

RESEARCH ARTICLE

Weighted Estimation of the Accelerated Failure Time Model in the Presence of Dependent Censoring

Youngjoo Cho^{1*}, Debashis Ghosh²

1 Department of Statistics, The Pennsylvania State University, University Park, Pennsylvania, 16802, United States of America, **2** Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado, Aurora, Colorado, 80045, United States of America

* yvc5154@psu.edu



Abstract

Independent censoring is a crucial assumption in survival analysis. However, this is impractical in many medical studies, where the presence of dependent censoring leads to difficulty in analyzing covariate effects on disease outcomes. The semicompeting risks framework offers one approach to handling dependent censoring. There are two representative estimators based on an artificial censoring technique in this data structure. However, neither of these estimators is better than another with respect to efficiency (standard error). In this paper, we propose a new weighted estimator for the accelerated failure time (AFT) model under dependent censoring. One of the advantages in our approach is that these weights are optimal among all the linear combinations of the previously mentioned two estimators. To calculate these weights, a novel resampling-based scheme is employed. Attendant asymptotic statistical results for the estimator are established. In addition, simulation studies, as well as an application to real data, show the gains in efficiency for our estimator.

OPEN ACCESS

Citation: Cho Y, Ghosh D (2015) Weighted Estimation of the Accelerated Failure Time Model in the Presence of Dependent Censoring. PLoS ONE 10(4): e0124381. doi:10.1371/journal.pone.0124381

Academic Editor: Xi Luo, Brown University, UNITED STATES

Received: October 19, 2014

Accepted: March 1, 2015

Published: April 24, 2015

Copyright: © 2015 Cho, Ghosh. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Third-party data are available from AIDS Clinical Trial Group 364. Requests for data may be sent to sdac.data@sdac.harvard.edu. Please also see the full AIDS Clinical Trial Group Access to Published Data page here: <https://actgnetwork.org/clinical-trials/access-published-data>.

Funding: This work was supported by National Institute of Health (NIH) grant CA 129102-05. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Introduction

In medical studies, it is very common that death or withdrawal of study and progression on disease of interest simultaneously occur in the study. For this case, death or withdrawal of study may censor the development of disease. This type of data structure is called 'semicompeting risks data' [4].

Semicompeting risks data have been widely studied in the past decade. Some researchers used a Gamma copula to estimate the association parameter between the event of interest and dependent censoring [2], [4]. There is a literature that extended the methodology of [2] to the case that a nuisance parameter exists and also considered a more general copula model [22].

On the other hand, other researchers used semiparametric regression to model the event of interest and dependent censoring jointly. One approach is an estimation procedure based on the accelerated failure time (AFT) model [13], [17]. They used the artificial censoring

Competing Interests: The authors have declared that no competing interests exist.

technique to adjust the bias of the usual estimator. While the estimating equation of [13] is a U-statistic of order one, that of [17] is a U-statistic of order 2.

However, none of these papers fully discussed optimality of the estimator. In this case, choosing an estimator that is optimal from an efficiency viewpoint is an important issue for consideration. Here, we adapt the idea of [25], which proposed an optimal estimator whose form is a linear combination of estimators for multivariate failure time data. They used idea of [24], which proposed using combinations of dependent tests in the presence of missing values. Idea of [24] is to create a test which can maximize power based on linear combination of test statistics. Approach of [25] is simple and flexible, so it is sensible to apply their method in our case.

In this paper, we propose a weighted estimator by using methodology from [25]. Our weighted estimator combines those of [13] and [17]. The structure of this paper is as follows. In methods section, we review estimators proposed by [13] and [17] briefly. In addition, we describe details on our new weighted estimator. In model checking section, model checking procedure is briefly discussed. In simulation studies section, results of simulation studies will be given. Application of our method to a real data example is presented in real data analysis section. Some discussion concludes discussion section.

Methods

Review of Model

Let X be time to the event of interest, D the time to dependent censoring and C the time to independent censoring. All these times are transformed on a logarithmic scale. Let $\tilde{X} = X \wedge D \wedge C$ and $\tilde{D} = D \wedge C$. Define $\delta = I(X \leq \tilde{D})$, $\Delta = I(D \leq C)$ and let \mathbf{Z} be covariates. The data contain n independent and identically distributed observations $(\tilde{X}_i, \tilde{D}_i, \mathbf{Z}_i, \delta_i, \Delta_i)$, $i = 1, \dots, n$. The model is

$$\begin{pmatrix} D_i = \mathbf{Z}_i^T \boldsymbol{\eta}_0 + \epsilon_i^D \\ X_i = \mathbf{Z}_i^T \boldsymbol{\theta}_0 + \epsilon_i^X \end{pmatrix}, \quad i = 1, \dots, n.$$

where $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$ are $k \times 1$ vectors, and $\epsilon_i \equiv (\epsilon_i^X, \epsilon_i^D)$ are error terms with an unknown joint distribution. In this case, we assume that the model is identifiable only in upper wedge $X < D$ [4], [17]. We assume that ϵ has unknown distribution H . The goal is to obtain an unbiased estimator of $\boldsymbol{\alpha} = (\boldsymbol{\eta}^T, \boldsymbol{\theta}^T)^T$ without nonparametrically estimating the distribution of ϵ_i , $i = 1, \dots, n$. We further assume that given \mathbf{Z} , C and (X, D) are independent, but X and D can be dependent given \mathbf{Z} . Now we are going to describe the procedures of [13] and [17] in turn.

Since \tilde{D} only depends on independent censoring, a standard rank regression approach is available for estimation [11], [13], [15], [17], [20], [23]. The estimating equation for $\boldsymbol{\eta}$ is given by

$$\mathbf{S}_n(\boldsymbol{\eta}) = n^{-1/2} \sum_{i=1}^n \Delta_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n \mathbf{Z}_j I\{\tilde{D}_j^*(\boldsymbol{\eta}) \geq \tilde{D}_i^*(\boldsymbol{\eta})\}}{\sum_{j=1}^n I\{\tilde{D}_j^*(\boldsymbol{\eta}) \geq \tilde{D}_i^*(\boldsymbol{\eta})\}} \right],$$

where $\tilde{D}_i^*(\boldsymbol{\eta}) = \tilde{D}_i - \mathbf{Z}_i^T \boldsymbol{\eta}$. The estimator of $\boldsymbol{\eta}$ can be obtained by solving $\mathbf{S}_n(\boldsymbol{\eta}) = 0$.

For estimation of $\boldsymbol{\theta}$, simply replacing $\tilde{D}_i - \mathbf{Z}_i^T \boldsymbol{\eta}$ to $\tilde{X}_i - \mathbf{Z}_i^T \boldsymbol{\theta}$ does not yield unbiased estimation of $\boldsymbol{\theta}$. This is because the cause-specific hazard function for $\tilde{X}_i - \mathbf{Z}_i^T \boldsymbol{\theta}$ depends on $\tilde{D}_i - \mathbf{Z}_i^T \boldsymbol{\theta}$, which violates the model assumption [13]. To fix this problem, many authors use artificial censoring techniques [3], [6], [7], [10], [13], [17]. In [13], a single constant term $g(\boldsymbol{\alpha})$ is

proposed so that the estimation equation will be unbiased for estimation of θ in the two-sample case. The form of $g(\alpha)$ is $g(\alpha) = \max_{1 \leq i \leq n} \{0, \mathbf{Z}_i^T(\theta - \eta)\}$. The proposed estimator in [13] is obtained by solving $U_n^L(\alpha) = 0$, where

$$\begin{aligned}
 U_n^L(\alpha) &= n^{-1/2} \sum_{i=1}^n \tilde{\delta}_i^*(\alpha) \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\alpha) \geq \tilde{X}_i^*(\alpha)\} \mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\alpha) \geq \tilde{X}_i^*(\alpha)\}} \right] \\
 \tilde{X}_i^*(\alpha) &= (X_i - \mathbf{Z}_i^T \theta) \wedge (D_i - \mathbf{Z}_i^T \eta - g(\alpha)) \wedge (C_i - \mathbf{Z}_i^T \eta - g(\alpha)) \\
 \tilde{\delta}_i^*(\alpha) &= I\{(X_i - \mathbf{Z}_i^T \theta) \leq (D_i - \mathbf{Z}_i^T \eta - g(\alpha)) \wedge (C_i - \mathbf{Z}_i^T \eta - g(\alpha))\},
 \end{aligned}
 \tag{1}$$

and $a \wedge b$ means minimum of a and b . In [17], pairwise comparisons of all the subjects is proposed so that each subject has different degree of the artificial censoring. The transformations suggested by [17] are

$$\begin{aligned}
 g_{ij}(\alpha) &= \max_{ij} \{0, \mathbf{Z}_i^T(\theta - \eta), \mathbf{Z}_j^T(\theta - \eta)\} \\
 \tilde{X}_{i(j)}^*(\alpha) &= (X_i - \mathbf{Z}_i^T \theta) \wedge (D_i - \mathbf{Z}_i^T \eta - g_{ij}(\alpha)) \wedge (C_i - \mathbf{Z}_i^T \eta - g_{ij}(\alpha)) \\
 \tilde{\delta}_{i(j)}^*(\alpha) &= I\{(X_i - \mathbf{Z}_i^T \theta) \leq (D_i - \mathbf{Z}_i^T \eta - g_{ij}(\alpha)) \wedge (C_i - \mathbf{Z}_i^T \eta - g_{ij}(\alpha))\} \\
 \phi_{ij}(\alpha) &= \tilde{\delta}_{i(j)}^*(\alpha) I\{\tilde{X}_{i(j)}^*(\alpha) \leq \tilde{X}_{j(i)}^*(\alpha)\} - \tilde{\delta}_{j(i)}^*(\alpha) I\{\tilde{X}_{j(i)}^*(\alpha) \leq \tilde{X}_{i(j)}^*(\alpha)\}.
 \end{aligned}
 \tag{2}$$

The proposed estimator according to [17] is obtained by solving $U_n^P(\alpha) = 0$, which is defined by

$$U_n^P(\alpha) = \frac{2n^{1/2}}{n(n-1)} \sum_{1 \leq i < j \leq n} (\mathbf{Z}_i - \mathbf{Z}_j) \phi_{ij}(\alpha).$$

Note that X and D are not observable, but we can express transformation (Eq 1) and (Eq 2) by using observable quantities [7], [13].

Weighted estimator

Given these two estimation procedures, it is natural to consider their efficiencies with respect to standard error. However, in this point of view, neither estimator is superior to the other. Moreover, these estimators may not be optimal estimators with respect to the standard error. There is an argument that estimator of [17] gains more efficiency than that of [13] because pairwise comparisons lead to less artificial censoring than that in [13]. However, this logic only holds when we look at performance of estimators in the view of bias and variance across the estimators in simulation study. Concentrating on standard error of an estimator in a single dataset, the estimator by [17] may not provide better estimator than that of [13]. This will be seen in the real data analysis section.

The reason for this is due to estimation procedure of [17]. As discussed [7], for n samples, the number of comparisons of [13] for artificial censoring is of order n , while that of [17] is of order n^2 . By definition of $g_{ij}(\alpha)$, different degrees of artificial censoring is applied to observations. It may lead more variation between observations, which makes standard error larger than that of [13].

Having discussed our data structure and estimators from [13] and [17], we now describe the proposed estimation in this paper. Let $\hat{\eta} = (\hat{\eta}_1, \dots, \hat{\eta}_k)^T$ be estimator of η_0 , $\hat{\theta}^L = (\hat{\theta}_1^L, \dots, \hat{\theta}_k^L)^T$ be estimator of θ_0 by [13] and $\hat{\theta}^P = (\hat{\theta}_1^P, \dots, \hat{\theta}_k^P)^T$ be estimator of θ_0 by [17]. $\hat{\theta}^L$ and $\hat{\theta}^P$ are asymptotically unbiased estimators of θ_0 .

We extend the scope of estimators which provide consistent estimation of θ_0 . The natural extension of estimators of [13] and [17] is to consider collections of estimators that are linear combination of these two estimators with sum of weights being 1. By choosing proper weights, we can expect that the variance of the new combined estimator is smaller than that of each individual estimator in $\hat{\theta}^L$ and $\hat{\theta}^P$.

The goal is to find weights such that the variance of the new estimator is smaller than the minimum of variance of the estimators by [13] and [17], which have good theoretical properties. To obtain the estimator that yields smallest variance with these properties, we can use the idea of [25], which was applied to the problem of modeling multivariate failure times.

In [25], the joint distribution of estimators $\hat{\gamma} = \{\hat{\gamma}_{mr}\}$ is considered, where $m = 1, \dots, k$ and $r = 1, \dots, R$. In this case, m indicates index of regression parameters and r stands for index of the r th event. For obtaining an optimal estimator, they applied arguments from [24] which derived a linear combination of test statistic to maximize power against every alternative hypothesis.

Let $\hat{\mathbf{H}}$ be the covariance matrix for the estimators $\hat{\gamma}$. Then we fix m and define $\hat{\mathbf{H}}_m$ be covariance matrix of $\hat{\gamma}_m = (\hat{\gamma}_{m1}, \dots, \hat{\gamma}_{mR})$. It can be obtained from the entire covariance matrix by selecting the part corresponding to $\hat{\gamma}$ for $r = 1, \dots, R$ under fixed m . Now we can define $\sum_{r=1}^R d_r \hat{\gamma}_{mr}$, where $\mathbf{d} = (d_1, d_2, \dots, d_R)$ satisfies $\sum_{r=1}^R d_r = 1$ [25]. Then $\mathbf{d} \equiv (\mathbf{e}^T \hat{\mathbf{H}}_m^{-1} \mathbf{e})^{-1} \hat{\mathbf{H}}_m^{-1} \mathbf{e}$ is a vector of weights which leads the best estimator among linear combinations of estimators of $\hat{\gamma}_m$ where \mathbf{e} is a vector consisting of R ones [24], [25].

We now apply the argument in previous paragraph to our model by considering the joint distribution of $\hat{\boldsymbol{\beta}} = \{\hat{\boldsymbol{\eta}}^T, (\hat{\boldsymbol{\theta}}^L)^T, (\hat{\boldsymbol{\theta}}^P)^T\}^T$. Let $\boldsymbol{\beta}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^L, \boldsymbol{\theta}_0^P)^T$ and $\mathbf{G}_n(\boldsymbol{\beta}) = [\mathbf{S}_n^T(\eta), \{\mathbf{U}_n^L(\boldsymbol{\alpha})\}^T, \{\mathbf{U}_n^P(\boldsymbol{\alpha})\}^T]^T$ where $[\mathbf{S}_n^T(\eta), \{\mathbf{U}_n^L(\boldsymbol{\alpha})\}^T, \{\mathbf{U}_n^P(\boldsymbol{\alpha})\}^T]^T$ are estimating equations for $\boldsymbol{\beta}_0$. The strong consistency and asymptotic joint distribution of three estimators, described in following theorems, play a crucial role in our methodology.

To prove asymptotic results, several regularity conditions are required. As stated in [7] and [17], define

$$F(a, b, c, d, e) = P(\epsilon_1^X - \epsilon_1^D \leq a, \epsilon_1^X - C_1 \leq b, \epsilon_1^X - \epsilon_2^X \leq c, \epsilon_1^X - \epsilon_2^D \leq d, \epsilon_1^X - C_2 \leq e | \mathbf{Z}_1, \mathbf{Z}_2)$$

Let $\boldsymbol{\alpha}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^T)^T$. Define

$$\begin{aligned} T_1(\mathbf{Z}_1, \mathbf{Z}_2) &= \frac{\partial F}{\partial a} \{g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_1^T \boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0), 0, g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_2^T \boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0)\} \\ &+ \frac{\partial F}{\partial b} \{g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_1^T \boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0), 0, g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_2^T \boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0)\} \\ &+ \frac{\partial F}{\partial d} \{g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_1^T \boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0), 0, g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_2^T \boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0)\} \\ &+ \frac{\partial F}{\partial e} \{g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_1^T \boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0), 0, g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_2^T \boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0)\} \\ &+ 2 \frac{\partial F}{\partial c} \{g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_1^T \boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0), 0, g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_2^T \boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0)\} \end{aligned}$$

and

$$T_2(\mathbf{Z}_1, \mathbf{Z}_2) = T_1(\mathbf{Z}_1, \mathbf{Z}_2) - 2 \frac{\partial F}{\partial c} \{g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_1^T \boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0), 0, g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_2^T \boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0)\}$$

From the Appendix in [17], the additional conditions are as follows:

1. The parameter space \mathcal{W} is compact, and the true parameter $\boldsymbol{\alpha}_0$ is an interior point of \mathcal{W} .

2. θ_0 is the only solution of the estimating equation $E\{n^{-1/2}U_n^P(\eta_0, \theta)\} = 0$.
3. $E(\|Z\|^2) < \infty$, where $\|\cdot\|$ is Euclidean norm and there exists positive constant K such that partial derivatives of F are bounded by K and there exists positive constant K^* such that marginal probability distribution of F is bounded by K^* almost surely.
4. $cov[(Z_1 - Z_2)\{T_1(Z_1, Z_2)\}^{1/2}]$ and $cov[(Z_1 - Z_2)\{T_2(Z_1, Z_2)\}^{1/2}]$ are positive definite.

In many parts of proofs, we adapt arguments from [13] and [17].

Theorem 1. By conditions of C1 – C3 in Appendix of [17] and conditions in [27], $\hat{\beta}$ is (strongly) consistent.

Proof. Let $\hat{\beta} = \{\hat{\eta}^T, (\hat{\theta}^L)^T, (\hat{\theta}^P)^T\}^T$. It suffices to show that $\hat{\eta}$, $\hat{\theta}^L$ and $\hat{\theta}^P$ are strongly consistent, respectively. Let $\alpha = (\eta^T, \theta^T)^T$. Note that we have compact region, say \mathcal{W} and we assume regularity conditions in [27]. By [27], there exists nonrandom function m_1 such that $\sup_{\eta \in \mathcal{N}_0} \|n^{-1/2} S_n(\eta) - m_1(\eta)\|$ converges to 0 with probability 1 where \mathcal{N}_0 is a neighborhood of η_0 . Thus $\hat{\eta}$ is strongly consistent. Similarly, we have another nonrandom function m_2 such that $\sup_{\alpha \in \mathcal{N}_1} \|n^{-1/2} U_n^L(\alpha) - m_2(\alpha)\|$ converges to 0 with probability 1 where \mathcal{N}_1 is a neighborhood of α_0 . Hence by [27], $\hat{\alpha}^L$ is strongly consistent.

For $\hat{\theta}^P$, by argument in Appendix of [17], note that by the U-statistics version of the law of large numbers, for all $\alpha \in \mathcal{W}$, $\|n^{-1/2} U_n^P(\alpha) - \gamma(\alpha)\|$ converges to 0 in probability where $\gamma(\alpha) = E\{n^{-1/2} U_n^P(\alpha)\}$. We can partition our compact space as $\mathcal{W}_1, \dots, \mathcal{W}_k$ so that $\mathcal{W} \in \cup_{j=1}^k \mathcal{W}_j$. Clearly, then for $\{\alpha^j \in \mathcal{W}_j, j = 1, \dots, k\}$, $\max_{1 \leq j \leq k} \|n^{-1/2} U_n^P(\alpha^j) - \gamma(\alpha^j)\|$ converges to 0 in probability. Then by Appendix of [17],

$$\begin{aligned} & \sup_{\|\alpha - \alpha^{**}\| \leq \xi} n^{-1/2} \|U_n^P(\alpha) - U_n^P(\alpha^{**})\| \\ & \leq \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \|Z_i - Z_j\| \sup_{\|\alpha - \alpha^{**}\| \leq \xi} |\phi_{ij}(\alpha) - \phi_{ij}(\alpha^{**})| \end{aligned}$$

and for all $\epsilon > 0$, there exists $\xi > 0$ such that

$$\lim_{n \rightarrow \infty} P\left(\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \|Z_i - Z_j\| \sup_{\|\alpha - \alpha^{**}\| \leq \xi} |\phi_{ij}(\alpha) - \phi_{ij}(\alpha^{**})| \geq \epsilon\right) = 0$$

Hence

$$\lim_{n \rightarrow \infty} P\left(\sup_{\|\alpha - \alpha^{**}\| \leq \xi} n^{-1/2} \|U_n^P(\alpha) - U_n^P(\alpha^{**})\| \geq \epsilon\right) = 0$$

Thus $\hat{\theta}^P$ is strongly consistent and clearly, $\hat{\beta}$ is strongly consistent.

Theorem 2. Assuming certain technical conditions from [27] and [17], $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically normal with mean zero vector and covariance matrix Σ_0 where $\Sigma_0 = \Gamma_0^{-1} \Omega_0 \Gamma_0^{-1}$, where Γ_0 is a nonsingular matrix and Ω_0 is the asymptotic covariance matrix of $G_n(\beta_0)$.

Proof. As consistency, we assume the same regularity conditions as in [27]. Let $\beta_0 = (\eta_0^T, \theta_0^L, \theta_0^P)^T$ and $G_n(\beta) = [S_n^T(\eta), \{U_n^L(\alpha)\}^T, \{U_n^P(\alpha)\}^T]^T$. Similar to [13], let $\lambda_0^{(1)}(t)$ be the cause-specific hazard function for the $\tilde{D}_i^*(\eta)$ and let $\lambda_0^{(2)}(t)$ be the cause-specific hazard

function for $\tilde{X}_i^*(\boldsymbol{\alpha})$ under dependent censoring. Define

$$M_{1i}(t) = N_{1i}(t; \boldsymbol{\eta}_0) - \int_{-\infty}^t I\{\tilde{D}_i^*(\boldsymbol{\eta}_0) \geq u\} \lambda_0^{(1)}(u) du \tag{3}$$

$$M_{2i}(t) = N_{2i}(t; \boldsymbol{\alpha}_0) - \int_{-\infty}^t I\{\tilde{X}_i^*(\boldsymbol{\alpha}_0) \geq u\} \lambda_0^{(2)}(u) du \tag{4}$$

Then M_{1i} and M_{2i} are martingales [5], [13]. By adapting a proof in the Appendix in [13], Rebdollo's martingale central limit theorem [5] gives

$$\mathbf{S}_n(\boldsymbol{\eta}_0) = n^{-1/2} \sum_{i=1}^n \int_{-\infty}^{\infty} \{\mathbf{Z}_i - \bar{\mathbf{Z}}^{(1)}(u)\} dM_{1i}(u) + o_p(1)$$

$$\mathbf{U}_n^L(\boldsymbol{\alpha}_0) = n^{-1/2} \sum_{i=1}^n \int_{-\infty}^{\infty} \{\mathbf{Z}_i - \bar{\mathbf{Z}}^{(2)}(u)\} dM_{2i}(u) + o_p(1)$$

where $\bar{\mathbf{Z}}^{(1)}(u) = \lim_{n \rightarrow \infty} [\sum_{j=1}^n I\{\tilde{D}_j^*(\boldsymbol{\eta}_0) \geq u\} \mathbf{Z}_j] / [\sum_{j=1}^n I\{\tilde{D}_j^*(\boldsymbol{\eta}_0) \geq u\}]$ and $\bar{\mathbf{Z}}^{(2)}(u) = \lim_{n \rightarrow \infty} [\sum_{j=1}^n I\{\tilde{X}_j^*(\boldsymbol{\alpha}_0) \geq u\} \mathbf{Z}_j] / [\sum_{j=1}^n I\{\tilde{X}_j^*(\boldsymbol{\alpha}_0) \geq u\}]$. From Appendix of [17],

$$\mathbf{U}_n^P(\boldsymbol{\alpha}_0) = n^{-1/2} \sum_{i=1}^n 2\mathbf{h}_1(\mathbf{V}_i, \boldsymbol{\alpha}_0) + o_p(1)$$

where $2\mathbf{h}_1(\mathbf{v}, \boldsymbol{\alpha}_0) = 2E[\mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\alpha}_0)]$. For $j = 1, \dots, n$, $M_{1j}(t)$ is the martingale associated with ϵ_j^D , while $M_{2j}(t)$ is the martingale associated with ϵ_j^X and $\mathbf{h}(\mathbf{V}_j, \mathbf{V}_j, \boldsymbol{\alpha}) = (\mathbf{Z}_i - \mathbf{Z}_j)\phi_{ij}(\boldsymbol{\alpha})$ [13], [17]. For $j = 1, \dots, n$, define

$$\mathbf{a}_{0j} = \int_{-\infty}^{\infty} \{\mathbf{Z}_j - \bar{\mathbf{Z}}^{(1)}(u)\} dM_{1j}(u) \quad \mathbf{a}_{1j} = \int_{-\infty}^{\infty} \{\mathbf{Z}_j - \bar{\mathbf{Z}}^{(2)}(u)\} dM_{2j}(u)$$

$$\mathbf{a}_{2j} = 2\mathbf{h}_1(\mathbf{V}_j, \boldsymbol{\alpha}_0),$$

By the Cramér-Wold theorem, $\mathbf{G}_n(\boldsymbol{\beta}_0)$ has an asymptotically normal distribution with mean zero and covariance matrix $\boldsymbol{\Omega}_0$, where

$$\boldsymbol{\Omega}_0 = E \begin{pmatrix} \mathbf{a}_{01}\mathbf{a}_{01}^T & \mathbf{a}_{01}\mathbf{a}_{11}^T & \mathbf{a}_{01}\mathbf{a}_{21}^T \\ \mathbf{a}_{11}\mathbf{a}_{01}^T & \mathbf{a}_{11}\mathbf{a}_{11}^T & \mathbf{a}_{11}\mathbf{a}_{21}^T \\ \mathbf{a}_{21}\mathbf{a}_{01}^T & \mathbf{a}_{21}\mathbf{a}_{11}^T & \mathbf{a}_{21}\mathbf{a}_{21}^T \end{pmatrix}$$

Note that $E\{n^{-1/2}\mathbf{U}_n^P(\boldsymbol{\alpha})\} = \boldsymbol{\gamma}(\boldsymbol{\alpha})$. As stated in the Appendix of [17], under conditions of N1 – N3 from [9], there exists an open neighborhood of $\boldsymbol{\alpha}_0$, say K_0 , such that

$$\sup_{\boldsymbol{\alpha} \in K_0} \frac{\|\mathbf{U}_n^P(\boldsymbol{\alpha}) - \mathbf{U}_n^P(\boldsymbol{\alpha}_0) - n^{1/2}\boldsymbol{\gamma}(\boldsymbol{\alpha})\|}{1 + n^{1/2}\|\boldsymbol{\gamma}(\boldsymbol{\alpha})\|} = o_p(1) \tag{5}$$

Using a Taylor series expansion of $\boldsymbol{\gamma}(\boldsymbol{\alpha})$ around $\boldsymbol{\alpha}_0$,

$$\boldsymbol{\gamma}(\boldsymbol{\alpha}) = \boldsymbol{\gamma}(\boldsymbol{\alpha}_0) + \frac{\partial \boldsymbol{\gamma}(\boldsymbol{\alpha})}{\partial \boldsymbol{\eta}} \Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} (\boldsymbol{\eta} - \boldsymbol{\eta}_0) + \frac{\partial \boldsymbol{\gamma}(\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|) \tag{6}$$

With these two results (Eq 5) and (Eq 6), by Appendix of [17],

$$U_n^p(\boldsymbol{\alpha}) = U_n^p(\boldsymbol{\alpha}_0) + n^{1/2} \frac{\partial \gamma(\boldsymbol{\alpha})}{\partial \boldsymbol{\eta}} \Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} (\boldsymbol{\eta} - \boldsymbol{\eta}_0) + n^{1/2} \frac{\partial \gamma(\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(1 + n^{1/2} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|) \quad (7)$$

From [27], we have that

$$S_n(\boldsymbol{\eta}) = S_n(\boldsymbol{\eta}_0) + n^{1/2} \mathbf{P}_0 (\boldsymbol{\eta} - \boldsymbol{\eta}_0) + o_p(1) \quad (8)$$

for any $\boldsymbol{\eta}$ in the small neighborhood of $\boldsymbol{\eta}_0$, where \mathbf{P}_0 is $k \times k$ nonsingular matrix. From the Appendix in [13], for $\mathbf{J}_{1n}(\boldsymbol{\alpha}) = [S_n^T(\boldsymbol{\eta}), \{U_n^l(\boldsymbol{\alpha})\}^T]^T$,

$$\mathbf{J}_{1n}(\boldsymbol{\alpha}) = \mathbf{J}_{1n}(\boldsymbol{\alpha}_0) + n^{1/2} \mathbf{L}_{10} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + o_p(1) \quad (9)$$

for any $\boldsymbol{\alpha}$ in the small neighborhood of $\boldsymbol{\alpha}_0$, where \mathbf{L}_{10} is defined as

$$\mathbf{L}_{10} = \begin{pmatrix} \mathbf{P}_0 & 0 \\ \mathbf{M}_0 & \mathbf{H}_0 \end{pmatrix}$$

is $2k \times 2k$ nonsingular matrix and \mathbf{M}_0 and \mathbf{H}_0 are $k \times k$ constant matrices. Define

$\mathbf{J}_{2n}(\boldsymbol{\alpha}) = [S_n^T(\boldsymbol{\eta}), \{U_n^p(\boldsymbol{\alpha})\}^T]^T$. Using expansion from [17], for any $\boldsymbol{\alpha}$ in the small neighborhood of $\boldsymbol{\alpha}_0$,

$$\mathbf{J}_{2n}(\boldsymbol{\alpha}) = \mathbf{J}_{2n}(\boldsymbol{\alpha}_0) + n^{1/2} \mathbf{L}_{20} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + o_p(1) \quad (10)$$

$$\mathbf{L}_{20} = \begin{pmatrix} \mathbf{P}_0 & 0 \\ \mathbf{R}_0 & \mathbf{V}_0 \end{pmatrix}$$

where $\mathbf{R}_0 = \frac{\partial \gamma(\boldsymbol{\alpha})}{\partial \boldsymbol{\eta}} \Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$ and $\mathbf{V}_0 = \frac{\partial \gamma(\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$. Combining expansions of (Eq 8), (Eq 9) and (Eq 10), we have

$$\mathbf{G}_n(\boldsymbol{\beta}) = \mathbf{G}_n(\boldsymbol{\beta}_0) + n^{1/2} \boldsymbol{\Gamma}_0 (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o_p(1)$$

for any $\boldsymbol{\beta}$ in the small neighborhood of $\boldsymbol{\beta}_0$, where $\boldsymbol{\Gamma}_0$ is defined as

$$\boldsymbol{\Gamma}_0 = \begin{pmatrix} \mathbf{P}_0 & 0 & 0 \\ \mathbf{M}_0 & \mathbf{H}_0 & 0 \\ \mathbf{R}_0 & 0 & \mathbf{V}_0 \end{pmatrix}$$

The results from [9] and [27], along with the consistency of $\hat{\boldsymbol{\beta}}$, imply that

$$n^{1/2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\boldsymbol{\Gamma}_0^{-1} \mathbf{G}_n(\boldsymbol{\beta}_0) + o_p(1)$$

By combining the above results with Slutsky's theorem, $n^{1/2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ has an asymptotically normal distribution with mean zero and covariance matrix $\boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Omega}_0 \boldsymbol{\Gamma}_0^{-1}$.

Theorem 2 implies the asymptotic normality of $\hat{\boldsymbol{\beta}}$ with the form of $\boldsymbol{\Sigma}_0$ being

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23} \\ \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33} \end{pmatrix}.$$

Let $\hat{\Sigma}$ be the estimated covariance matrix of Σ_0 . In this covariance matrix, $\hat{\Sigma}_{11}$ is a $k \times k$ covariance matrix for $\hat{\eta}$, $\hat{\Sigma}_{22}$ is a $k \times k$ covariance matrix for $\hat{\theta}^L$ and $\hat{\Sigma}_{33}$ is a $k \times k$ covariance matrix for $\hat{\theta}^P$. Moreover, $\hat{\Sigma}_{12}$ and $\hat{\Sigma}_{13}$ represent covariance terms between $\hat{\eta}$ and $\hat{\theta}^L$ and between $\hat{\eta}$ and $\hat{\theta}^P$, respectively. Define $\hat{\Sigma}_{23}$ as the covariance matrix between $\hat{\theta}^L$ and $\hat{\theta}^P$. Clearly, $\hat{\Sigma}_{21} = \hat{\Sigma}_{12}^T$, $\hat{\Sigma}_{31} = \hat{\Sigma}_{13}^T$ and $\hat{\Sigma}_{32} = \hat{\Sigma}_{23}^T$.

The issue remains of how to obtain the matrix corresponding to \hat{H}_m^{-1} in our context. Note that $\hat{\eta}$, $\hat{\theta}^L$ and $\hat{\theta}^P$ are correlated with each other. The estimating equation structure implies that $\hat{\theta}^L$ and $\hat{\theta}^P$ cannot be estimated separately from $\hat{\eta}$. Thus our matrix corresponding to \hat{H}_m^{-1} should include the effect of $\hat{\eta}$. To obtain the matrix, we need to invert whole matrix and extract submatrix corresponding to $\hat{\theta}^L$ and $\hat{\theta}^P$. There are two approaches to obtain submatrix.

The first approach is to invert $\hat{\Sigma}$ and obtain the submatrix of $\hat{\Sigma}^{-1}$ corresponding to $\hat{\theta}_m^L$ and $\hat{\theta}_m^P$. Let us denote this matrix as $\hat{\Sigma}_m^*$. Clearly, this matrix is 2×2 and also positive definite. Then we can calculate $\hat{c}_m = (\hat{c}_{m1}, \hat{c}_{m2})^T = (\mathbf{h}^T \hat{\Sigma}_m^* \mathbf{h})^{-1} \hat{\Sigma}_m^* \mathbf{h}$, where $\mathbf{h} = (1, 1)^T$. By using the form of the optimal estimator in [25], we obtain new weighted estimator for m th covariate, say $\hat{\theta}_m^{MWE}$, where

$$\hat{\theta}_m^{MWE} = \hat{c}_{m1} \hat{\theta}_m^L + \hat{c}_{m2} \hat{\theta}_m^P.$$

We can repeat this step for the other regression coefficients. Then we obtain $\hat{\theta}^{MWE} = (\hat{\theta}_1^{MWE}, \dots, \hat{\theta}_k^{MWE})^T$. In this first approach, weights are generated through using k number of 2×2 matrices. We can refer this first approach as ‘marginal approach’.

Sometimes it is desirable to consider entire covariates all at once when obtaining weights. The second approach is to obtain the corresponding submatrix of $\hat{\Sigma}^{-1}$ for $\{(\hat{\theta}^L)^T, (\hat{\theta}^P)^T\}^T$. We denote this matrix as $\hat{\Sigma}^{**}$. This approach is different from first one in that $\hat{\Gamma}_m$ consists of elements of the covariance matrix from $\hat{\theta}_m^L$ and $\hat{\theta}_m^P$ but now $\hat{\Sigma}^{**}$ has elements of covariance matrix from corresponding entire $\{(\hat{\theta}^L)^T, (\hat{\theta}^P)^T\}^T$. This approach reflects the effect of $\{(\hat{\theta}^L)^T, (\hat{\theta}^P)^T\}^T$ jointly on our new estimator. Let \mathbf{E} be a $2k \times k$ matrix such that

$$\mathbf{E} = \begin{pmatrix} 1, 0, \dots, 0 \\ 0, 1, \dots, 0 \\ \vdots \\ 0, 0, \dots, 1 \\ 1, 0, \dots, 0 \\ 0, 1, \dots, 0 \\ \vdots \\ 0, 0, \dots, 1 \end{pmatrix}.$$

\mathbf{E} is a multivariate extension of \mathbf{h} . Note that \mathbf{E} is concatenation of two $k \times k$ identity matrices by row. Entries that are 1 in these two $k \times k$ identity matrices are source of weights for $\hat{\theta}^L$ and

$\hat{\theta}^p$. The next step is to construct $\hat{\mathbf{B}}$, which is

$$\hat{\mathbf{B}} = \{(\mathbf{E}^T \hat{\Sigma}^{**} \mathbf{E})^{-1} \mathbf{E}^T \hat{\Sigma}^{**}\}^T$$

Then $\hat{\mathbf{B}}$ has the form

$$\begin{pmatrix} \hat{c}_{1,1}^* & \cdots & \hat{c}_{1,k}^* \\ \hat{c}_{2,1}^* & \cdots & \hat{c}_{2,k}^* \\ \vdots & \vdots & \vdots \\ \hat{c}_{(k+1),1}^* & \cdots & \hat{c}_{(k+1),k}^* \\ \vdots & \vdots & \vdots \\ \hat{c}_{2k,1}^* & \cdots & \hat{c}_{2k,k}^* \end{pmatrix}.$$

This matrix is a multivariate extension of \hat{c}_m from the first approach. This matrix is a contrast matrix in the sense that $\hat{c}_{m,m}^* + \hat{c}_{(k+m),m}^* = 1$ for the m th regression coefficient of $\hat{\theta}^L$ and $\hat{\theta}^p$. Moreover, $\hat{c}_{p,p}^* + \hat{c}_{(k+p),p}^* = 0$ for $p \neq m = 1, \dots, k$. Using a vector form, from this approach our new estimator, say $\hat{\theta}^{JWE}$,

$$\hat{\theta}^{JWE} = (\hat{\theta}_1^{JWE}, \dots, \hat{\theta}_k^{JWE})^T = (\hat{c}_{1,1}^* \hat{\theta}_1^L + \hat{c}_{(k+1),1}^* \hat{\theta}_1^p, \dots, \hat{c}_{k,k}^* \hat{\theta}_k^L + \hat{c}_{(2k),k}^* \hat{\theta}_k^p)^T.$$

We can also refer this approach as the ‘joint approach’.

Now the key step is to obtain $\hat{\Sigma}$. We use the resampling approach of [16], which was also used in [13] and [17]. Let $\hat{\alpha}^L = \{\hat{\eta}^T, (\hat{\theta}^L)^T\}^T$ and $\hat{\alpha}^p = \{\hat{\eta}^T, (\hat{\theta}^p)^T\}^T$. From [13] and [17], we have

$$\begin{aligned} \mathbf{W}_i^{(1)} &= \Delta_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}) \geq \tilde{D}_i^*(\hat{\eta})\} \mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}) \geq \tilde{D}_i^*(\hat{\eta})\}} \right] - \sum_{l=1}^n \frac{\Delta_l I\{\tilde{D}_l^*(\hat{\eta}) \geq \tilde{D}_i^*(\hat{\eta})\}}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}) \geq \tilde{D}_i^*(\hat{\eta})\}} \\ &\times \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}) \geq \tilde{D}_i^*(\hat{\eta})\} \mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}) \geq \tilde{D}_i^*(\hat{\eta})\}} \right], \end{aligned}$$

$$\begin{aligned} \mathbf{W}_i^{(2)} &= \tilde{\delta}_i^*(\hat{\alpha}^L) \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\alpha}^L) \geq \tilde{X}_i^*(\hat{\alpha}^L)\} \mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\alpha}^L) \geq \tilde{X}_i^*(\hat{\alpha}^L)\}} \right] - \sum_{l=1}^n \frac{\tilde{\delta}_l^*(\hat{\alpha}^L) I\{\tilde{X}_l^*(\hat{\alpha}^L) \geq \tilde{X}_i^*(\hat{\alpha}^L)\}}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\alpha}^L) \geq \tilde{X}_i^*(\hat{\alpha}^L)\}} \\ &\times \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\alpha}^L) \geq \tilde{X}_i^*(\hat{\alpha}^L)\} \mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\alpha}^L) \geq \tilde{X}_i^*(\hat{\alpha}^L)\}} \right], \end{aligned}$$

and

$$\mathbf{W}_i^{(3)} = \frac{2}{n-1} \sum_{j=1}^n (\mathbf{Z}_i - \mathbf{Z}_j) \phi_{ij}(\hat{\alpha}^p).$$

Define

$$\mathbf{W}_i = \begin{pmatrix} \mathbf{W}_i^{(1)} \\ \mathbf{W}_i^{(2)} \\ \mathbf{W}_i^{(3)} \end{pmatrix}.$$

A consistent estimator of $\mathbf{\Omega}_0$ is

$$\hat{\mathbf{\Omega}} = \frac{1}{n} \sum_{i=1}^n \mathbf{W}_i \mathbf{W}_i^T.$$

We then solve the estimating equation

$$\mathbf{G}_n(\boldsymbol{\beta}) = -n^{-1/2} \sum_{i=1}^n \mathbf{W}_i Q_i, \tag{11}$$

where Q_i ($i = 1, \dots, n$) represent standard normal random variables. Note that $\mathbf{G}_n(\boldsymbol{\beta}) = [\mathbf{S}_n^T(\eta), \{\mathbf{U}_n^L(\boldsymbol{\alpha})\}^T, \{\mathbf{U}_n^P(\boldsymbol{\alpha})\}^T]^T$ is joint estimating equation for $(\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^T, \boldsymbol{\theta}_0^T)^T$. By solving this equation, we obtain many realizations of $\hat{\boldsymbol{\beta}}^R$, say $\hat{\boldsymbol{\beta}}^R = \{(\hat{\eta}^*)^T, (\hat{\boldsymbol{\theta}}^{L*})^T, (\hat{\boldsymbol{\theta}}^{P*})^T\}^T$ where $\{(\hat{\eta}^*)^T, (\hat{\boldsymbol{\theta}}^{L*})^T, (\hat{\boldsymbol{\theta}}^{P*})^T\}^T$ are solutions from (Eq 11). The next theorem, combined with Theorem 2, justifies the resampling approach for calculating $\hat{\Sigma}$.

Theorem 3. Based on the technical conditions in [16], the unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is same asymptotically as the conditional distribution of $n^{1/2}(\hat{\boldsymbol{\beta}}^R - \hat{\boldsymbol{\beta}})$ where $\hat{\boldsymbol{\beta}}^R$ are realizations of $\hat{\boldsymbol{\beta}}$ from resampling.

Proof. Recall that for any $\boldsymbol{\beta}$ in the small neighborhood of $\boldsymbol{\beta}_0$, we have

$$\mathbf{G}_n(\boldsymbol{\beta}) = \mathbf{G}_n(\boldsymbol{\beta}_0) + n^{1/2}\Gamma_0(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o_p(1) \tag{12}$$

Note that $\hat{\boldsymbol{\beta}}^R$ are solutions of Eq (11). By conditioning on observed data and using expansion (Eq 12) as well as by adapting arguments in [13] and [16],

$$\mathbf{G}_n(\hat{\boldsymbol{\beta}}^R) = \mathbf{G}_n(\hat{\boldsymbol{\beta}}) + n^{1/2}\Gamma_0(\hat{\boldsymbol{\beta}}^R - \hat{\boldsymbol{\beta}}) + o_p(1)$$

and hence,

$$n^{1/2}(\hat{\boldsymbol{\beta}}^R - \hat{\boldsymbol{\beta}}) = -\Gamma_0^{-1}n^{-1/2} \sum_{i=1}^n \mathbf{W}_i Q_i + o_p(1)$$

Note that $n^{-1/2} \sum_{i=1}^n \mathbf{W}_i Q_i$ is asymptotically normal with covariance matrix Σ_0 . Then given observed data, distribution of $n^{1/2}(\hat{\boldsymbol{\beta}}^R - \hat{\boldsymbol{\beta}})$ is asymptotically normal with covariance matrix $\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1}$. Hence conditional distribution of $n^{1/2}(\hat{\boldsymbol{\beta}}^R - \hat{\boldsymbol{\beta}})$ on observed data is asymptotically same as unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$.

For $m = 1, \dots, k$ and $j = 1, \dots, M$, let $(\hat{\boldsymbol{\eta}}_m^*)^{(j)}$, $(\hat{\boldsymbol{\theta}}_m^{L*})^{(j)}$ and $(\hat{\boldsymbol{\theta}}_m^{P*})^{(j)}$ be j th realizations of an element $\hat{\boldsymbol{\eta}}_m$, $\hat{\boldsymbol{\theta}}_m^L$ and $\hat{\boldsymbol{\theta}}_m^P$ corresponding to m th covariate, respectively. The algorithm for the first approach is as follows.

1. By resampling, calculate the covariance matrix $\hat{\Sigma}$ using realizations $(\hat{\boldsymbol{\eta}}_m^*)^{(j)}$, $(\hat{\boldsymbol{\theta}}_m^{L*})^{(j)}$ and $(\hat{\boldsymbol{\theta}}_m^{P*})^{(j)}$, ($m = 1, \dots, k$ and $j = 1, \dots, M$).

2. From $\hat{\Sigma}^{-1}$, obtain the covariance matrix corresponding to $\hat{\theta}_m^L$ and $\hat{\theta}_m^P$, say $\hat{\Sigma}_m^*$.
3. Calculate $\hat{c}_m = (\hat{c}_{m1}, \hat{c}_{m2})^T = (\mathbf{h}^T \hat{\Sigma}_m^* \mathbf{h})^{-1} \hat{\Sigma}_m^* \mathbf{h}$ where $\mathbf{h} = (1, 1)^T$ and obtain the new estimate $\hat{\theta}_m^{MWE} = \hat{c}_{m1} \hat{\theta}_m^L + \hat{c}_{m2} \hat{\theta}_m^P$.
4. Repeat step 3 for all covariates.

The algorithm for the second approach is as follows.

1. By resampling, calculate the covariance matrix $\hat{\Sigma}$ using realizations $(\hat{\eta}_m^*)^{(j)}$, $(\hat{\theta}_m^{L*})^{(j)}$ and $(\hat{\theta}_m^{P*})^{(j)}$ ($m = 1, \dots, k$ and $j = 1, \dots, M$).
2. Obtain $\hat{\Sigma}^{**}$ from $\hat{\Sigma}$.
3. From $\hat{\Sigma}^{**}$ and \mathbf{E} , obtain $\hat{\mathbf{B}}$.
4. Calculate the new estimate $\hat{\theta}_m^{JWE} = \hat{c}_{m,m}^* \hat{\theta}_m^L + \hat{c}_{k+m,m}^* \hat{\theta}_m^P$, where $\hat{c}_{j,l}^*$ be the element of j th row and l th column of $\hat{\mathbf{B}}$.

By Theorem 1 and Theorem 2, our new estimators are consistent and asymptotically normal.

Model checking

For assessing the adequacy of the model, since our weight estimator is based on estimators from [13] and [17], it is reasonable to consider entire processes from [13] and [17]. In this case, we extend model checking technique from [13]. As defined in [13], Let $N_{1i}(t; \boldsymbol{\eta}) = \Delta_i I\{\tilde{D}_i^*(\boldsymbol{\eta}) \leq t\}$ and $N_{2i}(t; \boldsymbol{\alpha}) = \tilde{\delta}_i^*(\boldsymbol{\alpha}) I\{\tilde{X}_i^*(\boldsymbol{\alpha}) \leq t\}$, where $i = 1, \dots, n$. Then Nelson-Aalen estimators for the event of interest and dependent censoring are

$$\hat{\Lambda}_0^{(1)}(u; \boldsymbol{\eta}) = \int_{-\infty}^t \frac{\sum_{i=1}^n dN_{1i}(u; \boldsymbol{\eta})}{\sum_{j=1}^n I\{\tilde{D}_j^*(\boldsymbol{\eta}) \geq u\}} \quad \hat{\Lambda}_0^{(2)}(u; \boldsymbol{\alpha}) = \int_{-\infty}^t \frac{\sum_{i=1}^n dN_{2i}(u; \boldsymbol{\alpha})}{\sum_{j=1}^n I\{\tilde{X}_j^*(\boldsymbol{\alpha}) \geq u\}}$$

Note that by (Eq 3) and (Eq 4), martingale residuals are defined as

$$\hat{M}_{1i}(t; \hat{\boldsymbol{\eta}}) = N_{1i}(t; \hat{\boldsymbol{\eta}}) - \int_{-\infty}^t I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \geq u\} d\hat{\Lambda}_0^{(1)}(u, \hat{\boldsymbol{\eta}})$$

$$\hat{M}_{2i}(t; \hat{\boldsymbol{\alpha}}) = N_{2i}(t; \hat{\boldsymbol{\alpha}}) - \int_{-\infty}^t I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}) \geq u\} d\hat{\Lambda}_0^{(2)}(u, \hat{\boldsymbol{\alpha}}),$$

where $\hat{\boldsymbol{\alpha}}$ can be either $\hat{\boldsymbol{\alpha}}^L = \{\hat{\eta}^T, (\hat{\theta}^L)^T\}^T$ or $\hat{\boldsymbol{\alpha}}^P = \{\hat{\eta}^T, (\hat{\theta}^P)^T\}^T$. Then as defined in [13],

$$\mathbf{S}_n(s; \boldsymbol{\eta}) = n^{-1/2} \sum_{i=1}^n \mathbf{Z}_i \hat{M}_{1i}(s; \boldsymbol{\eta}) \quad \mathbf{U}_n(t; \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^n \mathbf{Z}_i \hat{M}_{2i}(t; \boldsymbol{\alpha}).$$

Then similar to [13] and [17], we can substitute $\hat{\boldsymbol{\eta}}$ on $\mathbf{S}_n(s; \boldsymbol{\eta})$, $\hat{\boldsymbol{\alpha}}^L$ and $\hat{\boldsymbol{\alpha}}^P$ on $\mathbf{U}_n(t; \boldsymbol{\alpha})$. $[\mathbf{S}_n^T(s; \hat{\boldsymbol{\eta}}), \{\mathbf{U}_n(t; \hat{\boldsymbol{\alpha}}^L)\}^T, \{\mathbf{U}_n(t; \hat{\boldsymbol{\alpha}}^P)\}^T]^T$ are called observed score processes with respect to dependent censoring and the event of interest, respectively [7], [13], [17]. We can construct

$[\hat{S}_n^T(s; \hat{\eta}^*), \{\hat{U}_n^L(t; \hat{\alpha}^{L*})\}^T, \{\hat{U}_n^P(v; \hat{\alpha}^{P*})\}^T]^T$ [13], [17], where

$$\hat{S}_n(s; \hat{\eta}^*) = n^{-1/2} \sum_{i=1}^n \int_{-\infty}^s \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}) \geq w\} \mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}) \geq w\}} \right] d\hat{M}_{1i}(w; \hat{\eta}) Q_i + \mathbf{S}_n(s; \hat{\eta}^*) - \mathbf{S}_n(s; \hat{\eta})$$

$$\hat{U}_n^L(t; \hat{\alpha}^{L*}) = n^{-1/2} \sum_{i=1}^n \int_{-\infty}^t \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\alpha}^L) \geq w\} \mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\alpha}^L) \geq w\}} \right] d\hat{M}_{2i}(w; \hat{\alpha}^L) Q_i + \mathbf{U}_n(t; \hat{\alpha}^{L*}) - \mathbf{U}_n(t; \hat{\alpha}^L)$$

$$\hat{U}_n^P(v; \hat{\alpha}^{P*}) = n^{-1/2} \sum_{i=1}^n \int_{-\infty}^v \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\alpha}^P) \geq w\} \mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\alpha}^P) \geq w\}} \right] d\hat{M}_{2i}(w; \hat{\alpha}^P) Q_i + \mathbf{U}_n(v; \hat{\alpha}^{P*}) - \mathbf{U}_n(v; \hat{\alpha}^P),$$

where $\hat{\alpha}^{L*} = \{(\hat{\eta}^*)^T, (\hat{\theta}^{L*})^T\}^T$ and $\hat{\alpha}^{P*} = \{(\hat{\eta}^*)^T, (\hat{\theta}^{P*})^T\}^T$. These three processes are called bootstrapped processes [7], [13], [17]. We can plot the observed process with bootstrapped processes by randomly selecting 20 or 30 observations. Standard tests for goodness of fit can be performed by calculating Kolmogorov-Smirnov type test statistics. Test statistics are then defined by $\sup_s \|\mathbf{S}_n(s; \hat{\eta})\|$, $\sup_t \|\mathbf{U}_n(t; \hat{\alpha}^L)\|$, and $\sup_v \|\mathbf{U}_n(v; \hat{\alpha}^P)\|$. To calculate the null distribution of the test statistics, first we obtain j th realizations of bootstrap samples $(\hat{\eta}^*)^{(j)}$, $(\hat{\theta}^{L*})^{(j)}$ and $(\hat{\theta}^{P*})^{(j)}$. Then we compute $BS_j = \sup_s \|\hat{S}_n(s; (\hat{\eta}^*)^{(j)})\|$, $BS_j^L = \sup_t \|\hat{U}_n^L(t; (\hat{\alpha}^{L*})^{(j)})\|$ and $BS_j^P = \sup_v \|\hat{U}_n^P(v; (\hat{\alpha}^{P*})^{(j)})\|$, respectively for $j = 1, \dots, M$, where $(\hat{\alpha}^{L*})^{(j)}$ and $(\hat{\alpha}^{P*})^{(j)}$ are j th realizations of bootstrap samples of $\hat{\alpha}^{L*}$ and $\hat{\alpha}^{P*}$. The p-values can be defined by

$$p_1 = \frac{1}{M} \sum_{j=1}^M I\{BS_j \geq \sup_s \|\mathbf{S}_n(s; \hat{\eta})\|\}$$

$$p_2 = \frac{1}{M} \sum_{j=1}^M I\{BS_j^L \geq \sup_t \|\mathbf{U}_n(t; \hat{\alpha}^L)\|\}$$

$$p_3 = \frac{1}{M} \sum_{j=1}^M I\{BS_j^P \geq \sup_v \|\mathbf{U}_n(v; \hat{\alpha}^P)\|\}.$$

[10]. If the p-value is smaller than predetermined level, we reject the null hypothesis, which means that data does not have appropriate fit on our bivariate model. Note that a multiple testing problem arises for testing the models for θ . We address this by adjusting p-values based on a Bonferroni correction with two tests.

Simulation Studies

We consider two simulation settings. In first simulation setting, the errors follow a bivariate normal distribution with mean (0,1.2) with variance 1 and correlation $\rho = 0,0.25$. The independent censoring time C is generated from $\log(U^*)$, where U^* has uniform distribution with minimum value 0 and maximum value 20. Covariate is $Z \sim \text{Bernoulli}(0.5)$, where $\text{Bernoulli}(0.5)$ is Bernoulli distribution with success probability 0.5. We run 500 simulation runs. Within each simulation run, 500 resampling runs are tried for covariance matrix calculation. Sample sizes are $N = 150$ and $N = 300$. If there is only one covariate in the model, the first and the second

Table 1. Simulation result when N = 150 and N = 300, ρ = 0 with covariate Bernoulli(0.5).

Estimators ¹	N = 150			
	Bias (Dmedian ²)	MSE ³ (Mediansq ⁴)	SEE ⁵ (Sdmedian ⁶)	Coverage ⁷
$\hat{\theta}^L$	0.018 (0.018)	0.04 (0.017)	0.204 (0.2)	0.95
$\hat{\theta}^P$	0.021 (0.014)	0.036 (0.017)	0.193 (0.19)	0.96
$\hat{\theta}^{WE}$	0.016 (0.006)	0.036 (0.015)	0.188 (0.185)	0.95
Estimators ¹	N = 300			
	Bias (Dmedian ²)	MSE ³ (Mediansq ⁴)	SEE ⁵ (Sdmedian ⁶)	Coverage ⁷
$\hat{\theta}^L$	-0.002 (-0.003)	0.017 (0.006)	0.140 (0.140)	0.95
$\hat{\theta}^P$	-0.001 (0.002)	0.016 (0.007)	0.133 (0.132)	0.95
$\hat{\theta}^{WE}$	-0.004 (-0.002)	0.016 (0.007)	0.130 (0.129)	0.94

¹ $\hat{\theta}^L$: the estimator by [13]; $\hat{\theta}^P$: the estimator by [17]; $\hat{\theta}^{WE}$: the weighted estimator by the proposed approach (Note that the marginal approach and the joint approach are equal in one variable case)

² median of difference of the estimator from true value

³ mean squared error

⁴ median of squared error

⁵ mean of standard error

⁶ median of standard error

⁷ 95% coverage rate

doi:10.1371/journal.pone.0124381.t001

method of the weighted estimation are equivalent. Let this common weighed estimator be $\hat{\theta}^{WE}$. We calculate bias (Bias), mean squared error (MSE), mean of standard error (SEE), 95% coverage rate (Coverage). The coverage is based on the normal approximation. Moreover, to evaluate robustness of estimators, we also compute median of difference of the estimator from true value (Dmedian), median of squared error of estimates (Mediansq), and median of standard errors (Sdmedian). Results are summarized on Table 1 and Table 2.

In second simulation setting, we generate Gamma random variable v with mean $\mu = 1$ and variance $\sigma^2 = 0$ or 1, then create $W = \exp(\epsilon^X)$, which is an exponential random variable with rate $4v^{-1}$ and $\exp(\epsilon^D)$ with an exponential random variable with rate v^{-1} . Then we generate time to the event of interest by $\exp(X) = \exp(\theta_0^T \mathbf{Z}) \exp(\epsilon^X)$ and time to the dependent censoring by $\exp(D) = \exp(\eta_0^T \mathbf{Z}) \exp(\epsilon^D)$ (By notation in our paper, X, D and C are already log-transformed times. Thus in this context, $\exp(X), \exp(D)$ and $\exp(C)$ are times in the original scale). The independent censoring time $\exp(C)$ has uniform distribution with minimum value 0 and maximum value 20. True parameter values are $\theta_0 = (0.5, 1)^T$ and $\eta_0 = (1, 0.5)^T$ and covariates are $Z_1 \sim U(0, 1)$, where $U(0, 1)$ is uniform distribution with minimum value 0 and maximum value 1 and $Z_2 \sim Bernoulli(0.5)$. We run 500 simulation runs. Within each simulation run, 500 resampling runs are tried for covariance matrix calculation. Let $\hat{\theta}^{MWE}$ be weighted estimators from calculating weights marginally (the first proposed method) and let $\hat{\theta}^{JWE}$ be weighted estimators from calculating weights jointly (the second proposed method). We compute the same quantities as we did in the first set of the simulation study. Results are summarized on Table 3 and Table 4.

In these simulation results, we can see that our weighted estimators have good results. In both cases, bias and mean squared error of our new estimator has similar performance compared to the estimators by [13] and [17]. Mean of standard errors and median of standard

Table 2. Simulation result when N = 150 and N = 300, $\rho = 0.25$ with covariate Bernoulli(0.5).

Estimators ¹	N = 150			
	Bias (Dmedian ²)	MSE ³ (Mediansq ⁴)	SEE ⁵ (Sdmedian ⁶)	Coverage ⁷
$\hat{\theta}^L$	0.005 (0.01)	0.036 (0.017)	0.198 (0.197)	0.95
$\hat{\theta}^P$	0.006 (0.007)	0.032 (0.015)	0.189 (0.188)	0.95
$\hat{\theta}^{WE}$	-0.001 (-0.006)	0.033 (0.016)	0.184 (0.183)	0.94
Estimators	N = 300			
	Bias (Dmedian ²)	MSE ³ (Mediansq ⁴)	SEE ⁵ (Sdmedian ⁶)	Coverage ⁷
$\hat{\theta}^L$	-0.003 (0.005)	0.018 (0.008)	0.138 (0.137)	0.95
$\hat{\theta}^P$	0.001 (0.007)	0.017 (0.007)	0.131 (0.131)	0.95
$\hat{\theta}^{WE}$	-0.003 (0.002)	0.017 (0.007)	0.129 (0.128)	0.95

¹ $\hat{\theta}^L$: the estimator by [13]; $\hat{\theta}^P$: the estimator by [17]; $\hat{\theta}^{WE}$: the weighted estimator by the proposed approach (Note that the marginal approach and the joint approach are equal in one variable case)

² median of difference of the estimator from true value

³ mean squared error

⁴ median of squared error

⁵ mean of standard error

⁶ median of standard error

⁷ 95% coverage rate

doi:10.1371/journal.pone.0124381.t002

errors are smaller than the estimators by [13] and [17]. Moreover, computation results for the median of difference of the estimators from true value and the median of squared error imply that our proposed estimator is comparable with the estimators from the original methods.

In the first simulation setting, the difference of standard error between our proposed estimator and $\hat{\theta}^L$ is bigger than the one between $\hat{\theta}^P$ and the proposed estimator. In the second simulation setting, the phenomenon is the opposite. Furthermore, in the first simulation setting, $\hat{\theta}^P$ has lower standard error on average than one of $\hat{\theta}^L$ while $\hat{\theta}^L$ have better efficiency (with respect to standard error) than ones by $\hat{\theta}^P$ in the second simulation setting. This simulation result verifies our claim, which means that neither estimator is better than another. Our proposed estimator takes advantage of smaller standard error with achieving small bias and correct coverage except $N = 150$ with $\sigma^2 = 1$ in the second simulation setting. In this scenario, empirical coverage of proposed estimators is lower than nominal 95% coverage. This is due to low coverage of $\hat{\theta}^L$. Since we combine $\hat{\theta}^L$ and $\hat{\theta}^P$, if one of them has low coverage, it is highly likely that the coverage of weighted estimator may also be below the nominal coverage.

Real data analysis

We applied our method to data from the AIDS Clinical Trial Group (ACTG) Study 364 [1], which was used in [17]. This multicenter randomized study investigated patients whose plasma RNA level is at least 500 copies per ml. Subjects were assigned to three treatments, nelfinavir (NFV), efavirenz (EFV), and combination of nelfinavir and efavirenz (NFV + EFV). Details about this study can be found in [1].

The two failure times are time to HIV RNA level greater than 2000 copies per ml and time to withdrawal of study. Let X be the first time when HIV RNA level is greater than 2000 copies

Table 3. Simulation result when N = 150 and N = 300, $\sigma^2 = 0$ with two covariates (Z_1 : U(0, 1), Z_2 : Bernoulli(0.5)).

Estimators ¹	N = 150							
	Bias (Dmedian ²)		MSE ³ (Mediansq ⁴)		SEE ⁵ (Sdmedian ⁶)		Coverage ⁷	
	Z ₁	Z ₂	Z ₁	Z ₂	Z ₁	Z ₂	Z ₁	Z ₂
$\hat{\theta}^L$	0.0001 (0.002)	0.002 (-0.005)	0.12 (0.052)	0.042 (0.018)	0.358 (0.352)	0.226 (0.222)	0.96	0.96
$\hat{\theta}^P$	-0.003 (0.003)	-0.003 (-0.002)	0.158 (0.074)	0.051 (0.023)	0.427 (0.418)	0.243 (0.241)	0.96	0.96
$\hat{\theta}^{MWE}$	0.003 (-0.007)	0.003 (0.003)	0.123 (0.053)	0.043 (0.019)	0.351 (0.349)	0.219 (0.218)	0.95	0.95
$\hat{\theta}^{JWE}$	0.003 (-0.007)	0.004 (0.001)	0.123 (0.055)	0.043 (0.018)	0.351 (0.349)	0.219 (0.217)	0.94	0.95
Estimators ¹	N = 300							
	Bias (Dmedian ²)		MSE ³ (Mediansq ⁴)		SEE ⁵ (Sdmedian ⁶)		Coverage ⁷	
	Z ₁	Z ₂	Z ₁	Z ₂	Z ₁	Z ₂	Z ₁	Z ₂
$\hat{\theta}^L$	-0.012 (-0.013)	0.004 (0.001)	0.065 (0.028)	0.02 (0.01)	0.257 (0.255)	0.148 (0.148)	0.95	0.96
$\hat{\theta}^P$	-0.014 (-0.017)	-0.001 (-0.015)	0.081 (0.035)	0.024 (0.013)	0.283 (0.281)	0.162 (0.162)	0.95	0.96
$\hat{\theta}^{MWE}$	-0.01 (-0.012)	0.003 (-0.001)	0.064 (0.031)	0.02 (0.01)	0.252 (0.251)	0.146 (0.146)	0.95	0.96
$\hat{\theta}^{JWE}$	-0.01 (-0.014)	0.003 (0.002)	0.064 (0.032)	0.02 (0.01)	0.251 (0.25)	0.146 (0.146)	0.95	0.96

¹ $\hat{\theta}^L$: the estimator by [13]; $\hat{\theta}^P$: the estimator by [17]; $\hat{\theta}^{MWE}$: the weighted estimator by the marginal approach; $\hat{\theta}^{JWE}$: the weighted estimator by the joint approach

² median of difference of the estimator from true value

³ mean squared error

⁴ median of squared error

⁵ mean of standard error

⁶ median of standard error

⁷ 95% coverage rate

doi:10.1371/journal.pone.0124381.t003

per ml and D be time to withdrawal of study. We considered four covariates and 194 observations. Z_1 takes value 1 if a patient receives EFV and 0 otherwise. Z_2 takes value 1 if a patient receives NFV + EFV and 0 otherwise. Z_3 is New3TC, which takes value 1 if lamivudine is given as a new nucleoside analogue therapy to a patient and 0 otherwise. Z_4 is logarithm of RNA level at the start of the study.

Table 5 and Table 6 show the point estimates and standard errors of $\hat{\eta}$, $\hat{\theta}^L$, $\hat{\theta}^P$, $\hat{\theta}^{MWE}$ and $\hat{\theta}^{JWE}$. Our method works well for the models with and without New3TC on all covariates. Some variables are seen to be statistically significant based on the weighted estimator while they are not by [13] or [17]. For example, let's consider effect of EFV to the time to first virologic failure. By Table 6, the estimated effect by using approach of [13] is 0.475 and its standard error is 0.250. From approach of [17], an estimate is 0.464 and its standard error is 0.281. Based on the fact that estimators are asymptotic normal, from Wald test using [13] and [17], EFV is not a statistically significant variable on 5% significant level. On the other hand, a weighted estimate

Table 4. Simulation result when N = 150 and N = 300, $\sigma^2 = 1$ with two covariates (Z_1 : U(0,1), Z_2 : Bernoulli(0.5)).

Estimators ¹	N = 150							
	Bias (Dmedian ²)		MSE ³ (Mediansq ⁴)		SEE ⁵ (Sdmedian ⁶)		Coverage ⁷	
	Z ₁	Z ₂	Z ₁	Z ₂	Z ₁	Z ₂	Z ₁	Z ₂
$\hat{\theta}^L$	-0.010 (-0.003)	-0.033 (-0.039)	0.273 (0.127)	0.086 (0.038)	0.441 (0.434)	0.315 (0.312)	0.90	0.96
$\hat{\theta}^P$	0.002 (-0.002)	-0.032 (-0.046)	0.295 (0.142)	0.095 (0.038)	0.559 (0.552)	0.325 (0.324)	0.95	0.96
$\hat{\theta}^{MWE}$	-0.009 (-0.008)	-0.030 (-0.031)	0.263 (0.128)	0.085 (0.041)	0.437 (0.432)	0.303 (0.301)	0.90	0.96
$\hat{\theta}^{JWE}$	-0.009 (-0.008)	-0.030 (-0.03)	0.262 (0.128)	0.086 (0.04)	0.437 (0.432)	0.303 (0.301)	0.90	0.96
Estimators ¹	N = 300							
	Bias (Dmedian ²)		MSE ³ (Mediansq ⁴)		SEE ⁵ (Sdmedian ⁶)		Coverage ⁷	
	Z ₁	Z ₂	Z ₁	Z ₂	Z ₁	Z ₂	Z ₁	Z ₂
$\hat{\theta}^L$	-0.024 (-0.038)	0.003 (-0.006)	0.133 (0.057)	0.04 (0.019)	0.345 (0.344)	0.211 (0.21)	0.94	0.96
$\hat{\theta}^P$	-0.016 (-0.016)	0.012 (0.003)	0.148 (0.059)	0.045 (0.02)	0.384 (0.382)	0.222 (0.222)	0.96	0.97
$\hat{\theta}^{MWE}$	-0.024 (-0.035)	0.007 (-0.003)	0.134 (0.058)	0.04 (0.019)	0.341 (0.341)	0.207 (0.207)	0.94	0.96
$\hat{\theta}^{JWE}$	-0.025 (-0.035)	0.007 (-0.002)	0.135 (0.058)	0.039 (0.018)	0.341 (0.341)	0.206 (0.207)	0.94	0.96

¹ $\hat{\theta}^L$: the estimator by [13]; $\hat{\theta}^P$: the estimator by [17]; $\hat{\theta}^{MWE}$: the weighted estimator by the marginal approach; $\hat{\theta}^{JWE}$: the weighted estimator by the joint approach

² median of difference of the estimator from true value

³ mean squared error

⁴ median of squared error

⁵ mean of standard error

⁶ median of standard error

⁷ 95% coverage rate

doi:10.1371/journal.pone.0124381.t004

Table 5. Point estimates with standard errors of covariates in AIDS study for model without New3TC (Standard errors are shown in parenthesis).

Covariates	$\hat{\eta}$	$\hat{\theta}^L$	$\hat{\theta}^P$	$\hat{\theta}^{MWE}$	$\hat{\theta}^{JWE}$
EFV ¹	0.753 (0.339)	0.115 (0.219)	0.375 (0.269)	0.168 (0.206)	0.2 (0.205)
NFV ² + EFV	0.674 (0.255)	1.128 (0.239)	1.091 (0.309)	1.120 (0.222)	1.114 (0.222)
log(RNA) ³	-0.544 (0.154)	-0.464 (0.215)	-0.531 (0.169)	-0.507 (0.163)	-0.511 (0.162)

¹ efavirenz

² nelfinavir

³ logarithm of RNA at the start of the study

doi:10.1371/journal.pone.0124381.t005

Table 6. Point estimates with standard errors of covariates in AIDS study for model with New3TC (Standard errors are shown in parenthesis).

Covariates	$\hat{\eta}$	$\hat{\theta}^L$	$\hat{\theta}^P$	$\hat{\theta}^{MWE}$	$\hat{\theta}^{JWE}$
EFV ¹	0.770 (0.278)	0.475 (0.250)	0.464 (0.281)	0.471 (0.222)	0.471 (0.222)
NFV ² + EFV	0.650 (0.260)	1.353 (0.277)	1.246 (0.338)	1.333 (0.263)	1.317 (0.261)
New3TC ³	0.927 (0.355)	1.449 (0.296)	1.374 (0.328)	1.431 (0.267)	1.420 (0.261)
log(RNA) ⁴	-0.631 (0.183)	-0.654 (0.289)	-0.661 (0.218)	-0.659 (0.216)	-0.660 (0.215)

¹ efavirenz

² nelfinavir

³ lamivudine as new nucleoside analogue therapy

⁴ logarithm of RNA at the start of the study

doi:10.1371/journal.pone.0124381.t006

using first approach is 0.471 and its standard error is 0.222. In this case, EFV is a statistically significant variable on 5% significant level.

Observed score process with bootstrapped processes for withdrawal of study with respect to Z_1 is shown in Fig 1. Fig 2 and Fig 3 show observed score processes and bootstrapped processes of the first virologic failure using $\hat{\alpha}^L$, $\hat{\alpha}^P$ with respect to Z_1 . These three plots are based on the model without New3TC. They are fluctuating around zero, so it seems that there is no graphical evidence for lack of fit. The p-value for the lack of fit tests of withdrawal is 0.952 and the first virologic failure using $\hat{\alpha}^L$ and $\hat{\alpha}^P$ are 0.918 and 0.959 respectively. With graphical checking, p-value indicates that there is no evidence for violation of the model assumption.

For purposes of interpretation, since D represents a standard survival time, the interpretation of $\hat{\eta}$ is in terms of covariate effect for survival time. However, since the observed time for X depends on D , interpretation of $\hat{\theta}$ is difficult. One way to interpret $\hat{\theta}$ is to assume that D does not exist and interpret the effect of $\hat{\theta}$ on X only. This approach is possible if there exists a reasonable extrapolation mechanism for X [18]. However, considering the estimation structure for θ , it is difficult to separate effect of $\hat{\theta}$ to X from effect of $\hat{\eta}$ to D .

Discussion

In this paper, we have proposed optimal estimators using combinations of the two estimators from [13] and [17]. Our methodology can be extended to a case of recurrent event with dependent censoring, which is extensively studied [6], [7], [10]. We are currently working on this extension.

Optimality of the estimator has been discussed in other contexts. Recently, there is a publication that proposed optimal additive functions based on score functions [14]. The main point of their method is to combine unbiased estimating functions. In our case, this would be combining estimating equations and new solution can be obtained by this estimating equation. Comparing performance of this solution and our proposed estimator is of interest. This will be left open to future research.

Another way of achieving optimality is to use generalized method of moment estimator [8]. This estimator is a linear combination of estimating functions [19]. In this case, the estimating functions have a greater dimension than the dimension of the parameter vector. The optimality is achieved by the linear combination. It is shown that the estimator from this linear combination of estimating functions is consistent and asymptotically normal [8]. In the literature of statistics, this idea is applied to generalized estimating equations [19]. The estimating functions

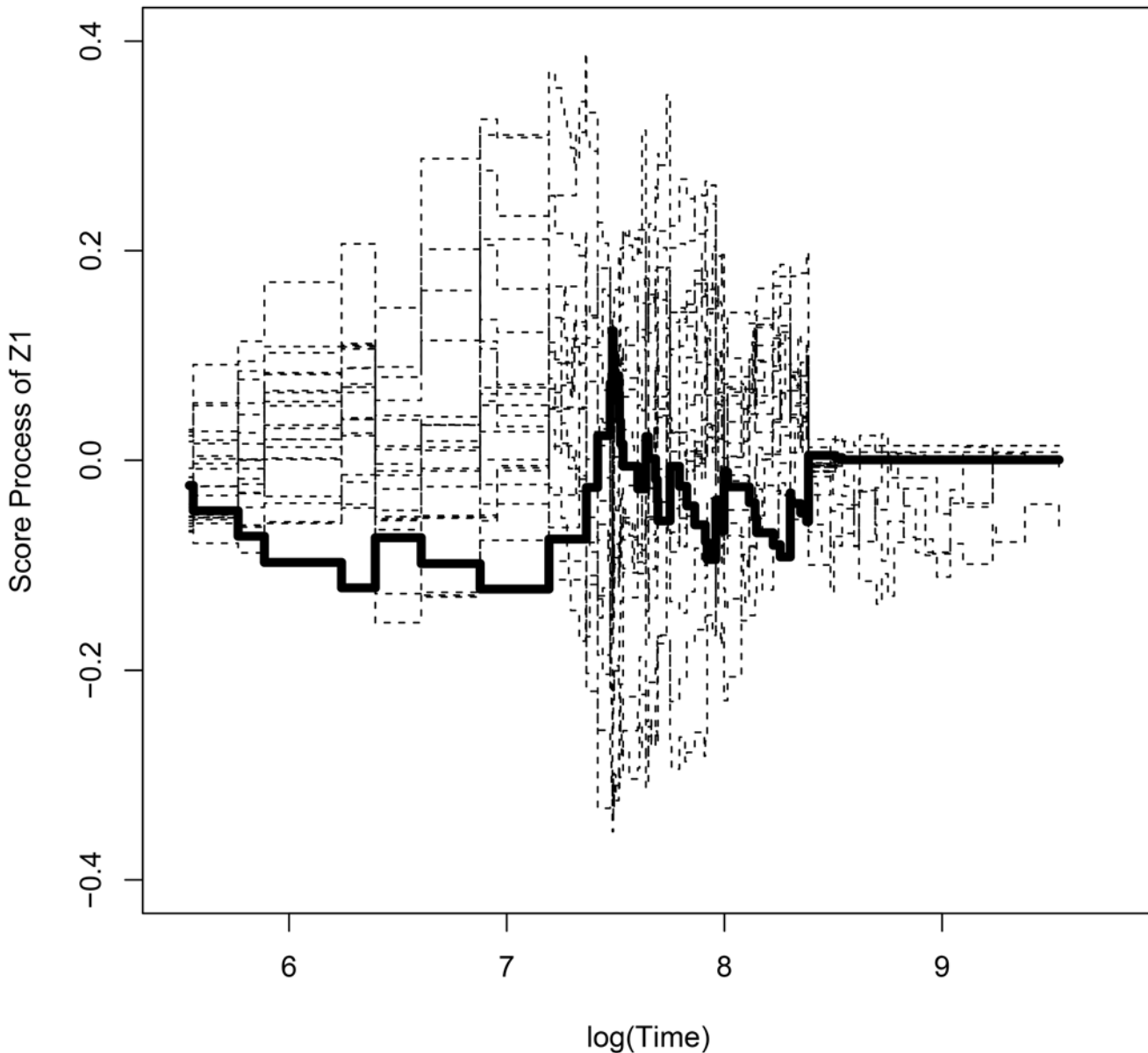


Fig 1. Plot of observed score process and bootstrapped processes of time to withdrawal of study with respect to Z_1 . The thickline is observed process and the dashed lines are bootstrapped processes.

doi:10.1371/journal.pone.0124381.g001

proposed by [19] are called quadratic inference function. Recently, the quadratic inference function is applied to Cox model [26].

[8] and [19] derived new estimating functions, while we combined two estimators directly. This idea of the generalized method of moments is very appealing, but the estimating functions of [13] and [17] are nonsmooth. Finding derivative for the linear combination of the estimating functions, which is a key in generalized method moments, is challenging for our work because we cannot find the derivatives in the estimating functions proposed by [13] and [17]. Applying the idea of [8] to AFT model will be interesting future research.

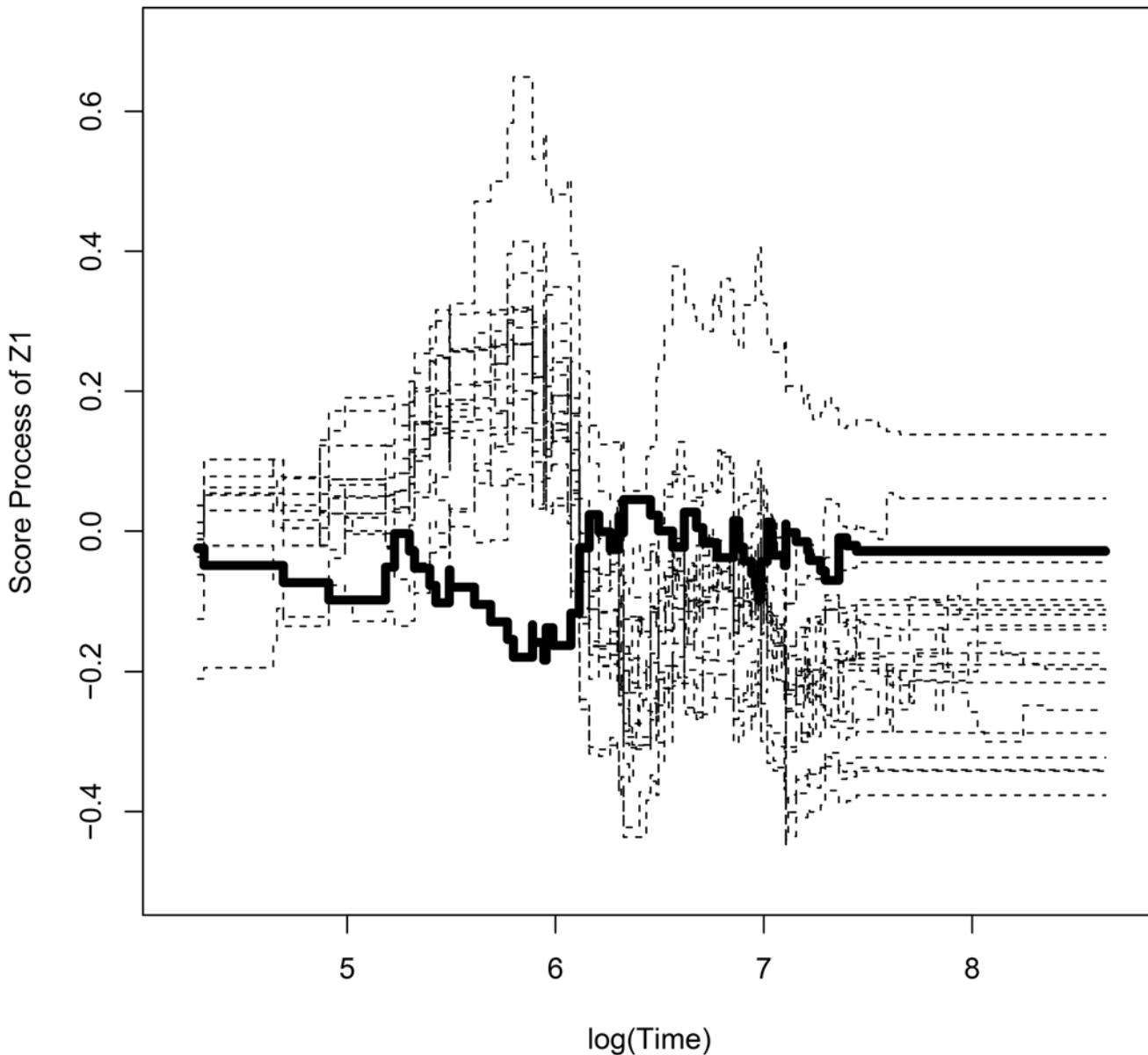


Fig 2. Plot of observed score process and bootstrapped processes of time to first virologic failure using $\hat{\theta}^L$ with respect to Z_1 . The thickline is observed process and the dashed lines are bootstrapped processes.

doi:10.1371/journal.pone.0124381.g002

Our estimating equations to obtain estimators involve nonsmooth functions of η and α . Many literatures used a linear programming approach for estimating θ [3], [11]. However, this linear programming method is very slow for computing estimators of θ . Thus this approach is very inefficient when implementing to solve (Eq 11) for estimation of Σ . Recently, an approach called a derivative free-spectral algorithm for nonlinear equations (DF-SANE) was proposed [12], and there is a publication that showed that this algorithm is better than the linear programming method using an example of estimating parameters of AFT models under independent censoring. [21]. However, under dependent censoring, the artificial censoring term leads

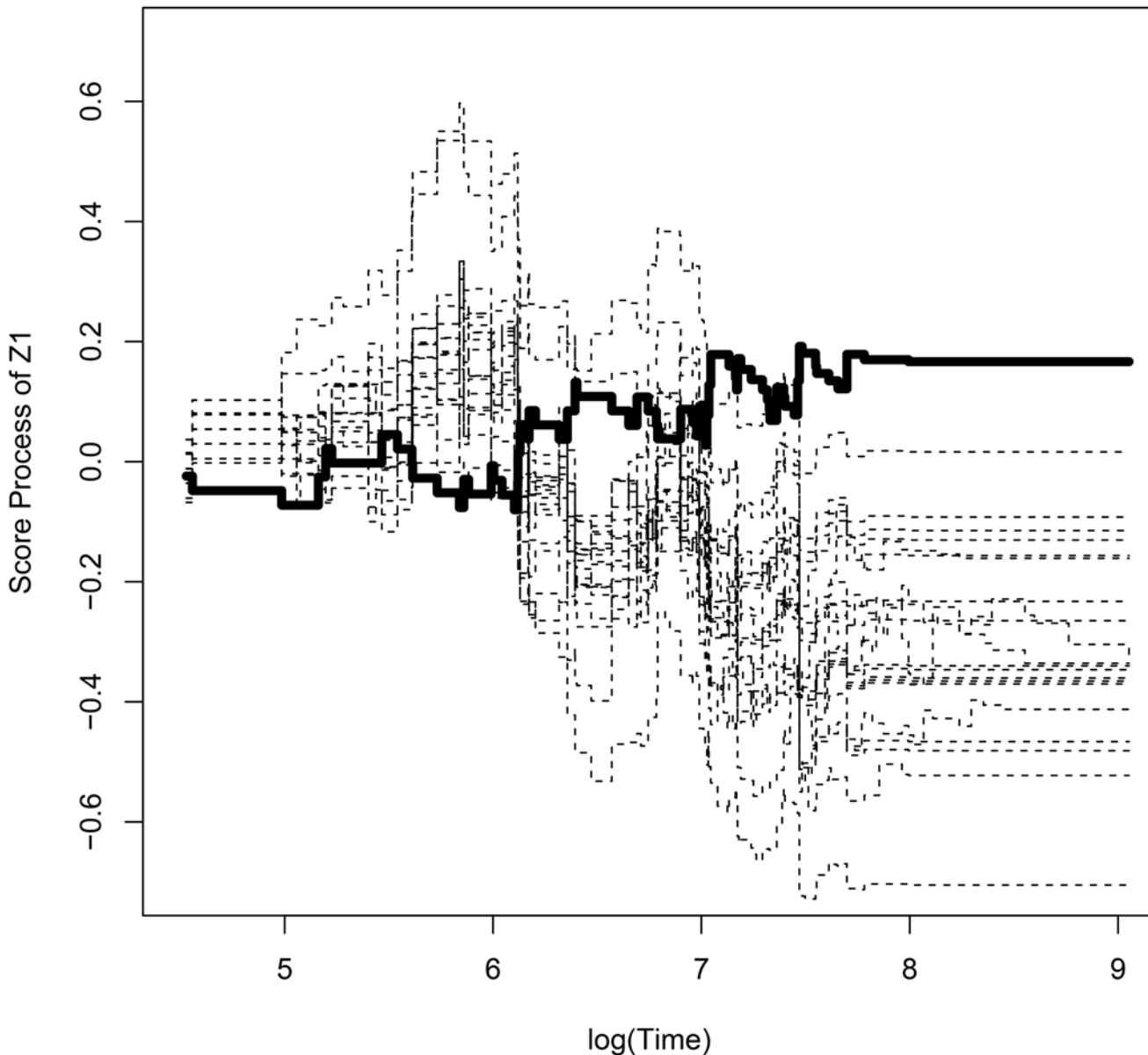


Fig 3. Plot of observed score process and bootstrapped processes of time to first virologic failure using $\hat{\alpha}^p$ with respect to Z_1 . The thickline is observed process and the dashed lines are bootstrapped processes.

doi:10.1371/journal.pone.0124381.g003

to numerical instability in estimating parameters and calculating resampled estimators. Moreover, this algorithm does not converge well under default tolerance settings using DF-SANE [21]. Thus using this algorithm requires changing the tolerance level. Developing efficient numerical algorithms for estimating parameters is an important topic for future research.

Author Contributions

Conceived and designed the experiments: DG. Performed the experiments: YC DG. Analyzed the data: YC. Contributed reagents/materials/analysis tools: YC DG. Wrote the paper: YC DG.

References

1. Albrecht MA, Bosch RJ, Hammer SM, Liou S-H, Kessler H, Para MF, et al. Nelfinavir, Efavirenz, or both after the failure of Nucleoside treatment of HIV infection. *N Engl J Med*. 2001; 345: 398–407. doi: [10.1056/NEJM200108093450602](https://doi.org/10.1056/NEJM200108093450602) PMID: [11496850](https://pubmed.ncbi.nlm.nih.gov/11496850/)
2. Day R, Bryant J, Lefkopoulou M. Adaptation of bivariate frailty models for prediction, with application to biological markers as prognostic indicators. *Biometrika*. 1997; 84: 45–56. doi: [10.1093/biomet/84.1.45](https://doi.org/10.1093/biomet/84.1.45)
3. Ding AA, Shi G, Wang W, Hsieh J-J. Marginal regression analysis for semi-competing risks data under dependent censoring. *Scand J Stat*. 2009; 36: 481–500. doi: [10.1111/j.1467-9469.2008.00635.x](https://doi.org/10.1111/j.1467-9469.2008.00635.x)
4. Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001; 88: 907–919. doi: [10.1093/biomet/88.4.907](https://doi.org/10.1093/biomet/88.4.907)
5. Fleming TR, Harrington DP. *Counting Processes and Survival Analysis*. 2nd ed. New York: Wiley; 2005.
6. Ghosh D, Lin DY. Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics*. 2003; 59: 877–885. doi: [10.1111/j.0006-341X.2003.00102.x](https://doi.org/10.1111/j.0006-341X.2003.00102.x) PMID: [14969466](https://pubmed.ncbi.nlm.nih.gov/14969466/)
7. Ghosh D. Semiparametric analysis of recurrent events: artificial censoring, truncation, pairwise estimation and inference. *Lifetime Data Anal*. 2010; 16: 509–524. doi: [10.1007/s10985-009-9150-4](https://doi.org/10.1007/s10985-009-9150-4) PMID: [20063182](https://pubmed.ncbi.nlm.nih.gov/20063182/)
8. Hansen LP. Large sample properties of generalized method of moments estimators. *Econometrica*. 1982; 50: 1029–1054. doi: [10.2307/1912775](https://doi.org/10.2307/1912775)
9. Honoré BE, Powell JL. Pairwise difference estimators of censored and truncated regression models. *J Econom*. 1994; 64: 241–278. doi: [10.1016/0304-4076\(94\)90065-5](https://doi.org/10.1016/0304-4076(94)90065-5)
10. Hsieh J-J, Ding AA, Wang W. Regression analysis for recurrent events data under dependent censoring. *Biometrics*. 2011; 67: 719–729. doi: [10.1111/j.1541-0420.2010.01497.x](https://doi.org/10.1111/j.1541-0420.2010.01497.x) PMID: [21039394](https://pubmed.ncbi.nlm.nih.gov/21039394/)
11. Jin Z, Lin DY, Wei LJ, Ying Z. Rank-based inference for the accelerated failure time model. *Biometrika*. 2003; 90: 341–353. doi: [10.1093/biomet/90.2.341](https://doi.org/10.1093/biomet/90.2.341)
12. La Cruz W, Martínez JM, Raydan M. Spectral residual method without gradient information for solving large-scale nonlinear systems of equations. *Mathematics of Computation*. 2006; 75: 1429–1448. doi: [10.1090/S0025-5718-06-01840-0](https://doi.org/10.1090/S0025-5718-06-01840-0)
13. Lin DY, Robins JM, Wei LJ. Comparing two failure time distributions in the presence of dependent censoring. *Biometrika*. 1996; 83: 381–393. doi: [10.1093/biomet/83.2.381](https://doi.org/10.1093/biomet/83.2.381)
14. Lindsay BG, Yi GY, Sun J. Issues and strategies in the selection of composite likelihoods. *Statistica Sinica*. 2011; 21: 71–105.
15. Louis TA. Nonparametric analysis of an accelerated failure time model. *Biometrika*. 1981; 68: 381–390. doi: [10.1093/biomet/68.2.381](https://doi.org/10.1093/biomet/68.2.381)
16. Parzen MI, Wei LJ, Ying Z. A resampling method based on pivotal estimating functions. *Biometrika*. 1994; 81: 341–350. doi: [10.1093/biomet/81.2.341](https://doi.org/10.1093/biomet/81.2.341)
17. Peng L, Fine JP. Rank estimation of accelerated lifetime models with dependent censoring. *J Am Stat Assoc*. 2006; 101: 1085–1093. doi: [10.1198/016214506000000131](https://doi.org/10.1198/016214506000000131)
18. Prentice RL, Kalbfleisch JD, Peterson AV Jr, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978; 34: 541–554. doi: [10.2307/2530374](https://doi.org/10.2307/2530374) PMID: [373811](https://pubmed.ncbi.nlm.nih.gov/373811/)
19. Qu A, Lindsay BG, and Li B. Improving generalised estimating equations using quadratic inference functions. *Biometrika*. 2000; 87: 823–836. doi: [10.1093/biomet/87.4.823](https://doi.org/10.1093/biomet/87.4.823)
20. Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. *Ann Stat*. 1990; 18: 354–372. doi: [10.1214/aos/1176347504](https://doi.org/10.1214/aos/1176347504)
21. Varadhan R, Gilbert PD. BB: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *J Stat Softw*. 2009; 32: 1–26.
22. Wang W. Estimating the association parameter for copula models under dependent censoring. *J R Stat Soc Series B Stat Methodol*. 2003; 65: 257–273. doi: [10.1111/1467-9868.00385](https://doi.org/10.1111/1467-9868.00385)
23. Wei LJ, Gail MH. Nonparametric estimation for a scale-change with censored observations. *J Am Stat Assoc*. 1983; 78: 382–388. doi: [10.1080/01621459.1983.10477981](https://doi.org/10.1080/01621459.1983.10477981)
24. Wei LJ, Johnson WE. Combining dependent tests with incomplete repeated measurements. *Biometrika*. 1985; 72: 359–364. doi: [10.1093/biomet/72.2.359](https://doi.org/10.1093/biomet/72.2.359)

25. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc.* 1989; 84: 1065–73. doi: [10.1080/01621459.1989.10478873](https://doi.org/10.1080/01621459.1989.10478873)
26. Xue L, Wang L, Qu A. Incorporating correlation for multivariate failure time data when cluster size is large. *Biometrics.* 2010; 66: 393–404. doi: [10.1111/j.1541-0420.2009.01307.x](https://doi.org/10.1111/j.1541-0420.2009.01307.x) PMID: [19673860](https://pubmed.ncbi.nlm.nih.gov/19673860/)
27. Ying Z. A large sample study of rank estimation for censored regression data. *Ann Stat.* 1993; 21: 76–99. doi: [10.1214/aos/1176349016](https://doi.org/10.1214/aos/1176349016)