Supplementary material for
“Maximum likelihood estimation for semiparametric regression models with multivariate interval-censored data”

BY DONGLIN ZENG, FEI GAO AND D. Y. LIN

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7420, USA
dzeng@bios.unc.edu  fgao@live.unc.edu  lin@bios.unc.edu

S.1. INCREASE OF THE LIKELIHOOD OVER EM ITERATIONS

We wish to show that the likelihood increases at each iteration of the EM algorithm. Let $O$ denote the random variables $(W_{ijk}, \xi_{ijk}, b_i) (i = 1, \ldots, n; j = 1, \ldots, J_i; k = 1, \ldots, K; q = 1, \ldots, m_k)$ with $t_{kj} < R_{ijk}^*$, and let $\chi$ denote the parameters $\theta$ and $\lambda_{kq}$’s. Clearly, the joint density of $O$ conditional on $(X_{ijk}, Z_{ijk}) (i = 1, \ldots, n; j = 1, \ldots, J_i; k = 1, \ldots, K)$, denoted by $f(O; \chi)$, with respect to a dominating measure is the exponential of expression (5).

Let $D$ consist of the maximum monitoring time for each subject such that the sum of $W_{ijkq}$ over $t_{kj}$ less than or equal to the monitoring time remains zero. Clearly, observing $(A_{ijk} = 0, B_{ijk} > 0 : i = 1, \ldots, n; j = 1, \ldots, J_i; k = 1, \ldots, K)$ is equivalent to observing $D = (L_{ijk} : i = 1, \ldots, n; j = 1, \ldots, J_i; k = 1, \ldots, K)$, so the likelihood for $D$ conditional on the covariates and monitoring times is the same as $L_n(\theta, A)$. Let $\chi^{(l)}$ denote the estimate of $\chi$ at the $l$th iteration in the EM algorithm. By definition, $\chi^{(l+1)}$ maximizes $E^{(l)} \{ \log f(O; \chi)/|D| \}$, where $E^{(l)}$ is the expectation under the conditional density with parameter value $\chi^{(l)}$. Because $E^{(l)} \{ \log f(O; \chi)/|D| \}$ is strictly concave in $\chi$ and $\chi^{(l+1)}$ is obtained by the one-step Newton-Raphson method, we see that $E^{(l)} \{ \log f(O; \chi^{(l+1)})/|D| \} \geq E^{(l)} \{ \log f(O; \chi^{(l)})/|D| \}$. Thus,

$$E^{(l)} \{ \log f(O|D, \chi^{(l+1)})/|D| \} + \log P^{(l+1)}(D) \geq E^{(l)} \{ \log f(O|D, \chi^{(l)})/|D| \} + \log P^{(l)}(D),$$

where $f(O|D, \chi^{(l)})$ is the conditional density of $O$ given $D$ with parameter value $\chi^{(l)}$, and $P^{(l)}$ pertains to $L_n(\theta, A)$ with parameter value $\chi^{(l)}$. Since

$$E^{(l)} \{ \log f(O|D, \chi^{(l)})/|D| \} - E^{(l)} \{ \log f(O|D, \chi^{(l+1)})/|D| \}$$

is the Kullback-Leibler distance between $f(O|D, \chi^{(l)})$ and $f(O|D, \chi^{(l+1)})$, which is non-negative, we conclude that $\log P^{(l+1)}(D) \geq \log P^{(l)}(D)$. That is, the likelihood is non-decreasing after each iteration of the EM algorithm.

S.2. ADDITIONAL SIMULATION STUDIES

S.2.1. Clustered Multiple-Events Data

To further evaluate the performance of the proposed methods, we conducted a series of simulation studies with clustered multiple-events data by combining the simulation schemes for clustered data and multiple events described in §5. Specifically, we considered model (1) with $K = 2$, $A_1(t) = \log(1 + 0.5t)$, $A_2(t) = 0.5t$, and $J_i$ being 1, 2 and 3 with probabilities 0.2, 0.7 and 0.1, respectively. We set $b_i = (b_{i1}, b_{21}, \ldots, b_{2J_i})^T$, where $b_{i1}$ is the cluster-specific random effect, and $b_{2j}$ ($j = 1, \ldots, J_i$) are subject-specific random effects. We chose $Z_{ijk}$ such
that $b_1^T Z_{ijk} = b_{i1} + b_{i2j}$, which is the sum of the cluster-specific random effect and the subject-specific random effect. We generated $b_{i1}$ and $b_{i2j}$ independently from $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$, respectively, where $\sigma_1^2 = 0.5$ and $\sigma_2^2 = 0.8$. In this case,

$$\Sigma_i(\gamma) = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_2^2 \end{bmatrix},$$

which is a $(J_i + 1) \times (J_i + 1)$ matrix, and $\gamma = (\sigma_1^2, \sigma_2^2)^T$. We generated two independent regression parameters for the first event cluster-level covariates, the first being $\text{Ber}(0.5)$ and the second being $\text{Un}(0,1)$. We set the regression parameters for the first event $(\beta_{11}, \beta_{12})$ to $(0.5, -0.5)$ and those of the second event $(\beta_{21}, \beta_{22})$ to $(0.4, 0.2)$. We adopted the class of logarithmic transformations indexed by parameter $r_k$ ($k = 1, 2$). We generated five potential examination times for each subject, the first being $\text{Un}(0,1)$, and the gap between any two successive examination times being $0.1 + \text{Un}(0,1)$. We assumed that the study ended at time 5, beyond which no examinations occurred. As shown in Table S.2, the proposed methods continue to perform well.

### S.2.2. Comparisons with Chen et al. (2009)

We used the simulation code kindly provided by Dr. M. H. Chen. Specifically, we generated two failure times $T_1$ and $T_2$ with hazard functions $0.04 t_1 e^{\beta X + b_1}$ and $0.02 t_2 e^{\beta X + b_2}$, respectively, where $X$ is $\text{Ber}(0.5)$, and $b_1$ and $b_2$ are $N(0, 0.01)$. We generated a sequence of potential examination times $s_j$ ($j = 1, \ldots, J$), where $J = 10$, and $s_j$ was chosen as the average of the $j/J$ quantiles of the distribution functions for the two failure times. We let the actual examination time be uniform over $\{s_1, \ldots, s_J\}$. We considered $n = 100$ or 200 and $\beta = -0.25$, 0 or 0.25 and generated 10,000 replicates for each combination.

We evaluated the proposed method and the parametric method of Chen et al. (2009). For the latter, we used Dr. Chen’s code, which sets the initial values in the algorithm to the true parameter values. Both algorithms converged in all replicates, and the results are summarized in Table S.3. Both the proposed and Chen et al. (2009)’s estimators are slightly biased when $n$ is small but the bias decreases as $n$ increases. There are some outliers in Chen et al. (2009)’s parameter estimates, the absolute values being greater than 1 in approximately 0.3% of the replicates. As a result, the standard error of Chen et al. (2009)’s estimator is considerably larger than that of the proposed estimator. If we use instead the robust standard error based on median absolute deviation, then Chen et al. (2009)’s estimator appears more efficient than the proposed estimator, as it should be since it is a parametric estimator.

In practice, the true parameter values are unknown. Thus, we also evaluated Chen et al. (2009)’s algorithm with our initial values. We found that the algorithm failed to converge within 1,000 iterations or produced zero standard error estimates in approximately 20% of the replicates.

### S.2.3. Comparisons with Chen et al. (2014)

We used the simulation code provided by Dr. M. H. Chen. Specifically, we generated two failure times $T_1$ and $T_2$ with cumulative hazard functions $0.1 t_1 e^{\beta_1 X + b_1}$ and $0.2 t_2 e^{\beta_2 X + b_2}$, respectively, where $X$ is $\text{Ber}(0.5)$, and $b_1$ and $b_2$ are $N(0, 0.16)$ with correlation 0.25. We generated a sequence of potential examination times $s_j$ ($j = 1, \ldots, J$), where $J = 5$, and $s_j$ was chosen as the average of the $j/J$ quantiles of the distribution functions for the two failure times. At each potential examination time, each subject was observed with probability 0.8. We considered...
The class of functions \{f \in \mathcal{M} : \text{Chen et al. (2014)’s algorithm converged to non-NA estimates}\}.

Chen et al. (2014)’s algorithm converged in all replicates, whereas the algorithm of Chen et al. (2014) sometimes converged to NA values. Table S.4 summarizes the simulation results based on the replicates in which Chen et al. (2014)’s algorithm converged to non-NA estimates. Chen et al. (2014)’s estimator tends to have slightly smaller standard error than the proposed estimator. This is not surprising since the former assumes that the examinations occur at a common set of time points for all study subjects and thus only needs to estimate the survival probabilities at those fixed time points. Chen et al. (2014)’s estimator tends to have larger bias than the proposed estimator. The slight bias by the proposed estimator is likely due to the fact that the simulated dependence structure is not the same as the one assumed by the model. We also examined Chen et al. (2014)’s algorithm with our initial values and found that approximately 40% of the replicates have NA estimates.

S.3. SOME USEFUL LEMMAS

**Lemma 1.** Under Conditions 1–6, the class \( \mathcal{M} \) is Glivenko-Cantelli.

**Proof.** For any \( \beta_1, \beta_2 \in \mathcal{B}, A_1, A_2 \in \mathcal{L}, k \in \{1, \ldots, K\} \) and \( t \in [0, \tau_k] \),

\[
Q_{jk}(t, b; \beta_1, A_{1k}) - Q_{jk}(t, b; \beta_2, A_{2k}) = \exp \left[ -G_k \left\{ \Lambda_k(\tau_k)F_{1k}(t, X_{jk}, Z_{jk}, b) \right\} \right] - \exp \left[ -G_k \left\{ \Lambda_k(\tau_k)F_{2k}(t, X_{jk}, Z_{jk}, b) \right\} \right],
\]

where

\[
F_{lk}(t, X_{jk}, Z_{jk}, b) = \frac{t^\beta e_{\tau}^{\gamma X_{jk}(s) + b^T Z_{jk}(s)}d\Lambda_k(s)}{\Lambda_k(\tau_k)}, \quad l = 1, 2.
\]

The class of functions \( \{e^{\beta^T X_{jk}(s) + b^T Z_{jk}(s)} : \beta \in \mathcal{B}\} \), with \( X_{jk}, Z_{jk} \) and \( b \) as random variables, is a VC class with VC-index \( V \). Thus, \( \mathcal{F}_k = \{F_{lk}(t, X_{jk}, Z_{jk}, b) : \beta \in \mathcal{B}, A_l \in \mathcal{L}\} \) is a convex hull of the VC class, with the \( L_2(\mathcal{P}) \)-bracketing number \( O \left( \exp \left( \epsilon^{-2V/(V+2)} \right) \right) \).

For any constant \( M_\Lambda \), if \( \Lambda_{1k}(\tau_k) > M_\Lambda \) and \( \Lambda_{2k}(\tau_k) > M_\Lambda \), then

\[
|Q_{jk}(t, b; \beta_1, A_{1k}) - Q_{jk}(t, b; \beta_2, A_{2k})| \leq \exp \left[ -G_k \left\{ \Lambda_k(\tau_k)F_{1k}(t, X_{jk}, Z_{jk}, b) \right\} \right] + \exp \left[ -G_k \left\{ \Lambda_k(\tau_k)F_{2k}(t, X_{jk}, Z_{jk}, b) \right\} \right] - 2 \exp \left\{ -G_k \left( M_\Lambda e^{-\tilde{M}/\|b\|} \right) \right\}.
\]

If \( \Lambda_{1k}(\tau_k) \leq M_\Lambda \) and \( \Lambda_{2k}(\tau_k) \leq M_\Lambda \), then

\[
|Q_{jk}(t, b; \beta_1, A_{1k}) - Q_{jk}(t, b; \beta_2, A_{2k})| \leq \max_{F \in \mathcal{F}_k, A \in \mathcal{L}, \Lambda_k(\tau_k) \leq M_\Lambda} \left( \exp \left[ -G_k \left\{ \Lambda_k(\tau_k)F(t, X_{jk}, Z_{jk}, b) \right\} \right] G_k \left\{ \Lambda_k(\tau_k)F(t, X_{jk}, Z_{jk}, b) \right\} \right) \times \left\{ |F_{1k}(t, X_{jk}, Z_{jk}, b) - F_{2k}(t, X_{jk}, Z_{jk}, b)| M_\Lambda + F_{2k}(t, X_{jk}, Z_{jk}, b) |\Lambda_k(\tau_k) - \Lambda_{2k}(\tau_k)| \right\} \times \left( |M_\Lambda | F_{1k}(t, X_{jk}, Z_{jk}, b) - F_{2k}(t, X_{jk}, Z_{jk}, b)| + e^{\tilde{M}/\|b\|} |\Lambda_k(\tau_k) - \Lambda_{2k}(\tau_k)| \right) \right\}.
\]
In the remaining scenario, we assume, without loss of generality, that $\Lambda_{1k}(\tau_k) \leq M_A$ and $\Lambda_{2k}(\tau_k) > M_A$. Then

\[
|Q_{jk}(t, b; \beta_1, \Lambda_{1k}) - Q_{jk}(t, b; \beta_2, \Lambda_{2k})| \\
\leq |\exp[-G_k \{\Lambda_{1k}(\tau_k)F_k(t, X_{jk}, Z_{jk}, b)\}] - \exp[-G_k \{M_A F_{1k}(t, X_{jk}, Z_{jk}, b)\}]| \\
+ |\exp[-G_k \{M_A F_{1k}(t, X_{jk}, Z_{jk}, b)\}] - \exp[-G_k \{\Lambda_{2k}(\tau_k)F_{2k}(t, X_{jk}, Z_{jk}, b)\}]| \\
\leq_{F \in F_k, A \in L, \Lambda_{1k}(\tau_k) \leq M_A} \exp[-G_k \{\Lambda_{1k}(\tau_k)F(t, X_{jk}, Z_{jk}, b)\}] G'_k \{\Lambda_{1k}(\tau_k)F(t, X_{jk}, Z_{jk}, b)\} \\
\times \left\{ e^{\tilde{M} + \tilde{M}\|b\|}|\Lambda_{1k}(\tau_k) - M_A| + 2 \exp \left\{ -G_k \left( M_A e^{\tilde{M} - \tilde{M}\|b\|} \right) \right\} \right\}.
\]

Because there exist $M_A/\epsilon$ $\epsilon$-brackets to cover $[0, M_A]$, the above results imply that there exists $O \left\{ \exp \left( e^{-2V/(V+2)} \right) \right\} \times M_A/\epsilon$ brackets

\[
\{F_{1k}(t, X_{jk}, Z_{jk}, b), F_{2k}(t, X_{jk}, Z_{jk}, b)\} \times \{\Lambda_{1k}(\tau_k), \Lambda_{2k}(\tau_k)\}
\]

such that

\[
|Q_{jk}(t, b; \beta_1, \Lambda_{1k}) - Q_{jk}(t, b; \beta_2, \Lambda_{2k})| \\
\leq_{F \in F_k, A \in L, \Lambda_{1k}(\tau_k) \leq M_A} \max \left\{ \exp[-G_k \{F(t, X_{jk}, Z_{jk}, b)\Lambda_{k}(\tau_k)\}] G'_k \{F(t, X_{jk}, Z_{jk}, b)\Lambda_{k}(\tau_k)\} \right\} \\
\times \left\{ M_A e^{\tilde{M} + \tilde{M}\|b\|} M_A \right\} \epsilon + 2 \exp \left\{ -G_k \left( M_A e^{-\tilde{M} - \tilde{M}\|b\|} M_A \right) \right\}.
\]

By Condition 6,

\[
\max_{F \in F_k, A \in L, \Lambda_{1k}(\tau_k) \leq M_A} \left\{ \exp[-G_k \{F(t, X_{jk}, Z_{jk}, b)\Lambda_{k}(\tau_k)\}] G'_k \{F(t, X_{jk}, Z_{jk}, b)\Lambda_{k}(\tau_k)\} \right\} = O_P(1).
\]

It also follows from Condition 6 that

\[
|Q_{jk}(t, b; \beta_1, \Lambda_{1k}) - Q_{jk}(t, b; \beta_2, \Lambda_{2k})| \leq O_P \left( M_A + e^{\tilde{M} + \tilde{M}\|b\|} \right) \epsilon + 2O_P(M_A^{-1/r_{x0}}).
\]

We choose $M_A = e^{-r_{x0}/(r_{x0}+1)}$ such that

\[
|Q_{jk}(t, b; \beta_1, \Lambda_{1k}) - Q_{jk}(t, b; \beta_2, \Lambda_{2k})| \leq O_P \left( e^{1/(r_{x0}+1)} + e \right) (1 + e^{\tilde{M} + \tilde{M}\|b\|}).
\]

We redefine $\epsilon$ as $e^{r_{x0}+1}$ such that there exist $O \left\{ e^{-2(r_{x0}+1)} \exp \left( e^{-2V(r_{x0}+1)/(V+2)} \right) \right\} \epsilon$-brackets to cover $Q_{jk} = \{Q_{jk}(t, b; \beta, \Lambda_k) : \beta \in B, A \in L \}$ in $L_2(\mathbb{P})$. Therefore, the class $Q_{jk}$ is Glivenko-Cantelli for any $j = 1, \ldots, J, k = 1, \ldots, K$, implying that $\mathcal{M}$ is Glivenko-Cantelli.

**Lemma 2.** Under Conditions 1–6, the class $\mathcal{M}^* = \{m(\theta, A) : \theta \in \Theta, A \in \mathcal{L}^*\}$ is Donsker, where $\mathcal{L}^* = \{A : A \in \mathcal{L}, \max_{1 \leq k \leq K} \Lambda_k(\tau_k) \leq M_A\}$, and $M_A$ is a finite constant.

**Proof.** For any $\beta_1, \beta_2 \in B, \Lambda_1, \Lambda_2 \in \mathcal{L}^*, k \in \{1, \ldots, K\}$ and $t \in [0, \tau_k],$

\[
Q_{jk}(t, b; \beta_1, \Lambda_{1k}) - Q_{jk}(t, b; \beta_2, \Lambda_{2k}) \\
\leq O_P(1) \left\{ M_A |F_{1k}(t, X_{jk}, Z_{jk}, b) - F_{2k}(t, X_{jk}, Z_{jk}, b)| + e^{\tilde{M} + \tilde{M}\|b\|}|\Lambda_{1k}(\tau_k) - \Lambda_{2k}(\tau_k)| \right\}.
\]

Thus, there exist $O \left\{ \exp \left( e^{-2V/(V+2)} \right) \right\} \times M_A/\epsilon$ $\epsilon$-brackets

\[
\{F_{1k}(t, X_{jk}, Z_{jk}, b), F_{2k}(t, X_{jk}, Z_{jk}, b)\} \times \{\Lambda_{1k}(\tau_k), \Lambda_{2k}(\tau_k)\},
\]

such that

\[
|Q_{jk}(t, b; \beta_1, \Lambda_{1k}) - Q_{jk}(t, b; \beta_2, \Lambda_{2k})| \leq O \left( M_A + e^{\tilde{M} + \tilde{M}\|b\|} \right) \epsilon.
\]
Therefore,

\[ \|Q_{jk}(t; b; \beta_1, \Lambda_{1k}) - Q_{jk}(t; b; \beta_2, \Lambda_{2k})\|_{L_2(\mathbb{P})} \leq O(\epsilon). \]

That is, the bracket number for \( M^* \) is

\[ N_{[]} (\epsilon, M^*, L_2(\mathbb{P})) = O \left\{ \exp \left( \epsilon^{-2V/(V+2)} \right) \right\} \times M_A / \epsilon. \]

The bracketing integral is finite, so \( M^* \) is Donsker.

**Lemma 3.** Under Conditions 1–7,

\[
E \left[ \sum_{j=1}^J \sum_{k=1}^K \sum_{l=0}^{M_{jk}} \left( \hat{\Lambda}_k(U_{jkl}) - \Lambda_{0k}(U_{jkl}) \right)^2 \right] = O_P(n^{-2/3}) + O \left( \|\hat{\gamma} - \gamma_0\|^2 + \|\hat{\beta} - \beta_0\|^2 \right).
\]

**Proof.** By Theorem 1, \( \hat{A} \) is consistent for \( A_0 \). Thus, there exists a finite constant \( M_A \) such that \( \hat{\Lambda}_k(r_k) \leq M_A \) for \( k = 1, \ldots, K \). It follows that the function \( m(\hat{\theta}, \hat{A}) \) belongs to the Donsker class \( M^* \) given in Lemma 2. Define

\[
\psi(\delta) = \int_0^\delta \sqrt{1 + \log N_{[]} (\epsilon, M^*, L_2(\mathbb{P}))} \, d\epsilon,
\]

which is less than \( O(\delta^{1/2}) \). By Lemma 1.3 of van de Geer (2000) and the mean-value theorem,

\[
\mathbb{P} \left\{ m(\hat{\theta}, \hat{A}) - m(\theta_0, A_0) \right\} \leq -H^2 \left\{ (\hat{\theta}, \hat{A}), (\theta_0, A_0) \right\},
\]

where \( A \lesssim B \) means that \( A \leq cB \) for a positive constant \( c \), and \( H \{(\theta, A), (\theta_0, A_0)\} \) is the Hellinger distance

\[
H \{(\theta, A), (\theta_0, A_0)\} = \left[ \int \left\{ L(\theta, A)^{1/2} - L(\theta_0, A_0)^{1/2} \right\}^2 \, d\mu \right]^{1/2},
\]

with respect to the dominating measure \( \mu \). By Theorem 3.4.1 of van der Vaart & Wellner (1996), there exists an \( r_n \) with \( r_n^2 \psi(1/r_n) \lesssim n^{1/2} \) such that \( H \{(\hat{\theta}, \hat{A}), (\theta_0, A_0)\} = O_P(1/r_n) \). In particular, we choose \( r_n \) in the order of \( n^{1/3} \) such that \( H \{(\hat{\theta}, \hat{A}), (\theta_0, A_0)\} = O_P(n^{-1/3}). \)

By the mean-value theorem,

\[
E \left( \left[ \int_{b}^{J} \prod_{j=1}^{J} \prod_{k=1}^{K} \left\{ D_{jk}(U_{jkl}) \right\} \phi(b; \Sigma(\hat{\gamma})) \, db - \int_{b}^{J} \prod_{j=1}^{J} \prod_{k=1}^{K} \left\{ D_{jk}(U_{jkl}) \right\} \phi(b; \Sigma_0) \right\} \, db \right)^2 \leq O_P(n^{-2/3}).
\]
Applying the mean-value theorem again, we have
\[
O_P(n^{-2/3}) + O(1) \| \gamma - \gamma_0 \|^2 + O(1) \| \beta - \beta_0 \|^2
\]
\[
\geq E \left[ \left( \int_{b} \prod_{j=1}^{J} \prod_{k=1}^{K} \{ D_{jk}(U_{jk}, b; \beta_0, \Lambda_k) \} \prod_{j=1}^{J} \prod_{k=1}^{K} \{ D_{jk}(U_{jk}, b; \beta_0, \Lambda_{0k}) \} \right)^2 \phi(b; \Sigma_0) db \right]^{1/2}
\]
\[
\geq c_0 E \left[ \left( \sum_{j=1}^{J} \sum_{k=1}^{K} \int_{b} \prod_{j'=1}^{J} \prod_{k'=1, k' \neq k}^{K} D_{j'k'}(U_{j'k'}, b; \beta_0, \Lambda_{0k'}) \right) \sum_{l=0}^{M_{jk}} \Delta_{jkl} \int_{0}^{T_k} B_{jk}(t, U_{jkl}, U_{jkl, t+1}, b; \beta_0, \Lambda_{0k}) dg_k(t) \phi(b; \Sigma_0) db \right]^{1/2}
\]
for some positive constant \( c_0 \). We define a metric space \( \mathcal{V} \) which consists of all functions \( g = (g_1, \ldots, g_K)^2 \) in \( BV[0, \tau_1] \times \cdots \times BV[0, \tau_K] \), the space of bounded variation spaces \( BV[0, \tau_k] \) \((k = 1, \ldots, K)\), with \( g_k(0) = 0 \), and define a norm as
\[
\| g \|_1 = E \left[ \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=0}^{M_{jk}} g_k(U_{jkl})^2 \right]^{1/2}
\]
In addition, we define a seminorm
\[
\| g \|_2 = E \left[ \left( \sum_{j=1}^{J} \sum_{k=1}^{K} \int_{b} \prod_{j'=1}^{J} \prod_{k'=1, k' \neq k}^{K} D_{j'k'}(U_{j'k'}, b; \beta_0, \Lambda_{0k'}) \right) \sum_{l=0}^{M_{jk}} \Delta_{jkl} \int_{0}^{T_k} B_{jk}(t, U_{jkl}, U_{jkl, t+1}, b; \beta_0, \Lambda_{0k}) dg_k(t) \phi(b; \Sigma_0) db \right]^{1/2}
\]
If \( \| g \|_2 = 0 \) for some \( g \in \mathcal{V} \), then
\[
\sum_{j=1}^{J} \sum_{k=1}^{K} \int_{b} \prod_{j'=1}^{J} \prod_{k'=1, k' \neq k}^{K} D_{j'k'}(U_{j'k'}, b; \beta_0, \Lambda_{0k'}) \right) \sum_{l=0}^{M_{jk}} \Delta_{jkl} \int_{0}^{T_k} B_{jk}(t, U_{jkl}, U_{jkl, t+1}, b; \beta_0, \Lambda_{0k}) dg_k(t) \phi(b; \Sigma_0) db = 0
\]
with probability 1. For any \( j \in \{1, \ldots, J\}, k \in \{1, \ldots, K\} \) and \( l, j' \in \{0, \ldots, M_{j'k'}\} \) \((j' = 1, \ldots, J; k' = 1, \ldots, K)\), we evaluate the above equation at all possible values of \( \Delta_{jkl} \) with \((j', k') \in C_{j} = \{1, \ldots, j\} \times \{1, \ldots, k\} \) and \( l = l_{j'k'}, \ldots, M_{j'k'} \) and take the sum of the resulting equations. We then consider all possible values of \( \Delta_{jkl} \) with \((j', k') \notin C_{j} \) and \( l = 0, \ldots, M_{j'k'} \) and take the sum of the resulting equations to obtain
\[
\int_{b} \left\{ \prod_{j'=1}^{J} \prod_{k'=1}^{K} Q_{j'k'}(U_{j'k'}, b; \beta_0, \Lambda_{0k}) \right\} \sum_{j'=1}^{J} \sum_{k'=1}^{K} G_{k'} \left\{ \int_{0}^{T_k} U_{j'k'} e^{\beta_0 T} \Lambda_{j'k'}(t) + b_{j'k'} d\Lambda_{0k'}(t) \right\}
\]
This equality holds for any \( U_{j'k'} \). Thus, for any \( t' < t \) and \( 1 \leq j' < j \leq J, 1 \leq k' < k \leq K \),

\[
\int_0^t \left\{ \prod_{j'=1}^j \prod_{k'=1}^k Q_{j',k'}(t'; b; \beta_0, \Lambda_{0k'}) \right\} \sum_{j'=1}^j \sum_{k'=1}^k G_k' \left\{ \int_0^t e^{(T_{j'k'} + b')Z_{j'k'}(t)} d\Lambda_{0k'}(t) \right\} \times \int_0^t e^{(T_{j'k'} + b')Z_{j'k'}(t)} d\Lambda_{0k'}(t) \phi(b; \Sigma_0) db = 0.
\]

By Condition 8,

\[
G_k' \left\{ \int_0^t e^{(T_{j'k'} + b')Z_{j'k'}(s)} d\Lambda_{0k'}(s) \right\} \times \int_0^t e^{(T_{j'k'} + b')Z_{j'k'}(s)} d\Lambda_{0k'}(s) = 0
\]

for any \( j \in \{1, \ldots, J\} \), \( k \in \{1, \ldots, K\} \) and \( t \in [0, \tau_k] \). The term \( G_k' \left\{ \int_0^t e^{(T_{j'k'} + b')Z_{j'k'}(s)} d\Lambda_{0k'}(s) \right\} \) is bounded away from zero, such that \( \int_0^t e^{(T_{j'k'} + b')Z_{j'k'}(s)} d\Lambda_{0k'}(s) = 0 \) almost surely for any \( t \in [0, \tau_k] \). Therefore, \( g_k(t) = 0 \) for any \( t \in [0, \tau_k] \) and \( k \in \{1, \ldots, K\} \). This implies that \( \| \cdot \|_2 \) is a norm in \( V \). By the Cauchy-Schwarz inequality, for any \( g \in V \),

\[
\|g\|_2 \leq \left\{ \sum_{j=1}^J \sum_{k=1}^K M_{jk} \Delta_{jkl} \right\} \times \int_0^T \left\{ \prod_{j'=1}^j \prod_{k'=1}^k D_{j'k'}(U_{j'k'}, b; \beta_0, \Lambda_{0k'}) \right\} \times \left[ \int_0^T B_{jkl}(t, U_{jkl}, U_{jkl,t+1}; b; \beta_0, \Lambda_{0k}) dt \phi(b; \Sigma_0) db \right]^2 \right\}^{1/2} \leq c_1 \|g\|_1,
\]

where \( c_1 \) is a finite constant. By the bounded inverse theorem in the Banach space, \( \|g\|_2 \geq c'_1 \|g\|_1 \) for some constant \( c'_1 \). Therefore,

\[
O_P(n^{-2/3}) + O(1) \| \gamma - \gamma_0 \|^2 + O(1) \| \hat{\beta} - \beta_0 \|^2 \geq c_0 c'_1 \left\{ \sum_{j=1}^J \sum_{k=1}^K M_{jk} \left\{ \hat{\Lambda_k}(U_{jkl}) - \Lambda_{0k}(U_{jkl}) \right\}^2 \right\}.
\]

The lemma thus holds.

### References

Chen, M. H., Tong, X. & Sun, J. (2009). A frailty model approach for regression analysis of multivariate current status data. *Statist. Med.*, 28, 3424–3436.

Chen, M. H., Chen, L. C., Lin, K. H. & Tong, X. (2014). Analysis of multivariate interval censoring by Diabetic Retinopathy Study. *Commun. Statist. B*, 43, 1825–1835.

Van de Geer, S. A. (2000). *Empirical Processes in M-Estimation*. Cambridge: Cambridge University Press.

Van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.
Table S1: Summary statistics for simulation studies with multiple events

| $r_1 = r_2$ | $n = 100$ | $n = 200$ | $n = 400$ |
|-------------|-----------|-----------|-----------|
|             | Bias      | SE        | SEE       | CP | Bias      | SE        | SEE       | CP | Bias      | SE        | SEE       | CP |
| 0           | $\beta_{11} = 0.5$ | 0.031      | 0.340     | 0.329     | 95 | 0.016      | 0.231     | 0.227     | 95 | 0.009      | 0.160     | 0.159     | 95 |
|             | $\beta_{12} = -0.5$ | -0.030     | 0.586     | 0.577     | 95 | -0.016     | 0.399     | 0.395     | 95 | -0.004     | 0.275     | 0.274     | 95 |
|             | $\beta_{21} = 0.4$ | 0.025      | 0.311     | 0.302     | 95 | 0.012      | 0.209     | 0.208     | 95 | 0.007      | 0.146     | 0.145     | 95 |
|             | $\beta_{22} = 0.2$ | 0.014      | 0.539     | 0.531     | 95 | 0.007      | 0.364     | 0.361     | 95 | 0.007      | 0.252     | 0.250     | 95 |
|             | $\sigma^2 = 0.5$ | 0.029      | 0.361     | 0.394     | 96 | 0.018      | 0.235     | 0.263     | 96 | 0.012      | 0.159     | 0.171     | 96 |
| 0.5         | $\beta_{11} = 0.5$ | 0.031      | 0.400     | 0.389     | 95 | 0.015      | 0.273     | 0.268     | 95 | 0.009      | 0.189     | 0.187     | 95 |
|             | $\beta_{12} = -0.5$ | -0.032     | 0.685     | 0.683     | 95 | -0.017     | 0.469     | 0.467     | 95 | -0.004     | 0.324     | 0.324     | 95 |
|             | $\beta_{21} = 0.4$ | 0.023      | 0.371     | 0.364     | 95 | 0.011      | 0.250     | 0.250     | 95 | 0.005      | 0.176     | 0.176     | 95 |
|             | $\beta_{22} = 0.2$ | 0.010      | 0.641     | 0.637     | 96 | 0.002      | 0.439     | 0.435     | 95 | 0.005      | 0.305     | 0.302     | 95 |
|             | $\sigma^2 = 0.5$ | 0.030      | 0.448     | 0.483     | 95 | 0.017      | 0.296     | 0.319     | 95 | 0.011      | 0.203     | 0.215     | 96 |
| 1           | $\beta_{11} = 0.5$ | 0.033      | 0.459     | 0.448     | 95 | 0.016      | 0.312     | 0.308     | 95 | 0.009      | 0.215     | 0.214     | 95 |
|             | $\beta_{12} = -0.5$ | -0.035     | 0.785     | 0.787     | 96 | -0.019     | 0.535     | 0.536     | 95 | -0.005     | 0.370     | 0.371     | 95 |
|             | $\beta_{21} = 0.4$ | 0.024      | 0.434     | 0.425     | 95 | 0.012      | 0.294     | 0.292     | 95 | 0.005      | 0.206     | 0.204     | 95 |
|             | $\beta_{22} = 0.2$ | 0.007      | 0.749     | 0.744     | 96 | 0.001      | 0.512     | 0.508     | 95 | 0.004      | 0.356     | 0.352     | 95 |
|             | $\sigma^2 = 0.5$ | 0.035      | 0.559     | 0.605     | 94 | 0.019      | 0.372     | 0.401     | 94 | 0.012      | 0.255     | 0.272     | 95 |

SE, SEE and CP stand for empirical standard error, mean standard error estimator and empirical coverage percentage of the 95% confidence interval, respectively. For $\sigma^2$, bias and SEE are based on the median instead of the means, and the confidence interval is based on the log transformation. Each entry is based on 10,000 replicates.
Table S2: Summary statistics for simulation studies with clustered multiple-events data

|                  | $n = 100$ |                  | $n = 200$ |
|------------------|----------|-----------------|----------|
| $r_1 = r_2$      |          | Bias  | SE    | SEE  | CP  | Bias  | SE    | SEE  | CP  |
| 0                |          | 0.022 | 0.293 | 0.285 | 95  | 0.010 | 0.202 | 0.198 | 95  |
| $\beta_{11} = 0.5$ |          | -0.023 | 0.505 | 0.508 | 95  | -0.015 | 0.342 | 0.348 | 95  |
| $\beta_{12} = -0.5$ |          | 0.019 | 0.275 | 0.269 | 95  | 0.010 | 0.189 | 0.187 | 95  |
| $\beta_{21} = 0.4$ |          | 0.002 | 0.480 | 0.478 | 95  | 0.001 | 0.328 | 0.328 | 95  |
| $\sigma^2_1 = 0.5$ |          | 0.013 | 0.243 | 0.311 | 98  | -0.004 | 0.165 | 0.200 | 98  |
| $\sigma^2_2 = 0.8$ |          | 0.042 | 0.307 | 0.367 | 96  | 0.020 | 0.206 | 0.237 | 96  |
| 0.5              |          | 0.020 | 0.327 | 0.325 | 95  | 0.008 | 0.227 | 0.225 | 95  |
| $\beta_{11} = 0.5$ |          | -0.022 | 0.571 | 0.577 | 95  | -0.015 | 0.386 | 0.394 | 95  |
| $\beta_{12} = -0.5$ |          | 0.016 | 0.309 | 0.309 | 95  | 0.008 | 0.216 | 0.213 | 95  |
| $\beta_{21} = 0.4$ |          | -0.001 | 0.542 | 0.546 | 95  | 0.000 | 0.371 | 0.373 | 95  |
| $\sigma^2_1 = 0.5$ |          | -0.026 | 0.283 | 0.367 | 98  | -0.014 | 0.197 | 0.236 | 98  |
| $\sigma^2_2 = 0.8$ |          | 0.041 | 0.382 | 0.451 | 95  | 0.023 | 0.264 | 0.295 | 96  |
| 1                |          | 0.020 | 0.362 | 0.365 | 95  | 0.008 | 0.252 | 0.251 | 95  |
| $\beta_{11} = 0.5$ |          | -0.024 | 0.637 | 0.646 | 95  | -0.018 | 0.433 | 0.439 | 95  |
| $\beta_{12} = -0.5$ |          | 0.016 | 0.349 | 0.350 | 95  | 0.009 | 0.241 | 0.241 | 95  |
| $\beta_{21} = 0.4$ |          | 0.000 | 0.607 | 0.616 | 95  | -0.002 | 0.419 | 0.420 | 95  |
| $\sigma^2_1 = 0.5$ |          | -0.038 | 0.330 | 0.433 | 98  | -0.018 | 0.234 | 0.278 | 97  |
| $\sigma^2_2 = 0.8$ |          | 0.044 | 0.470 | 0.556 | 94  | 0.023 | 0.327 | 0.368 | 95  |

SE, SEE and CP stand for empirical standard error, mean standard error estimator and empirical coverage percentage of the 95% confidence interval, respectively. For $\sigma^2$, bias and SEE are based on the median instead of the means, and the confidence interval is based on the log transformation. Each entry is based on 10,000 replicates.

Table S3: Summary statistics for simulation studies comparing the proposed and Chen et al. (2009)’s methods

|                  | Proposed                  | Chen et al. (2009) |
|------------------|---------------------------|--------------------|
| $n$              | Bias  | SE    | RSE  | SEE  | CP  | Bias  | SE    | RSE  | SEE  | CP  | MSE |
| 100              | 0.25  | 0.046 | 0.288 | 0.284 | 0.282 | 0.954 | 0.085 | 0.045 | 0.297 | 0.267 | 0.256 | 0.942 | 0.090 |
|                  | -0.003 | 0.284 | 0.274 | 0.279 | 0.951 | 0.081 | 0.027 | 0.309 | 0.256 | 0.254 | 0.943 | 0.096 |
|                  | -0.25  | -0.051 | 0.286 | 0.282 | 0.283 | 0.955 | 0.085 | 0.022 | 0.322 | 0.259 | 0.261 | 0.946 | 0.104 |
| 200              | 0.25  | 0.029 | 0.184 | 0.181 | 0.184 | 0.952 | 0.035 | 0.019 | 0.203 | 0.176 | 0.176 | 0.945 | 0.041 |
|                  | 0.00  | 0.001 | 0.182 | 0.185 | 0.182 | 0.955 | 0.033 | 0.021 | 0.770 | 0.175 | 0.175 | 0.950 | 0.593 |
|                  | -0.25  | -0.028 | 0.186 | 0.187 | 0.186 | 0.953 | 0.035 | 0.018 | 0.231 | 0.181 | 0.181 | 0.945 | 0.054 |

SE, RSE, SEE, CP and MSE stand for standard error, robust standard error based on median absolute deviation, mean standard error estimator, coverage probability of the 95% confidence interval and mean squared error, respectively. Each entry is based on 10,000 replicates.
Table S4: Summary statistics for simulation studies comparing the proposed and Chen et al. (2014)’s methods

| #rep | $\beta_1 = 0$ | $\beta_2 = 0$ | $\beta_1 = 0.5$ | $\beta_2 = 0.5$ | $\beta_1 = 1$ | $\beta_2 = 0.5$ | $\beta_1 = 0$ | $\beta_2 = 0$ | $\beta_1 = 0.5$ | $\beta_2 = 0.5$ | $\beta_1 = 1$ | $\beta_2 = 0.5$ |
|------|---------------|---------------|-----------------|-----------------|---------------|---------------|---------------|---------------|----------------|----------------|---------------|---------------|
|      | Bias          | SE            | SEE             | CP              | MSE           | Bias          | SE            | SEE             | CP              | MSE           | Bias          | SE            | SEE             | CP              | MSE           |
| 9918 | 0.001         | 0.259         | 0.254           | 0.950           | 0.067         | 0.001         | 0.254         | 0.244           | 0.943           | 0.064         | 0.000         | 0.230         | 0.220           | 0.945           | 0.053         |
| 9895 | -0.001        | 0.238         | 0.236           | 0.955           | 0.057         | 0.000         | 0.230         | 0.220           | 0.945           | 0.053         | -0.024        | 0.257         | 0.246           | 0.946           | 0.067         |
| 9913 | -0.006        | 0.246         | 0.238           | 0.945           | 0.060         | -0.028        | 0.235         | 0.224           | 0.937           | 0.056         | -0.053        | 0.260         | 0.251           | 0.936           | 0.071         |
|      | -0.015        | 0.273         | 0.264           | 0.943           | 0.075         | -0.053        | 0.260         | 0.251           | 0.936           | 0.071         | -0.025        | 0.240         | 0.227           | 0.940           | 0.058         |
| 9976 | -0.002        | 0.174         | 0.176           | 0.956           | 0.030         | 0.000         | 0.174         | 0.171           | 0.949           | 0.030         | -0.024        | 0.165         | 0.161           | 0.944           | 0.025         |
| 9965 | 0.002         | 0.160         | 0.161           | 0.953           | 0.026         | 0.003         | 0.159         | 0.155           | 0.944           | 0.025         | -0.034        | 0.179         | 0.172           | 0.936           | 0.033         |
| 9978 | -0.023        | 0.180         | 0.177           | 0.944           | 0.033         | -0.042        | 0.162         | 0.157           | 0.937           | 0.028         | -0.064        | 0.182         | 0.175           | 0.920           | 0.037         |
|      | -0.027        | 0.165         | 0.163           | 0.947           | 0.028         | -0.070        | 0.166         | 0.159           | 0.934           | 0.029         | -0.037        | 0.166         | 0.159           | 0.934           | 0.029         |

SE, SEE, CP and MSE stand for standard error, mean standard error estimator, coverage probability of the 95% confidence interval and mean squared error, respectively. #rep is the number of replicates with non-NA estimates by Chen et al. (2014)’s method.

Fig. S1: Estimation of $\Lambda_1(t)$ for multiple events data. The solid and dashed curves pertain to the true values and mean estimates, respectively. Each estimate is based on 10,000 replicates.
Fig. S2: Estimation of $\Lambda_2(t)$ for multiple events data. The solid and dashed curves pertain to the true values and mean estimates, respectively. Each estimate is based on 10,000 replicates.
Fig. S3: Log-likelihood at the nonparametric maximum likelihood estimates as a function of $r_1$ and $r_2$ in the logarithmic families for the Atherosclerosis Risk in Communities Study.