabetic Retinopathy Study.

| Margin        | Effects        | JS     | IND     | EX      |
|---------------|----------------|--------|---------|---------|
|               |                | EST    | SE      | EST     | SE      | EST    | SE      |
|                |                |        |         |         |         |        |         |
| pooled risk    | age            | -0.010 | 0.012   | -0.010  | 0.013   | -0.010 | 0.014   |
| diabetes       |                | -0.140 | 0.349   | -0.065  | 0.440   | -0.065 | 0.369   |
| treatment      |                | 0.520  | 0.197   | 0.545   | 0.330   | 0.542  | 0.263   |
| interaction    |                | 1.116  | 0.301   | 0.961   | 0.466   | 0.964  | 0.410   |
| different      |                |        |         |         |         |        |         |
| left risk      | age            | -0.042 | 0.016   | -0.037  | 0.019   | -0.036 | 0.020   |
| diabetes       |                | 0.825  | 0.463   | 0.706   | 0.554   | 0.702  | 0.544   |
| treatment      |                | 0.925  | 0.422   | 0.645   | 0.549   | 0.652  | 0.489   |
| interaction    |                | 1.719  | 0.650   | 1.742   | 0.855   | 1.739  | 0.820   |
| right risk     | age            | 0.011  | 0.014   | 0.009   | 0.016   | 0.009  | 0.018   |
| diabetes       |                | -0.770 | 0.432   | -0.640  | 0.528   | -0.639 | 0.656   |
| treatment      |                | 0.383  | 0.326   | 0.481   | 0.381   | 0.477  | 0.446   |
| interaction    |                | 0.752  | 0.476   | 0.600   | 0.639   | 0.603  | 0.646   |
| partial common |                |        |         |         |         |        |         |
| left age       | diabetes       | 0.892  | 0.406   | 0.848   | 0.607   | 0.846  | 0.621   |
| diabetes       |                | -0.039 | 0.015   | -0.036  | 0.021   | -0.036 | 0.022   |
| right age      | diabetes       | 0.011  | 0.015   | 0.009   | 0.019   | 0.009  | 0.017   |
| diabetes       |                | -0.870 | 0.435   | -0.837  | 0.499   | -0.835 | 0.574   |
| common treatment | risk group   | 0.630  | 0.227   | 0.606   | 0.250   | 0.607  | 0.267   |
| interaction    |                | 1.067  | 0.318   | 1.014   | 0.344   | 1.014  | 0.409   |
|                |                |        |         |         |         |        |         |
yielded a p-value of 0.907, confirming the visual observation. Our joint model also allows hypothesis testing of equal coefficients for each covariate across the two margins with Wald-type tests. The coefficients of treatment, risk group, and treatment-diabetes interaction were found to be not significantly different across the two margins, with p-values 0.400, 0.278, and 0.147, respectively. The coefficients of age and diabetes were found to be significantly different across the two margins, with p-values 0.036 and 0.042, respectively.

We then fit an bivariate AFT model with identical error margins, same coefficients for treatment, risk group and treatment-diabetes interaction, and different coefficients for age and diabetes. This is one of the many models with intermediate complexity between the first model and the second model. Results are summarized in the last section of Table 3. This time, the shared coefficients of treatment, risk group, and treatment-diabetes interaction remained significant as before. An interesting finding is that the difference between the coefficient of diabetes (0.846 versus −0.835) is significantly nonzero with a p-value 0.002, suggesting that the adult diabetes have sooner onset of vision loss in right eye than in left eye. This finding has not been reported in existing analyses.

The second application is a colon cancer study (Lin, 1994). Through randomization, 315, 310 and 304 patients with stage C colon cancer received observation, levamisole alone (Lev), and levamisole combined with fluorouracil (Lev + 5FU), respectively. Lin (1994) considered bivariate models for the time to first recurrence and the time to death. The research interest was the effectiveness of the treatment in prolonging the time to recurrence and time to death. Gender and age are available as covariates besides treatment.

In this application, the error distributions and regression coefficients have no reason to be identical across margins. We report results with different error margin and different regression coefficients in Table 4. Since all covariates are at the cluster level, the exchangeable and independent working covariance structure give the same results (e.g., Hin et al., 2007). The Kaplan–Meier survival curves for the two error margins are shown in Figure 1, which clearly exhibits no similarity; a naive log-rank test gives p-value 0.0008. The treatment of levamisole combined with fluorouracil
appears to have a significant positive effect on both event times. The gender and age are found not
to be significant for either time. The estimated difference between the combined treatment effect
on recurrence and on death (0.931 versus 0.307) has a standard error 0.103, suggesting that the
combined treatment has a higher effect on recurrence than on death.

6 Discussion

The working covariance structure of the proposed GEE approach is different from that in a gen-
eralized linear model setting, where the variance is assumed to be a function of the mean. The
errors at each margin are assumed to be independent and identically distributed, and hence have
the same variance. This assumption may be relaxed by imposing a structure on the variance of
the errors. For instance, in model (1), we replace $\epsilon_{ik}$ with $\sigma_{ik}\nu_{ik}$, where $\nu_{ik}$’s are independent and
identically distributed for $i = 1, \ldots, n$ with mean zero and variance one, and the scale $\sigma_{ik}$ may be
described by a regression model with covariates. Such specification leads to heteroskedasticity in
errors and merits further investigation.

For applications like the DRS study, where there are reasons to impose identical distribution
across margins, a rigorous test to compare the survival curves of the residuals would be desirable.
We used naive tests ignoring the fact that the residuals were calculated based on estimated regres-
sion coefficients. A rigorous test procedure should take into account of the variation caused by the
estimation procedure.

A Sketch of the Proofs

We impose the following regularity conditions:

A1: $\|X_i\| \leq B$ for all $i = 1, \ldots, n$ and some nonrandom constant $B$, where $\| \cdot \|$ is matrix norm.

A2: The density function of $F_{k,\beta}$ exists such that $\int_{-\infty}^{\infty} t^2 dF_{k,\beta}(t) < \infty$, for $k = 1, \ldots, K$. 
A3: The distribution function $F_{k,\beta}$ is twice differentiable with density $f_{k,\beta}$ such that

$$\int_{-\infty}^{\infty} \left( \frac{f'_{k,\beta}(t)}{f_{k,\beta}(t)} \right)^2 dF_{k,\beta}(t) < \infty$$

where $1 \leq k \leq K$, and both $f_{k,\beta}(t)$ and $f'_{k,\beta}(t)$ are bounded functions.

A4: $E[\exp(\theta c_{ik})] + \sup_{k \in \{1, \ldots, K\}} E[\exp(\theta c_{ik}^{-1})] < \infty$ for some $\theta > 0$, where $a^+ = |a|I_{a > 0}$.

A5: $\sup_{|b| < \infty; -\infty < t < \infty} \sum_{i=1}^{\infty} \sum_{k=1}^{K} \Pr(t \leq C_{ik} - X^\top_{ik} b \leq t + h) = O(nh)$ as $h \to 0$ and $nh \to \infty$.

A6: As $n \to \infty$, $\hat{\alpha}_n$ is bounded and is $n^{1/2}$ consistent to $\alpha_0$ given $\beta$.

A7: As $n \to \infty$, initial estimator $b_n$ is $n^{1/2}$ consistent to $\beta_0$ and $\sqrt{n}(b_n - \beta_0)$ is asymptotically normal with zero mean.

A8: The slope matrices $n^{-1}\partial U_n/\partial \beta$ and $n^{-1}\partial U_n/\partial b$ evaluated at $(\beta_0, \beta_0, \alpha_0)$ converge to nondegenerate, finite limit $A$ and $B$, respectively.

A9: The derivative $\partial \Omega^{-1}_i(\alpha)/\partial \alpha$ is finite for all $i = 1, 2, \ldots, n$.

Conditions A1–A5 are standard and ensure the existence of the solution of equation (2) (Lai and Ying, 1991). It is natural to assume that the working covariance matrix $\Omega$ in equation (4) is a symmetric positive definite matrix. Then there exist a $K \times K$ nonsingular matrix, $\Gamma$, such that $\Omega(\alpha_0) = \Gamma^{1/2}\Gamma^{-1/2}$. Let $X_i = \Gamma^{-1/2}X_i$, $T_i = \Gamma^{-1/2}Y_i$, $C_i = \Gamma^{-1/2}C_i$, and $\omega_i = \Gamma^{-1/2}e_i$. Then equation (4) evaluated at $\alpha = \alpha_0$ can be viewed as equation (2) with the transformed data $X_i$ and $Y_i = \min(Y_i, C_i)$, with error $\omega_i, i = 1, \ldots, n$. The existence of the solution to equation (4) can be verified by the same arguments as in Lai and Ying (1991), with assumptions similar to A1 to A5 on the transformed data. The consistency and asymptotic normality of the estimator given $\alpha = \alpha_0$ follow from the same arguments as in Jin et al. (2006a).

The extra complexity here comes from the fact that equation (4) is solved at $\alpha = \hat{\alpha}_n$, an estimator of $\alpha_0$. Under condition A9, the $i$th term in the summation of $\partial U_n/\partial \alpha$ evaluated at $(\beta_0, \beta_0, \alpha_0)$ is a linear function of $Y_i(\beta_0) - X^\top_{ik} \beta_0, i = 1, \ldots, n$, with expectation zero. By the law of large number, $n^{-1}\partial U_n/\partial \alpha$ evaluated at $(\beta_0, \beta_0, \alpha_0)$ converges to zero in probability.

### A.1 Proof of Theorem 1

At the solution $\hat{\beta}_n^{(1)}$ given $b_n$ and $\hat{\alpha}_n$, we have $n^{-1}U_n(\hat{\beta}_n^{(1)}, b_n, \hat{\alpha}_n) = 0$. Taylor expansion at $(\beta_0, \beta_0, \alpha_0)$ gives

$$0 = \frac{1}{n} U_n(\beta_0, \beta_0, \alpha_0) + \frac{1}{n} \frac{\partial}{\partial \beta} [U_n(\beta_0, \beta_0, \alpha_0)] (\hat{\beta}_n^{(1)} - \beta_0)$$

$$+ \frac{1}{n} \frac{\partial}{\partial b} [U_n(\beta_0, \beta_0, \alpha_0)] (b_n - \beta_0) + \frac{1}{n} \frac{\partial}{\partial \alpha} [U_n(\beta_0, \beta_0, \alpha_0)] (\hat{\alpha}_n - \alpha_0) + o_p(n^{-1/2})$$

$$= \frac{1}{n} U_n(\beta_0, \beta_0, \alpha_0) + A_n(\hat{\beta}_n^{(1)} - \beta_0) + B_n(b_n - \beta_0) + C_n(\hat{\alpha}_n - \alpha_0) + o_p(n^{-1/2}). \quad (10)$$

With regularity conditions A1–A5, the first term converges in probability to zero by the law of large number. The convergence of $b_n$ and $\alpha_n$ in A6 and A7, combined with the limit condition in A8 and A9, then gives consistency of $\hat{\beta}_n^{(m)}$ to $\beta_0$. By induction, $\hat{\beta}_n^{(m)}$ is consistent for $\beta_0$ at every $m$. 
A.2 Proof of Theorem 2

Under regularity conditions \( \sqrt{n}(\beta_n^{(1)} - \beta_0) \) can be expressed as

\[
\sqrt{n}(\beta_n^{(1)} - \beta_0) = [A_n]^{-1} \left[ \frac{1}{\sqrt{n}} U_n(\beta_0, \beta_0, \alpha_0) + B_n \sqrt{n}(b_n - \beta_0) + C_n \sqrt{n}(\alpha_n - \alpha_0) \right] + o_p(1). \tag{11}
\]

With condition A9, \( C_n \) converges to zero in probability, and, hence, with \( \sqrt{n} \) consistency of \( \hat{\alpha}_n \),

\[
C_n \sqrt{n}(\hat{\alpha}_n - \alpha_0) = o_p(1).
\]

Equation (11) is then asymptotically equivalent to

\[
[A_n]^{-1} \left[ \frac{1}{\sqrt{n}} U_n(\beta_0, \beta_0, \alpha_0) + B_n \sqrt{n}(b_n - \beta_0) \right].
\]

With the assumption that \( b_n - \beta_0 \) is asymptotically normal, there exist some nonrandom functions \( \eta_i \) with zero mean such that,

\[
\sqrt{n}(b_n - \beta_0) = n^{-1/2} \sum_{i=1}^{n} \eta_i + o_p(\|b_n - \beta_0\|).
\]

On the other hand, \( U_n(\beta_0, \beta_0, \alpha_0) \) is a sum of independent and identically distributed quantities with zero mean, denoted by \( \phi_i \)'s, \( i = 1, \ldots, n \). Equation (11) reduces to

\[
\sqrt{n}(\beta_n^{(1)} - \beta_0) = [A_n]^{-1} \left[ n^{-1/2} \sum_{i=1}^{n} (\phi_i + B_n \eta_i) \right] + o_p(\|b_n - \beta_0\|).
\]

By multivariate central limit theorem for sums of independent random vectors, the asymptotic distribution for \( \beta_n^{(1)} \) is zero mean multivariate normal as \( n \to \infty \). The limit covariance matrix \( \Sigma \) have the form \( A^{-1} \Phi A^{-1} \), where \( \Phi = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \nu_i \nu_i^\top \) with \( \nu_i = \phi_i + B \eta_i \). Induction then implies that \( \beta_n^{(m)} \) is multivariate normal for every \( m \).

References

Brown, B. M. and Wang, Y.-G. (2005). Standard errors and covariance matrices for smoothed rank estimators. *Biometrika* 92, 149–158.

Brown, B. M. and Wang, Y.-G. (2007). Induced smoothing for rank regression with censored survival times. *Statistics in Medicine* 26, 828–836.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* 66, 429–436.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B, Methodological* 34, 187–220.

Diabetic Retinopathy Study Research Group (1976). Preliminary report on effects of photocoagulation therapy. *American Journal of Ophthalmology* 81, 383–396.

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* 52, 203–223.
Hin, L.-Y., Carey, V. J., and Wang, Y.-G. (2007). Criteria for working correlation structure selection in GEE. *The American Statistician* **61**, 360–364. doi:10.1198/000313007X245122.

Hornsteiner, U. and Hamerle, A. (1996). A combined GEE/Buckley-James method for estimating an accelerated failure time model of multivariate failure times. Discussion Paper 47, Ludwig-Maximilians-Universität München, Collaborative Research Center 386.

Huang, Y. (2002). Calibration regression of censored lifetime medical cost. *Journal of the American Statistical Association* **97**, 318–327.

Huster, W. J., Brookmeyer, R., and Self, S. G. (1989). Modelling paired survival data with covariates. *Biometrics* **45**, 145–156.

Jin, Z., Lin, D. Y., Wei, L. J., and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–353.

Jin, Z., Lin, D. Y., and Ying, Z. (2006a). On least-squares regression with censored data. *Biometrika* **93**, 147–161.

Jin, Z., Lin, D. Y., and Ying, Z. (2006b). Rank regression analysis of multivariate failure time data based on marginal linear models. *Scandinavian Journal of Statistics* **33**, 1–23.

Johnson, L. M. and Strawderman, R. L. (2009). Induced smoothing for the semiparametric accelerated failure time model: Asymptotics and extensions to clustered data. *Biometrika* **96**, 577–590.

Lai, T. L. and Ying, Z. (1991). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *The Annals of Statistics* **19**, 1370–1402.

Lee, E. W. and Wei, Z., L. J. aand Ying (1993). Linear regression analysis for highly stratified failure time data. *Journal of the American Statistical Association* **88**, 557–565.

Li, H. and Yin, G. (2009). Generalized method of moments estimation for linear regression with clustered failure time data. *Biometrika* **96**, 293–306.

Liang, K.-Y., Self, S. G., and Chang, Y.-C. (1993). Modelling marginal hazards in multivariate failure time data. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* **55**, 441–453.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine* **13**, 2233–2247.

Prentice, R. L. (1978). Linear rank tests with right censored data (Corr: V70 p304). *Biometrika* **65**, 167–180.

Qu, A., Lindsay, B. G., and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–836.
Ritov, Y. (1990). Estimation in a linear regression model with censored data. *The Annals of Statistics* **18**, 303–328.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In Jewell, N., Dietz, K., and Farewell, V. (editors), *AIDS Epidemiology — Methodological Issues*, pages 297–331. Boston, MA: Birkhäuser.

Spiekerman, C. F. and Lin, D. Y. (1996). Checking the marginal Cox model for correlated failure time data. *Biometrika* **83**, 143–156.

Strawderman, R. L. (2005). The accelerated gap times model. *Biometrika* **92**, 647–666.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis* **45**, 89–103.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian Journal of Statistics* **23**, 461–471.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics* **18**, 354–372.

Wang, Y.-G. and Fu, L. (2011). Rank regression for accelerated failure time model with clustered and censored data. *Computational Statistics and Data Analysis* **55**, 2334–2343.

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

Zhou, M. (1992). *M*-estimation in censored linear models. *Biometrika* **79**, 837–841.

Department of Statistics, University of Connecticut, 215 Glenbrook Rd. U-4120, Storrs, CT 06269, U.S.A.

E-mail: (steven.chiou@uconn.edu and jun.yan@uconn.edu)

Division of Biostatistics, School of Public Health, University of Minnesota, A460 Mayo Building, MMC 303, 420 Delaware St., S.E. Minneapolis, MN 55455

E-mail: (junghikim0@gmail.com)

Institute for Public Health Research, University of Connecticut Health Center, 99 Ash Street, 2nd Floor, MC 7160, East Hartford, CT 06108

E-mail: (jun.yan@uconn.edu)