Hsiang Yu, Yu-Jen Cheng* and Ching-Yun Wang

Semiparametric Regression Estimation for Recurrent Event Data with Errors in Covariates under Informative Censoring

DOI 10.1515/ijb-2016-0001

Abstract: Recurrent event data arise frequently in many longitudinal follow-up studies. Hence, evaluating covariate effects on the rates of occurrence of such events is commonly of interest. Examples include repeated hospitalizations, recurrent infections of HIV, and tumor recurrences. In this article, we consider semiparametric regression methods for the occurrence rate function of recurrent events when the covariates may be measured with errors. In contrast to the existing works, in our case the conventional assumption of independent censoring is violated since the recurrent event process is interrupted by some correlated events, which is called informative drop-out. Further, some covariates may be measured with errors. To accommodate for both informative censoring and measurement error, the occurrence of recurrent events is modelled through an unspecified frailty distribution and accompanied with a classical measurement error model. We propose two corrected approaches based on different ideas, and we show that they are numerically identical when estimating the regression parameters. The asymptotic properties of the proposed estimators are established, and the finite sample performance is examined via simulations. The proposed methods are applied to the Nutritional Prevention of Cancer trial for assessing the effect of the plasma selenium treatment on the recurrence of squamous cell carcinoma.

Keywords: informative censoring, measurement error, surrogate covariate, recurrent event data

1 Introduction

In many longitudinal follow-up studies, recurrent event data are collected when subjects experience an event multiple times. For example, patients with superficial bladder cancer may experience tumor recurrence many times; patients with cystic fibrosis may experience repeated lung exacerbations; patients with chronic granulomatous disease may experience repeated pyogenic infections [1, 2]. Models for recurrent event data can be categorized into two different classes: time-to-event or gap time models. In time-to-event models, interest focuses on the occurrence rate of an event over time [3–5]. In gap time models, interest lies in the gap time between two consecutive events [6].

In this study, we focus on the time-to-event models. The time-to-event models may be constructed on the basis of an intensity function [7] or a rate function [3, 8]. The intensity function uniquely determines the probability structure of the recurrent event process. However, it needs to specify the occurrence of an event given the prior event history correctly. On the other hand, the rate function allows for arbitrary dependence among the recurrent events and provides a direct interpretation on the occurrence rate without conditioning on the prior event history. Our primary focus is to assess the average effects of treatments or risk factors, that is, we are mainly interested in the inference of the rate function. Lawless and Nadeau [9] estimated the cumulative rate function nonparametrically, and applied their approach to industrial warranty data. In addition, Hu and Lagakos [8] proposed a nonparametric method to study the rate function of viral load changing process for HIV infected patients. Nevertheless, all of the above approaches need to assume non-informative censoring or the observation mechanism is independent of the recurrent process. In practice, the assumption is usually violated; for example, when the recurrent event process is interrupted by some

***Corresponding author: Yu-Jen Cheng,** Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan,
E-mail: ycheng@stat.nthu.edu.tw
Hsiang Yu, Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan
Ching-Yun Wang, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N PO Box 19024 M2B500, Seattle, WA 98109, USA

terminal events that are related to the recurrent events. A potential remedy is to consider a frailty model which allows dependence between the recurrent event process and the informative drop-out through a non-negative frailty variable. In general, the distribution of the frailty variable is assumed to be known [10] and thus the likelihood-based approach [11] is preferred. More recently, Kalbfleisch et al. [12] proposed a weighted estimating equation approach with the weight specified by a gamma frailty distribution. However, in general it is not easy to verify the frailty distribution due to invisibility of the frailty variables. To avoid specification of the frailty distribution, Wang et al. [13] and Wang and Huang [14] considered a conditional likelihood approach, where the unobserved frailty variables are "conditioned away" in their proposed estimating equations.

The aforementioned approaches, nevertheless, require that the covariates are correctly measured. In many epidemiologic or medical studies, the covariates may suffer from measurement errors. For example, baseline plasma selenium level is an important predictor for the occurrence of skin cancers in the Nutritional Prevention of Cancer (NPC) trial study [15]. However, the true value of plasma selenium level can never be measured because of intrinsic biological variability or limited instrumental precision. Instead, the values we observed are contaminated with measurement errors. The most convenient approach is to treat the observed covariates as the true covariates in the regular estimating procedure, which is also referred to as the naive approach. However, the naive estimator obtained from this approach is generally known to be inconsistent ([16], Chapter 3). In survival and longitudinal data analysis, intensive research has been done to deal with measurement error problems. For Cox regression, Prentice [17] proposed likelihood approaches with normal measurement error and rare disease assumptions. Wang et al. [18] applied regression calibration to the partial score function, and investigated the performance of the regression calibration estimator through simulation studies; whereas, Nakamura [19] constructed unbiased estimating equations based on the concept of corrected scores. For nonlinear mixed models, Wu [20], Liu and Wu [21] and Wu et al. [22] proposed estimating approaches for longitudinal response data when the covariates are measured with errors, which can also handle censoring in the response and missing data. In recurrent event analysis, nonetheless, little has been addressed for measurement error problems. Under a normal measurement error assumption, Jiang et al. [23] proposed a moment corrected method to adjust for the bias of a naive estimator under a semi-parametric model. However, their approach not only requires the assumption of non-informative censoring but also assumes that the censoring distribution is independent of the covariates.

The present study is motivated by the NPC trial study, which aimed to assess the efficacy of oral supplement of plasma selenium in preventing the development of skin cancers such as squamous cell carcinoma (SCC). This clinical trial began in 1983 and had included approximately 1,300 patients with dermatologic cancer histories. Nearly half of the patients in the NPC trial were randomly assigned to the placebo or treatment group respectively. Patients in the treatment arm were supposed to take 200 μg of plasma selenium supplement per day. In the study period, the patients in the trial might experience SCC events repeatedly. Each incidence of a new SCC was diagnosed and recorded by certified doctors. The medical records were reviewed by the clinical coordinators at the semi-annual visit, the annual contact or by self-report to ensure the completeness of the data. At the time of randomization, many prognostic risk factors of SCC were recorded including the baseline plasma selenium level. As we mentioned, the plasma selenium level may be measured with errors. In the original study, Clark et al. [15] did not take measurement error into account and found a nonsignificant negative plasma selenium effect on developing SCC. The result contradicted the evidence of the previous studies which showed high correlation between plasma selenium level and several kinds of cancer. Later, many studies focused on the effect of plasma selenium level on the recurrences of SCC by assuming an independent censoring assumption, some of which also took measurement error into account [23]. However, we found a significant negative relationship between the censoring time and the SCC occurrence rate. This implies that the independent censoring assumption is not satisfied. Therefore, the existing methods are not appropriate for the NPC trial data.

This paper is organized as follows. In Section 2, statistical models for recurrent events and measurement errors are given. In Section 3, we propose a regression calibration method and a moment corrected method to correct the measurement errors in the presence of informative censoring. The simulation results

are given in Section 4 to investigate the finite sample performance of the proposed methods. Then, we applied the proposed methods to the NPC trial data to evaluate the effect of selenium on the recurrence of SCC in Section 5, and concluded with a discussion in Section 6. The regularity conditions and technical proofs are provided in the Appendix and the Supplementary Information.

2 Model illustration

2.1 Recurrent event model

Assume that there are n independent individuals in the cohort. Let subscript i be the index for a subject, $i = 1, \ldots, n$. For the ith subject, let $N_i(t)$ denote the number of recurrent events occurring up to t within a fixed time period $[0, \tau]$, where the recurrent event process could be observed beyond τ. Let $\boldsymbol{Z}_i$ be a $q \times 1$ vector of covariates that is precisely measured and $\boldsymbol{X}_i$ be a $p \times 1$ vector of covariates that can be measured with errors. Let $\mathscr{E}$ denote expectation over the samples, v_i be the unobserved frailty variable with mean $\mathscr{E}(v_i|\boldsymbol{X}_i, \boldsymbol{Z}_i) = \mu_v$ which does not depend on $(\boldsymbol{X}_i, \boldsymbol{Z}_i)$, and C_i be the informative censoring time, $i = 1, \ldots, n$. Suppose that conditional on $(v_i, X_i, \boldsymbol{Z}_i)$, $N_i(t)$ follows a Poisson process with a multiplicative intensity function

$$\lambda(t|v_i, \boldsymbol{X}_i, \boldsymbol{Z}_i) = v_i \lambda_0(t) e^{\boldsymbol{\beta}_X' \boldsymbol{X}_i + \boldsymbol{\beta}_Z' \boldsymbol{Z}_i} \tag{1}$$

where $\lambda_0(t)$ is a baseline function and $(\boldsymbol{\beta}_X', \boldsymbol{\beta}_Z')'$ is a vector of regression parameters. Note that when v is given, model (1) is also a rate function due to the assumption of the Poisson process. In general, regression parameters can be estimated by either a likelihood-based approach or by solving a set of unbiased estimating equations. If the distribution of the frailty variable, v, is assumed and the true covariates can be observed, then the standard procedure of the likelihood-based approaches can be conducted by integrating v out ([24], Chapter 3). There are several popular choices for the frailty distribution such as gamma, log-normal, and positive stable distribution. Balakrishnan and Peng [25] advocated using the generalized gamma distribution as the frailty distribution since it includes many distributions (e. g., Weibull, log-normal, gamma, positive stable distribution) as special cases. Recently, Mazroui et al. [26] and Zeng et al. [27] proposed a joint frailty model with two independent frailty variables to distinguish the dependence within the recurrent events and the association between the recurrent event process and terminal events. However, the determination of the frailty distribution usually depends on computational convenience instead of biological reasons or data characteristics. Further, Balakrishnan and Peng [25] pointed out that an inappropriate frailty distribution may result in large bias in the estimation.

Alternatively, we can construct a set of unbiased estimating equations based on the cumulative rate function. According to model (1), the cumulative rate function up to time t is

$$\mathscr{E}(N_i(t)|\boldsymbol{X}_i, Z_i) = \mathscr{E}(\mathscr{E}(N_i(t)|v_i, \boldsymbol{X}_i, \boldsymbol{Z}_i)|\boldsymbol{X}_i, \boldsymbol{Z}_i) = \Lambda_0(t) e^{\alpha_0 + \boldsymbol{\beta}_X' \boldsymbol{X}_i + \boldsymbol{\beta}_Z' \boldsymbol{Z}_i}, \forall t \in [0, \tau] \tag{2}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ and $\alpha_0 = \log(\mu_v)$. It should be noted that an advantage of using estimating equations over a likelihood-based approach is to avoid misspecification of the frailty distribution. However, to solve estimating equations based on eq. (2), $\Lambda_0(t)$ needs to be known and the true covariates need to be observed. Both deficiencies motivate us to consider the recurrent event process with an unspecified distribution of the frailty variable and an unknown $\Lambda_0(t)$ in this article.

2.2 Measurement error model

For subject i, let $\boldsymbol{W}_{ij}$ be the jth replicated surrogate measurement of the true covariate vector $\boldsymbol{X}_i$, and k_i be the number of the replicates of $\boldsymbol{W}_i$. Assume that the surrogate measurement satisfies the classical measurement error model,

$$\boldsymbol{W}_{ij} = \boldsymbol{X}_i + \mathbf{U}_{ij}, i = 1, \ldots, n, j = 1, \ldots, k_i,$$

where $\boldsymbol{U}_{ij}$ are random errors. Suppose that U_{ij} are independent of $(v_i, \boldsymbol{X}_i, \boldsymbol{Z}_i)$ and C_i, which implies that the measurement errors are non-differential. In other words, $\boldsymbol{W}_i$ provides no additional information about the event process when the true covariate $\boldsymbol{X}_i$ is given ([16], Chapter 2). Let μ_s and Σ_s be the mean and covariance matrix of a random vector $\boldsymbol{s}$, Σ_{sh} be the covariance matrix of two random vectors $(\boldsymbol{s}, \boldsymbol{h})$, and $\boldsymbol{\gamma} = (\mu_X, \mu_Z, \Sigma_U, \Sigma_X, \Sigma_Z, \Sigma_{XZ})$ be the parameter of the distribution of $\boldsymbol{X}$ given $(\boldsymbol{W}, \boldsymbol{Z})$. We assume that $\boldsymbol{X}$ given $(\overline{\boldsymbol{W}}, \boldsymbol{Z})$ follows a multivariate normal distribution with mean

$$\mathscr{E}(\boldsymbol{X}|\overline{\boldsymbol{W}}, \boldsymbol{Z}, \boldsymbol{\gamma}) = \boldsymbol{\mu}_X + \begin{pmatrix} \Sigma_X & \Sigma_{XZ} \end{pmatrix} \begin{pmatrix} \Sigma_X + \Sigma_U/k & \Sigma_{XZ} \\ \Sigma_{ZX} & \Sigma_Z \end{pmatrix}^{-1} \begin{pmatrix} \overline{\boldsymbol{W}} - \boldsymbol{\mu}_X \\ \boldsymbol{Z} - \boldsymbol{\mu}_Z \end{pmatrix}$$

and variance

$$\Sigma(\boldsymbol{\gamma}) = \Sigma_X - \begin{pmatrix} \Sigma_X & \Sigma_{XZ} \end{pmatrix} \begin{pmatrix} \Sigma_X + \Sigma_U/k & \Sigma_{XZ} \\ \Sigma_{ZX} & \Sigma_Z \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_X \\ \Sigma_{ZX} \end{pmatrix}.$$

As in Carroll et al. ([16], Chapter 4), the formula given above is the best linear approximation of $\mathscr{E}(\boldsymbol{X}|\overline{\boldsymbol{W}}, \boldsymbol{Z}, \boldsymbol{\gamma})$, and it can also be applied when $\boldsymbol{Z}$ is discrete.

3 Correction for errors-in-variable

Assume the observed data $\{(C_i, (T_{i1}, \ldots, T_{im_i}), \{\boldsymbol{W}_{i1}, \ldots, \boldsymbol{W}_{ik_i}\}, \boldsymbol{Z}_i), i = 1, \ldots, n\}$ are independent and identically distributed (iid), where T_{ij} denotes the observed event times for $j = 1, \ldots, m_i$, and m_i denotes the number of recurrent events occurred before C_i for subject i. As we mentioned in Section 2.1, C is conditionally independent of the recurrent event process $N(t)$ given $(\boldsymbol{v}, \boldsymbol{X}, \boldsymbol{Z})$. Then, by eq. (2) we have

$$\mathscr{E}(N(C)\Lambda_0^{-1}(C)|\boldsymbol{X}, \boldsymbol{Z}) = \mathscr{E}(\mathscr{E}(N(C)\Lambda_0^{-1}(C)|v, C, \boldsymbol{X}, \boldsymbol{Z})|\boldsymbol{X}, \boldsymbol{Z}) = e^{\alpha_0 + \boldsymbol{\beta}_X' \boldsymbol{X} + \boldsymbol{\beta}_Z' \boldsymbol{Z}}.$$

If $\Lambda_0(t)$ and $\boldsymbol{X}$ are known, the estimating equations $\sum_{i=1}^n (1, \boldsymbol{X}_i', \boldsymbol{Z}_i')\{m_i \Lambda_0^{-1}(C_i) - e^{\alpha_0 + \boldsymbol{\beta}_X' \boldsymbol{X}_i + \boldsymbol{\beta}_Z' \boldsymbol{Z}_i}\} = 0$ for $(\boldsymbol{\beta}_X, \boldsymbol{\beta}_Z)$ are unbiased. In practice, they cannot be implemented since $\boldsymbol{X}_i$ is unobserved and $\Lambda_0(t)$ is unknown. To deal with the unknown function $\Lambda_0(t)$, we start with the conditional likelihood function of $(T_{i1}, \ldots, T_{im_i})$ given $(C_i, v_i, m_i, \boldsymbol{X}_i, \boldsymbol{Z}_i)$. Under the Poisson process assumption, such a conditional likelihood can be constructed from a set of iid random variables with truncated density $\prod_{j=1}^{m_i} \lambda_0(T_{ij})/\Lambda_0(C_i) I(0 \le T_{ij} \le C_i)$. Define a rescaled baseline function $\phi(t) \equiv \lambda_0(t)/\Lambda_0(\tau)$ and $\Phi(t) = \int_0^t \phi(u)du = \Lambda_0(t)/\Lambda_0(\tau)$ for $t \in [0, \tau]$, where $\Phi(\tau) = 1$. The conditional likelihood is given by $\prod_{i=1}^n P(T_{i1}, \ldots, T_{im_i}|C_i, v_i, m_i, \boldsymbol{X}_i, \boldsymbol{Z}_i)$ which is proportional to $\prod_{i=1}^n \prod_{j=1}^{m_i} \{\phi(T_{ij})/\Phi(C_i)\}$. As pointed out by Wang et al. [13], the conditional likelihood shares the same form as the nonparametric likelihood for right-truncated data. Thus, $\Phi(t)$ can be consistently estimated by the product limit estimator

$$\widehat{\Phi}(t) = \prod_{T_{(l)} > t} \left(1 - \frac{n_{(l)}}{N_{(l)}}\right),$$

where $\{T_{(l)}\}$ are the ordered and distinct values of $\{T_{ij}\}_{i=1,\ldots,n; j=1,\ldots,m_i}$, $n_{(l)}$ is the number of events occurred at $T_{(l)}$, and $N_{(l)}$ is the number of events which satisfy $T_{ij} \le T_{(l)} \le C_i$. Note that the non-parametric estimation of Φ does not require any information from the covariates and the unobserved frailty variable. Hence, $\widehat{\Phi}(t)$ is a consistent estimator even if $\boldsymbol{X}$ is measured with errors or the frailty distribution is unspecified.

For the issue of identifiability, let $\mu_v = 1$ without loss of generality. The expectation of the event number divided by the rescaled baseline function before time C is

$$\mathscr{E}(N(C)\Phi^{-1}(C)|\boldsymbol{X}, \boldsymbol{Z}) = \mathscr{E}(\mathscr{E}(N(C)\Phi^{-1}(C)|C, v, \boldsymbol{X}, \boldsymbol{Z})|\boldsymbol{X}, \boldsymbol{Z}) = e^{\beta_0 + \boldsymbol{\beta}_X' \boldsymbol{X} + \boldsymbol{\beta}_Z' \boldsymbol{Z}},$$

where $\boldsymbol{\beta}_0 = \log(\Lambda_0(\tau))$. With the above equation, we can construct the unbiased estimating equations by $\Phi(t)$ instead of the unknown $\Lambda_0(t)$. After replacing the unknown $\boldsymbol{X}$ with the average of the replicates $\overline{\boldsymbol{W}}_i = \sum_{j=1}^{k_i} \boldsymbol{W}_{ij}/k_i$, we can obtain the naive estimating equations

$$U_N(\boldsymbol{b}) = n^{-1}\sum_{i=1}^{n}\begin{pmatrix}1\\ \overline{\boldsymbol{W}}_i\\ \boldsymbol{Z}_i\end{pmatrix}\left\{m_i\widehat{\Phi}^{-1}(C_i) - e^{b_0+\boldsymbol{b}_X'\overline{\boldsymbol{W}}_i+\boldsymbol{b}_Z'\boldsymbol{Z}_i}\right\} = 0. \tag{3}$$

Then, the naive estimator $\widehat{\boldsymbol{\beta}}_N = (\widehat{\beta}_{N,0}, \widehat{\boldsymbol{\beta}}_{N,X}', \widehat{\boldsymbol{\beta}}_{N,Z}')'$ is obtained by solving eq. (3) and $\Lambda_0(t)$ can be estimated by $\widehat{\Lambda}_0^N(t) = \widehat{\Phi}(t)\exp(\widehat{\beta}_{N,0})$. Due to the measurement error in $\overline{\boldsymbol{W}}$, it can be shown that $\widehat{\boldsymbol{\beta}}_N$ does not converge to the true parameter $\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_X', \boldsymbol{\beta}_Z')'$. Based on eq. (3), we develop a regression calibration method and a moment corrected method to adjust for covariate measurement errors in the following subsections.

3.1 Regression calibration approach

The regression calibration (RC) method is based on the assumption that the induced model of the response conditioning on $(\overline{\boldsymbol{W}}, \boldsymbol{Z})$ can be well approximated by the underlying model with $\boldsymbol{X}$ being replaced by the conditional mean $\mathscr{E}(\boldsymbol{X}|\overline{\boldsymbol{W}}, \boldsymbol{Z})$. The RC estimator is obtained by treating $\mathscr{E}(\boldsymbol{X}|\overline{\boldsymbol{W}}, \boldsymbol{Z})$ as the true covariate $\boldsymbol{X}$ in the standard estimating procedure ([16], Chapter 4). Although the RC method generally yields to an inconsistent estimator in non-linear models, it is still valuable with the advantage of computational efficiency and limited bias under some conditions [16, 17].

Under our framework, the RC method substitutes $\overline{\boldsymbol{W}}$ with $\mathscr{E}(\boldsymbol{X}|\overline{\boldsymbol{W}}, \boldsymbol{Z}, \boldsymbol{\gamma})$ in eq. (3). If the measurement error covariance matrix Σ_U is known, we can estimate the other components of $\boldsymbol{\gamma}$ by using the observed data without replicates. If not, replicated data is needed to estimate Σ_U [16, 18, 28]. By the method of moments, the estimator $\hat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ can be obtained by solving the equations $n^{-1}\sum_{i=1}^{n}\Psi_i(\boldsymbol{\gamma}) = 0$ where $\Psi_i(\boldsymbol{\gamma})$ is given in Appendix A. Then, the RC estimator $\widehat{\boldsymbol{\beta}}_R = (\widehat{\beta}_{R,0}, \widehat{\boldsymbol{\beta}}_{R,X}', \widehat{\boldsymbol{\beta}}_{R,Z}')'$ can be obtained by solving the equations

$$U_R(\boldsymbol{b}) = n^{-1}\sum_{i=1}^{n}\begin{pmatrix}1\\ \mathscr{E}(\boldsymbol{X}_i|\overline{\boldsymbol{W}}_i, \boldsymbol{Z}_i, \widehat{\boldsymbol{\gamma}})\\ \boldsymbol{Z}_i\end{pmatrix}\left\{m_i\widehat{\Phi}^{-1}(C_i) - e^{b_0+\boldsymbol{b}_X'\mathscr{E}(\boldsymbol{X}_i|\overline{\boldsymbol{W}}_i,\boldsymbol{Z}_i,\widehat{\boldsymbol{\gamma}})+\boldsymbol{b}_Z'\boldsymbol{Z}_i}\right\} = 0. \tag{4}$$

Coincidently, the conditional expectation of $m\Phi^{-1}(C)$ given the observed covariate $(\overline{\boldsymbol{W}}, \boldsymbol{Z})$ is $\exp(\beta_0 + \boldsymbol{\beta}_X'\Sigma(\boldsymbol{\gamma})\boldsymbol{\beta}_X/2 + \boldsymbol{\beta}_X'\mathscr{E}(\boldsymbol{X}|\overline{\boldsymbol{W}}, \boldsymbol{Z}, \boldsymbol{\gamma}) + \boldsymbol{\beta}_Z'\boldsymbol{Z})$. Thus, the RC estimator $\widehat{\boldsymbol{\beta}}_R$ converges to a limit $\boldsymbol{\beta}_R = (\beta_0 + \boldsymbol{\beta}_X'\Sigma(\boldsymbol{\gamma})\boldsymbol{\beta}_X/2, \boldsymbol{\beta}_X', \boldsymbol{\beta}_Z')'$. The result implies that the RC estimator is consistent for the regression coefficients but not for the intercept.

Note that $\widehat{\beta}_{R,0}$ converges to $\beta_0 + \boldsymbol{\beta}_X'\Sigma(\boldsymbol{\gamma})\boldsymbol{\beta}_X/2$. Let $\widehat{\Sigma}$ be the estimator of $\Sigma(\boldsymbol{\gamma})$ which is calculated as $\widehat{\Sigma}_X - (\widehat{\Sigma}_X - \widehat{\Sigma}_{XZ}\widehat{\Sigma}_Z^{-1}\widehat{\Sigma}_{ZX})(\widehat{\Sigma}_{\overline{W}} - \widehat{\Sigma}_{XZ}\widehat{\Sigma}_Z^{-1}\widehat{\Sigma}_{ZX})^{-1}(\widehat{\Sigma}_X - \widehat{\Sigma}_{XZ}\widehat{\Sigma}_Z^{-1}\widehat{\Sigma}_{ZX}) - \widehat{\Sigma}_{XZ}\widehat{\Sigma}_Z^{-1}\widehat{\Sigma}_{ZX}$ where $\widehat{\Sigma}_{\overline{W}} = \widehat{\Sigma}_X + \widehat{\Sigma}_U\sum_{i=1}^{n}(nk_i)^{-1}$. The RC estimator of $\Lambda_0(t)$ can be adjusted as $\widehat{\Lambda}_0^R(t) = \widehat{\Phi}(t)\exp(\widehat{\boldsymbol{\beta}}_{R,0} - \widehat{\boldsymbol{\beta}}_{R,X}'\widehat{\Sigma}\widehat{\boldsymbol{\beta}}_{R,X}/2)$ which converges to $\Lambda_0(t)$. In the Supplementary Information, we show that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_R)$ is asymptotically normally distributed with mean zero and variance $A^{-1}\Sigma_g\{A^{-1}\}'$, where A and Σ_g are defined in Proposition 1 in Appendix A. The covariance matrix estimation of the RC estimator is also given in Appendix B.

3.2 Moment corrected approach

The moment corrected (MC) method is motivated by the bias-correction method proposed by Stefanski [29]. Under the classical measurement error model, Stefanski [29] showed that the naive estimator converges to a limit which is a function of the true parameter and the error variance. Accordingly, the bias of the naive estimator can be corrected based on the relationship between the limit of the naive estimator and the true parameter.

Based on this idea, we can show that the naive estimator $\widehat{\boldsymbol{\beta}}_N$ converges to a limit $\boldsymbol{\beta}_N = (\beta_{N,0}, \boldsymbol{\beta}'_{N,X}, \boldsymbol{\beta}'_{N,Z})'$ which satisfies

$$\mathscr{E}\{U_N(\boldsymbol{\beta}_N)|\overline{\boldsymbol{W}}, \boldsymbol{Z}\} = \mathscr{E}\begin{pmatrix} 1 \\ \overline{\boldsymbol{W}} \\ \boldsymbol{Z} \end{pmatrix}\left\{\mathscr{E}(m\Phi^{-1}(C)|\overline{\boldsymbol{W}}, \boldsymbol{Z}) - e^{\beta_{N,0}+\boldsymbol{\beta}'_{N,X}\overline{\boldsymbol{W}}+\boldsymbol{\beta}'_{N,Z}\boldsymbol{Z}}\right\} = 0. \tag{5}$$

In the Supplementary Information, we have shown that the root of eq. (5) is unique. As described in Section 2.2, we assume that $\boldsymbol{X}$ given $(\overline{\boldsymbol{W}}, \boldsymbol{Z})$ follows a multivariate normal distribution. For the convenience of derivation, we re-parametrize the conditional mean as $\mathscr{E}(\boldsymbol{X}|\overline{\boldsymbol{W}}, \boldsymbol{Z}, \boldsymbol{\gamma}) = \eta_0 + \eta_W\overline{\boldsymbol{W}} + \eta_Z\boldsymbol{Z}$, where I_p denotes an identity matrix of size p, $\eta_0 = (\mathrm{I}_p - \eta_W)\boldsymbol{\mu}_X - \eta_Z\boldsymbol{\mu}_Z$, $\eta_W = (\Sigma_X - \Sigma_{XZ}\Sigma_Z^{-1}\Sigma_{ZX})(\Sigma_{\overline{W}} - \Sigma_{XZ}\Sigma_Z^{-1}\Sigma_{ZX})^{-1}$, and $\eta_Z = \{\mathrm{I}_p - (\Sigma_X - \Sigma_{XZ}\Sigma_Z^{-1}\Sigma_{ZX})(\Sigma_{\overline{W}} - \Sigma_{XZ}\Sigma_Z^{-1}\Sigma_{ZX})^{-1}\}\Sigma_{XZ}\Sigma_Z^{-1}$. By the non-differential error assumption, it follows that $\mathscr{E}(m\Phi^{-1}(C)|\overline{\boldsymbol{W}}, \boldsymbol{Z}) = \mathscr{E}(\mathscr{E}(m\Phi^{-1}(C)|\boldsymbol{X}, \boldsymbol{Z})|\overline{\boldsymbol{W}}, \boldsymbol{Z}) = \exp(\beta_0 + \boldsymbol{\beta}'_X\mathscr{E}(\boldsymbol{X}|\overline{\boldsymbol{W}}, \boldsymbol{Z}, \boldsymbol{\gamma}) + \boldsymbol{\beta}'_X\Sigma(\boldsymbol{\gamma})\boldsymbol{\beta}_X/2 + \boldsymbol{\beta}'_Z\boldsymbol{Z})$. Thus, we can easily show that the unique root $\boldsymbol{\beta}_N$ of eq. (5) is related to the true parameter $\boldsymbol{\beta}$ as $\beta_{N,0} = \beta_0 + \boldsymbol{\beta}'_X\eta_0 + \boldsymbol{\beta}'_X\Sigma(\boldsymbol{\gamma})\boldsymbol{\beta}_X/2$, $\boldsymbol{\beta}_{N,X} = \eta'_W\boldsymbol{\beta}_X$ and $\boldsymbol{\beta}_{N,Z} = \boldsymbol{\beta}_Z + \eta'_Z\boldsymbol{\beta}_X$. Specifically, $\boldsymbol{\beta}_N = D(\boldsymbol{\beta}, \eta)$ is a one to one function of the true parameter $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}'_X, \boldsymbol{\beta}'_Z)'$ when the nuisance parameter η is given. Therefore, substituting the estimators of $\boldsymbol{b}_N$ and η in the inverse function D^{-1} results in the moment corrected estimator

$$\widehat{\boldsymbol{\beta}}_M = D^{-1}(\widehat{\boldsymbol{\beta}}_N, \widehat{\eta}) = \begin{pmatrix} \widehat{\beta}_{N,0} - \widehat{\boldsymbol{\beta}}'_{N,X}\widehat{\eta}_W^{-1}\widehat{\eta}_0 - \widehat{\boldsymbol{\beta}}'_{N,X}\widehat{\eta}_W^{-1}\widehat{\Sigma}\{\widehat{\eta}'_W\}^{-1}\widehat{\boldsymbol{\beta}}_{N,X}/2 \\ \{\widehat{\eta}'_W\}^{-1}\widehat{\boldsymbol{\beta}}_{N,X} \\ \widehat{\boldsymbol{\beta}}_{N,Z} - \widehat{\eta}'_Z\{\widehat{\eta}'_W\}^{-1}\widehat{\boldsymbol{\beta}}_{N,X} \end{pmatrix},$$

where $\widehat{\boldsymbol{\beta}}_M = (\widehat{\beta}_{M,0}, \widehat{\beta}_{M,X}, \widehat{\beta}_{M,Z})$ and $\widehat{\eta}_0 = (\mathrm{I}_p - \widehat{\eta}_W)\widehat{\boldsymbol{\mu}}_X - \widehat{\eta}_Z\widehat{\boldsymbol{\mu}}_Z$, $\widehat{\eta}_W = (\widehat{\Sigma}_X - \widehat{\Sigma}_{XZ}\widehat{\Sigma}_Z^{-1}\widehat{\Sigma}_{ZX})(\widehat{\Sigma}_{\overline{W}} - \widehat{\Sigma}_{XZ}\widehat{\Sigma}_Z^{-1}\widehat{\Sigma}_{ZX})^{-1}$, $\widehat{\eta}_Z = \{\mathrm{I}_p - (\widehat{\Sigma}_X - \widehat{\Sigma}_{XZ}\widehat{\Sigma}_Z^{-1}\widehat{\Sigma}_{ZX})(\widehat{\Sigma}_{\overline{W}} - \widehat{\Sigma}_{XZ}\widehat{\Sigma}_Z^{-1}\widehat{\Sigma}_{ZX})^{-1}\}\widehat{\Sigma}_{XZ}\widehat{\Sigma}_Z^{-1}$. Since $\widehat{\boldsymbol{\beta}}_{M,0}$ is consistent for the true intercept $\boldsymbol{\beta}_0$, $\Lambda_0(t)$ can also be consistently estimated by $\widehat{\Lambda}_0^M(t) = \widehat{\Phi}(t)\exp(\widehat{\beta}_{M,0})$. In summary, the estimating procedure of the MC method is

1. Solve eq. (3) and $\sum_{i=1}^n \Psi_i(\boldsymbol{\gamma}) = 0$ illustrated in Appendix A to obtain the naive estimator $\widehat{\boldsymbol{\beta}}_N$ and $\widehat{\boldsymbol{\gamma}}$.
2. Apply $\widehat{\boldsymbol{\beta}}_N$ and $\widehat{\eta} = \eta(\widehat{\boldsymbol{\gamma}})$ to the function D^{-1} to obtain the MC estimator $\widehat{\boldsymbol{\beta}}_M = D^{-1}(\widehat{\boldsymbol{\beta}}_N, \widehat{\eta})$.

In the Supplementary Information, we show that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})$ is asymptotically normally distributed with mean zero and covariance matrix $B^{-1}\Sigma_h\{B^{-1}\}'$ where B and Σ_h are defined in Proposition 2 in Appendix A. The covariate matrix estimation of the MC estimator is also illustrated in Appendix C.

An important feature of the MC estimator is that it is numerically identical to the RC estimator for the regression parameter $(\boldsymbol{\beta}'_X, \boldsymbol{\beta}'_Z)'$ but not for the intercept β_0. That is, the estimating equations for the two estimators will have exactly the same root for the regression parameters. The proof of $\widehat{\beta}_{M,X} = \widehat{\beta}_{R,X}$ and $\widehat{\beta}_{M,Z} = \widehat{\beta}_{R,Z}$ is provided in Appendix D.

4 Simulation study

In this section, we evaluate the performance of the RC and MC methods with the naive approach under the semi-parametric model via the simulation studies. Additionally, the corrected partial likelihood (CPL) approach proposed by Jiang et al. [23] is also listed for comparison. The CPL estimator takes measurement error into account but assumes non-informative and covariate-independent censoring.

We consider a regression model with a continuous covariate X and a discrete covariate Z. Let $X \sim N(0, \sigma_X^2 = 1/3)$ be the error-prone covariate which is unobserved, while $Z \sim Bin(0.5)$ be a random treatment assignment and is precisely obtained. For subject i, we generate k_i repeated surrogates $W_{ij} = X_i + U_{ij}$ for X_i where k_i is generated from a discrete uniform distribution ranging from 1 to 4 and

$U_{ij} \sim N(0, \sigma_U^2)$. With the repeated surrogates, we estimate the nuisance parameter $\boldsymbol{\gamma}$ by solving $\sum_{i=1}^{n} \Psi_i(\boldsymbol{\gamma})/n = 0$, where Ψ is shown in the Appendix. We conduct the simulations with reliability ratio (RR) $\sigma_X^2/(\sigma_X^2 + \sigma_U^2) = 0.8$ and 0.5. The reliability ratio is used to represent the magnitude of the error contamination, and lower reliability ratio indicates higher error contamination. We generate $v_i^\star$ from a mixture model of which $v_i^\star$ follows a uniform distribution ranging from 0.5 to 1.5 when $Z_i = 0$, and follows a uniform distribution ranging from 1.5 to 4 otherwise. Then, the frailty variable is $v_i = \exp(-Z\log(2.75))v_i^\star$. When (v_i, X_i, Z_i) is given, the recurrent event process $\{N_i(t)\}$ is generated with intensity function $\lambda(t|v_i, X_i, Z_i) = v_i\lambda_0(t)\exp(\beta_X X_i + \beta_Z Z_i)$ in which $\lambda_0(t) = (t-6)^3/360 + 0.6, t \in [0, \tau]$, $\tau = 10$ for $i = 1, \ldots, n$. We consider two distinct coefficient parameters $(\beta_X, \beta_Z) = (\log(1.5), \log(1.5))$ and $(\beta_X, \beta_Z) = (\log(3), \log(1.5))$. To show the robustness of the proposed estimators, the first two scenarios are conducted under different censoring time settings. In Scenario 1, we let the censoring time C depend on W. When $W_{i1} > 0$, C_i is generated from an exponential distribution with mean $10v_i^{-1}$ and is truncated after $\tau = 10$; otherwise, C_i is generated from an exponential distribution with mean $0.5v_i^{-1}$ and is truncated after $\tau = 10$. In Scenario 2, we let the censoring time C depend on X. We generate C_i from the mixed exponential distribution in the same way as in Scenario 1 with W_i replaced by X_i. In addition, we conduct two cases to investigate the sensitivity of the conditional normal assumption imposed on the covariate X. In Scenario 3, X is uniformly distributed over the interval $\left(-\sqrt{3\sigma_X^2}, \sqrt{3\sigma_X^2}\right)$ and Z is allowed to be correlated with X. Let $Z^\star = X + \varepsilon$ where $\varepsilon \sim N(0, \sigma_X^2)$, and $Z = 1$ if $Z^\star \le 0$ and $Z = 0$ otherwise. The other variables are generated the same as those in Scenario 2. Further, a non-normal measurement error case is considered in Scenario 4. We generate measurement error U from a skew normal distribution with mean 0, variance σ_U^2 and skewness parameter $\alpha = -2$, and X from $N(0, \sigma_X^2 = 1/3)$. The remaining variables are generated the same as those in Scenario 3. A total of 200 replicates with sample sizes $n = 300$ and $n = 600$ are generated in each simulation configuration. In the tables, BIAS denotes the average bias, ASE denotes the average standard error estimation, ESD denotes the empirical sample standard deviation, and CP and CL denote the coverage probability and average interval length of the 95% confidence interval based on 200 runs. The standard errors of the proposed estimators are obtained by taking the square roots of the diagonal elements from the sandwich variance estimators given in Appendices B and C.

The results of Scenarios 1 to 4 are demonstrated in Tables 1 to 4. In general, the naive estimator for the error-prone covariate X has large biases and disastrous coverage probabilities as shown in all tables. This phenomenon is due to the common attenuation effect. The degree of bias becomes critical when the error-prone covariate effect is large and the reliability ratio is low. In Scenarios 1 and 2, the naive estimation of the effect of Z is not affected by the measurement errors since X and Z are generated to be mutually independent. In Scenario 3 in which X and Z are correlated, the naive estimator for β_Z also has low coverage probabilities which is shown in Tables 3 and 4. Further, the numerical equivalence of the RC and MC estimators is also seen in the simulation results.

Table 1 demonstrates the results when C depends on W. Comparing to the proposed estimators, we can see that the CPL estimator generally has larger but not significant biases when reliability ratio becomes lower (RR = 0.5). However, when C depends on X, the coverage probabilities of the CPL estimator for β_X dramatically decline due to the substantial biased problem which is presented in Table 2. The bias problem becomes more serious as the coefficient parameter β_X increases or the reliability ratio decreases. Table 3 shows the results when X follows a uniform distribution. We can see that the coverage probabilities of the CPL estimators for (β_X, β_Z) are both nearly zero in the setting with a large coefficient parameter, a large sample size and a low reliability ratio. In contrast, the proposed methods have good performance with at least 92% coverage probabilities and limited biases even if the conditional normal assumption on X is violated. In Table 4, it can be seen that the proposed estimators still have good performance in terms of bias and coverage probability compared to the CPL estimator. However, when the sample size increases to $n = 2000$, the coverage probabilities of the 95% confidence intervals for the proposed estimators may be lower than 90%.

Table 1: Censoring time depends on W; X follows a normal distribution, and X and Z are independent.

		$n = 300$				$n = 600$			
		Naive	**RC**	**MC**	**CPL**	**(Naive**	**RC**	**MC**	**CPL**
						$(\beta_X, \beta_Z) = (\log(1.5), \log(1.5))$; **RR = 0.8**			
β_X	BIAS × 10^3	−83	−1	−1	24	−71	13	13	39
	ASE × 10^3	137	172	172	127	96	120	120	87
	ESD × 10^3	133	167	167	120	94	117	117	86
	CP	0.93	0.97	0.97	0.96	0.91	0.94	0.94	0.91
	CL × 10^3	537	675	676	497	375	470	470	343
β_Z	BIAS × 10^3	−2	−2	−2	13	−4	−4	−4	5
	ASE × 10^3	164	164	164	106	114	114	114	74
	ESD × 10^3	157	157	157	102	120	120	120	75
	CP	0.97	0.96	0.96	0.96	0.96	0.96	0.96	0.95
	CL × 10^3	643	644	644	416	446	447	447	292
						$(\beta_X, \beta_Z) = (\log(1.5), \log(1.5))$; **RR = 0.5**			
β_X	BIAS × 10^3	−216	−20	−20	−89	−198	11	11	49
	ASE × 10^3	102	209	209	150	78	156	156	117
	ESD × 10^3	103	211	211	139	69	142	142	115
	CP	0.43	0.92	0.92	0.89	0.23	0.96	0.96	0.925
	CL × 10^3	424	868	869	650	304	612	613	460
β_Z	BIAS × 10^3	5	7	7	14	7	8	8	12
	ASE × 10^3	162	163	164	108	117	117	117	77
	ESD × 10^3	168	169	169	107	115	117	117	79
	CP	0.94	0.94	0.94	0.94	0.96	0.96	0.96	0.93
	CL × 10^3	650	655	655	427	458	460	460	300
						$(\beta_X, \beta_Z) = (\log(3), \log(1.5))$; **RR = 0.8**			
β_X	BIAS × 10^3	−241	−24	−24	3	−214	7	7	23
	ASE × 10^3	130	163	163	136	92	115	115	100
	ESD × 10^3	129	161	161	139	94	119	119	99
	CP	0.53	0.95	0.95	0.94	0.34	0.96	0.96	0.96
	CL × 10^3	508	639	639	532	360	451	451	392
β_Z	BIAS × 10^3	22	25	25	10	8	9	9	2
	ASE × 10^3	146	146	146	110	103	103	103	79
	ESD × 10^3	147	148	148	116	99	102	102	79
	CP	0.95	0.95	0.95	0.94	0.95	0.95	0.95	0.94
	CL × 10^3	573	574	574	433	402	403	403	311
						$(\beta_X, \beta_Z) = (\log(3), \log(1.5))$; **RR = 0.5**			
β_X	BIAS × 10^3	−539	34	34	67	−551	−4	−4	12
	ASE × 10^3	106	221	222	226	75	154	155	158
	ESD × 10^3	111	228	228	234	76	146	146	151
	CP	0.00	0.94	0.94	0.95	0.00	0.96	0.96	0.96
	CL × 10^3	417	867	868	887	295	606	606	619
β_Z	BIAS × 10^3	7	8	8	15	−7	−8	−8	2
	ASE × 10^3	155	159	159	130	110	112	112	95
	ESD × 10^3	157	163	163	138	119	121	121	92
	CP	0.95	0.94	0.94	0.94	0.94	0.94	0.94	0.95
	CL × 10^3	607	622	623	511	430	438	439	371

Note: BIAS denotes the average of $\widehat{\beta} - \beta$ from 200 samplings, ASE denotes the average standard error from 200 samplings, ESD denotes the empirical standard deviation from 200 samplings, CP denotes the coverage probability of Wald 95 % confidence interval, CL denotes the average length of Wald 95 % confidence interval from 200 samplings.

Table 2: Censoring time depends on X; X follows a normal distribution, and X and Z are independent.

		$n = 300$				$n = 600$			
		Naive	**RC**	**MC**	**CPL**	**Naive**	**RC**	**MC**	**CPL**
						$(\beta_X, \beta_Z) = (\log(1.5), \log(1.5))$; **RR** = **0.8**			
β_X	BIAS × 10^3	−80	2	3	−14	−102	−25	−20	−30
	ASE × 10^3	133	167	167	118	96	120	120	84
	ESD × 10^3	136	171	171	131	93	118	117	87
	CP	0.93	0.94	0.95	0.92	0.73	0.97	0.97	0.90
	CL × 10^3	523	655	653	462	375	470	471	331
β_Z	BIAS × 10^3	16	15	15	13	−20	−19	−21	−6
	ASE × 10^3	161	161	161	105	115	115	115	75
	ESD × 10^3	161	162	161	118	109	108	110	74
	CP	0.95	0.95	0.95	0.89	0.97	0.97	0.97	0.97
	CL × 10^3	632	632	632	412	451	451	451	294
						$(\beta_X, \beta_Z) = (\log(1.5), \log(1.5))$; **RR** = **0.5**			
β_X	BIAS × 10^3	−216	−20	−20	−89	−198	11	11	49
	ASE × 10^3	102	209	209	150	78	156	156	117
	ESD × 10^3	103	211	211	139	69	142	142	115
	CP	0.43	0.92	0.92	0.89	0.23	0.96	0.96	0.925
	CL × 10^3	401	821	821	587	289	586	586	409
β_Z	BIAS × 10^3	5	7	7	14	7	8	8	12
	ASE × 10^3	162	163	164	108	117	117	117	77
	ESD × 10^3	168	169	169	107	115	117	117	79
	CP	0.94	0.94	0.94	0.94	0.96	0.96	0.96	0.93
	CL × 10^3	523	655	653	462	375	470	471	331
						$(\beta_X, \beta_Z) = (\log(3), \log(1.5))$; **RR** = **0.8**			
β_X	BIAS × 10^3	−225	−8	−8	−91	−216	4	4	−95
	ASE × 10^3	126	159	159	133	88	111	111	94
	ESD × 10^3	124	157	157	141	84	106	106	89
	CP	0.58	0.96	0.96	0.87	0.33	0.95	0.95	0.83
	CL × 10^3	496	622	622	520	346	434	434	368
β_Z	BIAS × 10^3	3	4	4	7	3	3	3	4
	ASE × 10^3	143	144	144	109	101	102	102	78
	ESD × 10^3	147	146	146	109	98	97	97	82
	CP	0.95	0.94	0.94	0.94	0.98	0.98	0.98	0.93
	CL × 10^3	562	563	563	429	398	398	398	305
						$(\beta_X, \beta_Z) = (\log(3), \log(1.5))$; **RR** = **0.5**			
β_X	BIAS × 10^3	−558	−6	−6	−238	−552	−1	−1	−229
	ASE × 10^3	99	207	208	195	70	144	144	139
	ESD × 10^3	100	202	202	186	68	147	147	133
	CP	0.00	0.97	0.97	0.74	0.00	0.95	0.95	0.59
	CL × 10^3	389	812	814	763	273	565	565	546
β_Z	BIAS × 10^3	6	6	6	11	−10	−12	−12	1
	ASE × 10^3	151	154	154	125	106	108	108	88
	ESD × 10^3	154	158	158	115	110	115	115	95
	CP	0.93	0.95	0.95	0.96	0.97	0.94	0.94	0.92
	CL × 10^3	591	603	604	489	416	424	424	345

Note: BIAS denotes the average of $\widehat{\beta} - \beta$ from 200 samplings, ASE denotes the average standard error from 200 samplings, ESD denotes the empirical standard deviation from 200 samplings, CP denotes the coverage probability of Wald 95 % confidence interval, CL denotes the average length of Wald 95 % confidence interval from 200 samplings.

Table 3: Censoring time depends on X; X follows a uniform distribution, and X and Z are correlated.

		$n = 300$				$n = 600$			
		Naive	**RC**	**MC**	**CPL**	**Naive**	**RC**	**MC**	**CPL**
						$(\beta_X, \beta_Z) = (\log(1.5), \log(1.5))$; **RR = 0.8**			
β_X	BIAS × 10^3	−114	6	6	−71	−119	−2	−2	−78
	ASE × 10^3	151	213	213	132	108	152	152	93
	ESD × 10^3	163	230	230	142	104	147	147	85
	CP	0.88	0.94	0.94	0.88	0.83	0.95	0.95	0.91
	CL × 10^3	591	834	834	519	422	595	595	363
β_Z	BIAS × 10^3	98	12	12	81	81	−3	−3	88
	ASE × 10^3	200	223	223	138	141	157	157	94
	ESD × 10^3	216	239	239	138	139	157	157	94
	CP	0.92	0.93	0.93	0.92	0.91	0.96	0.96	0.84
	CL × 10^3	785	875	875	541	552	616	616	370
						$(\beta_X, \beta_Z) = (\log(1.5), \log(1.5))$; **RR = 0.5**			
β_X	BIAS × 10^3	−253	2	2	−170	−244	24	24	−152
	ASE × 10^3	109	297	298	152	77	207	207	107
	ESD × 10^3	125	340	340	153	74	207	207	100
	CP	0.37	0.93	0.93	0.76	0.10	0.95	0.95	0.69
	CL × 10^3	426	1166	1168	595	302	811	811	419
β_Z	BIAS × 10^3	178	−7	−7	178	180	−13	−13	153
	ASE × 10^3	195	275	275	134	137	190	190	95
	ESD × 10^3	208	303	303	122	134	186	186	91
	CP	0.85	0.92	0.92	0.75	0.76	0.94	0.94	0.59
	CL × 10^3	766	1077	1077	524	537	744	744	372
						$(\beta_X, \beta_Z) = (\log(3), \log(1.5))$; **RR = 0.8**			
β_X	BIAS × 10^3	−351	−42	−42	−306	−364	−66	−66	−324
	ASE × 10^3	138	197	197	141	98	139	139	100
	ESD × 10^3	141	199	199	140	96	133	133	87
	CP	0.25	0.93	0.93	0.42	0.05	0.95	0.95	0.08
	CL × 10^3	542	771	772	552	385	544	545	391
β_Z	BIAS × 10^3	237	15	15	209	240	26	26	215
	ASE × 10^3	189	209	209	148	134	147	147	101
	ESD × 10^3	215	235	235	142	138	149	149	102
	CP	0.75	0.93	0.93	0.70	0.58	0.94	0.94	0.46
	CL × 10^3	743	819	819	581	527	577	577	398
						$(\beta_X, \beta_Z) = (\log(3), \log(1.5))$; **RR = 0.5**			
β_X	BIAS × 10^3	−722	−89	−89	−548	−715	−85	−85	−554
	ASE × 10^3	96	273	274	169	67	187	187	121
	ESD × 10^3	87	254	254	159	69	191	191	117
	CP	0.00	0.94	0.93	0.11	0.00	0.91	0.92	0.01
	CL × 10^3	377	1071	1073	664	263	732	733	476
β_Z	BIAS × 10^3	470	11	11	349	487	35	35	368
	ASE × 10^3	188	261	261	162	133	180	180	115
	ESD × 10^3	190	248	248	138	150	193	193	107
	CP	0.31	0.95	0.96	0.40	0.05	0.94	0.94	0.10
	CL × 10^3	738	1021	1021	636	522	707	705	449

Note: BIAS denotes the average of $\widehat{\beta} - \beta$ from 200 samplings, ASE denotes the average standard error from 200 samplings, ESD denotes the empirical standard deviation from 200 samplings, CP denotes the coverage probability of Wald 95 % confidence interval, CL denotes the average length of Wald 95 % confidence interval from 200 samplings.

Table 4: Censoring time depends on X; U follows a skew normal distribution, and X and Z are correlated.

		$n = 300$				$n = 600$			
		Naive	**RC**	**MC**	**CPL**	**Naive**	**RC**	**MC**	**CPL**
						$(\beta_X, \beta_Z) = (\log(1.5), \log(1.5))$; **RR** = **0.8**			
β_X	BIAS × 10^3	−118	−10	−10	−43	−99	16	16	−44
	ASE × 10^3	150	207	207	127	104	142	142	91
	ESD × 10^3	150	208	208	118	103	142	142	84
	CP	0.89	0.95	0.95	0.95	0.84	0.95	0.95	0.94
	CL × 10^3	589	811	811	499	406	558	559	357
β_Z	BIAS × 10^3	97	26	26	84	50	−25	−25	75
	ASE × 10^3	198	217	217	130	138	150	150	89
	ESD × 10^3	197	217	217	124	133	146	146	89
	CP	0.89	0.95	0.95	0.90	0.94	0.94	0.94	0.88
	CL × 10^3	777	851	851	511	541	588	588	349
						$(\beta_X, \beta_Z) = (\log(1.5), \log(1.5))$; **RR** = **0.5**			
β_X	BIAS × 10^3	−244	5	5	−109	−231	32	32	−115
	ASE × 10^3	111	285	285	155	77	195	195	110
	ESD × 10^3	114	296	296	137	79	202	202	103
	CP	0.42	0.94	0.94	0.92	0.16	0.93	0.94	0.82
	CL × 10^3	437	1115	1118	606	302	763	764	432
β_Z	BIAS × 10^3	179	16	16	149	134	−36	−36	139
	ASE × 10^3	191	254	255	130	133	174	174	88
	ESD × 10^3	187	252	252	120	128	174	174	88
	CP	0.84	0.94	0.94	0.83	0.85	0.95	0.95	0.68
	CL × 10^3	747	997	998	508	522	681	682	346
						$(\beta_X, \beta_Z) = (\log(3), \log(1.5))$; **RR** = **0.8**			
β_X	BIAS × 10^3	−271	40	40	−115	−261	51	51	−121
	ASE × 10^3	137	189	190	142	96	133	133	100
	ESD × 10^3	144	195	195	144	92	129	129	93
	CP	0.50	0.93	0.92	0.87	0.21	0.94	0.94	0.78
	CL × 10^3	536	742	743	558	377	522	522	394
β_Z	BIAS × 10^3	157	−46	−46	148	173	−29	−29	166
	ASE × 10^3	185	201	201	136	131	142	142	95
	ESD × 10^3	172	188	188	122	131	144	144	93
	CP	0.89	0.96	0.96	0.83	0.74	0.95	0.95	0.59
	CL × 10^3	725	790	790	534	512	556	556	373
						$(\beta_X, \beta_Z) = (\log(3), \log(1.5))$; **RR** = **0.5**			
β_X	BIAS × 10^3	−623	98	98	−295	−615	99	99	−273
	ASE × 10^3	109	290	291	202	79	215	216	148
	ESD × 10^3	112	305	305	172	88	220	220	150
	CP	0.00	0.95	0.95	0.70	0.00	0.93	0.93	0.48
	CL × 10^3	428	1139	1142	791	309	805	808	582
β_Z	BIAS × 10^3	419	−45	−45	342	411	−56	−56	335
	ASE × 10^3	185	255	255	147	130	179	180	103
	ESD × 10^3	189	269	269	140	128	185	185	88
	CP	0.39	0.93	0.93	0.36	0.11	0.93	0.93	0.08
	CL × 10^3	725	998	1000	576	509	703	705	405

Note: BIAS denotes the average of $\widehat{\beta} - \beta$ from 200 samplings, ASE denotes the average standard error from 200 samplings, ESD denotes the empirical standard deviation from 200 samplings, CP denotes the coverage probability of Wald 95 % confidence interval, CL denotes the average length of Wald 95 % confidence interval from 200 samplings.

To summarize, the simulation study reveals that the proposed methods can effectively correct the bias due to measurement errors even when the conditional normal assumption of X is violated. However, the CPL estimator is sensitive to the assumption of the independence between the censoring time and the covariates, and is also sensitive to the distributional assumption imposed on the covariates. The both assumptions may not be verified since X is unobservable. The simulation study also shows that the naive approach which ignores measurement errors in the covariates in general will cause a large bias. We note that the proposed estimators are not consistent in Scenarios 3 and 4 because of a violation of the normal assumption imposed on X given (W, Z). Hence, the corresponding coverage probabilities obtained from the 95% confidence intervals may be lower than 90% when the sample size is large (such as $n = 2000$), especially under a skewed measurement error distribution.

5 Data analysis

In this section, we apply the proposed methods to the NPC trial dataset to assess the effect of plasma selenium treatment on SCC recurrences. This randomized, double-blinded clinical trial recruited 1312 patients with histories of skin cancer, including 653 and 659 patients in the treatment and placebo groups respectively, and the study period had lasted up to 12 years.

Many critical risk factors for SCC were recorded at baseline, particularly the plasma selenium level. As we mentioned, the plasma selenium level is measured with error due to the measuring instrument or temporary biological fluctuation. Some patients in the placebo group had more than one plasma selenium measurement that can be treated as replicates. However, the repeat plasma selenium measurements of the treatment group patients can not represent the baseline value. Therefore, the treatment group patients had only one baseline plasma selenium measurement. Multiple occurrences of SCC can be observed for each patient because each new incidence of SCC was diagnosed and recorded during the follow-up time.

In this analysis, we consider two covariates: the baseline plasma selenium measurement, and the treatment assignment indicator. The latter is our primary covariate of interest, whereas the former is an important predictor for adjusting the model but is contaminated with measurement errors. After taking logarithm, the plasma selenium measurement follows a normal distribution (shown in Figure 1). Thus, we let X be the logarithm of the baseline plasma selenium measurement (abbreviated as log(selenium)), and Z be the treatment assignment. We assume that the recurrence of SCC follows a non-homogeneous Poisson process with intensity function $\lambda(t|v, X, Z) = v\lambda_0(t)\exp(\beta_X X + \beta_Z Z)$, where the frailty variable v accounts for the correlations among the SCC recurrences and between the SCC event process and informative censoring time. Here, X is independent of Z since the NPC trial is a randomized clinical trial. Assume that X given W follows a conditional normal distribution. By using the replicate data, the variance of X given W is estimated by $\widehat{\sigma}^2 = \widehat{\sigma}_U^2\widehat{\sigma}_X^2/\widehat{\sigma}_W^2 = 0.156^2 \cdot 0.133^2/0.205^2 = 0.101^2$.

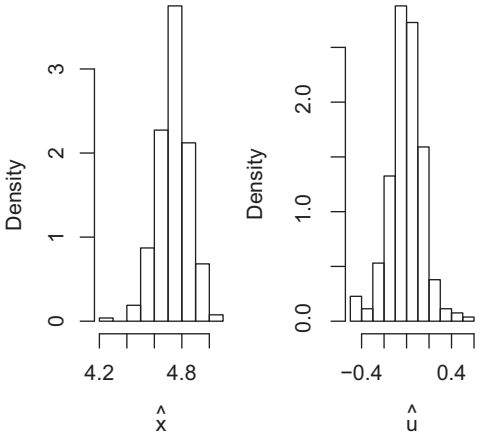

Figure 1: Histograms of estimated true covariate $\{\widehat{X}_i\}$ and estimated error terms $\{\widehat{U}_i\}$ by using 292 placebo grouped patients with more than 10 plasma selenium measurements.

To verify the distributional assumptions imposed on the covariates, a subset consisting of 292 placebo-grouped patients with 10 or more selenium measurements is used. Because the numbers of replicates of these patients are large enough, the average of replicates should be very close to the true value of the plasma selenium level. Thus, we estimate X_i by $\widehat{X}_i = \sum_{j=1}^{k_i} W_{ij}/k_i$ and U_i by $\widehat{U}_i = W_{i1} - \widehat{X}_i$, in which X_i is the true plasma selenium level of the ith patient in the subset. Figure 1 shows the histograms of $\widehat{X}$ and $\widehat{U}$, which suggest the marginal normal distributions for X and U. The correlation between $\widehat{X}$ and $\widehat{U}$ is only −0.069 with P-value = 0.234. Under the assumption of normality, the non-significant correlation implies the independence between the two variables. Therefore, the conditional normal assumption of X is appropriate for the NPC dataset.

The patients in the trial were arranged to receive the dermatologic examination periodically. Let C_i denote the last examination time from the randomization for subject i, and $\tau = 149.5(months)$ denote the maximum time of C_i's. Fifty five patients without any record of dermatologic examination and SCC event are excluded from the data analysis. The existing recurrent event studies [15, 23] for the NPC data assumed that the censoring is non-informative, which might be improper. Figure 2 shows the weighted average of the SCC recurrences versus time for subjects in the four selected risk sets ($t_1 = 54.9$, $t_2 = 86.3$, $t_3 = 115.5$, $t_4 = 135.2$). Note that for a subject i the number of SCC recurrences by time t is calculated as $N_i(t \wedge C_i)$, where $a \wedge b = \min(a, b)$. If the censoring time is independent of the SCC recurrence, we expect that all lines should be close to each other. However, it can be seen that the subjects stayed in the trial longer (censoring time after 115.5 months and 135.2 months) tended to have fewer SCC recurrences in the early and middle stages. The result implies that the independent censoring assumption is not satisfied and the proposed methods are necessary.

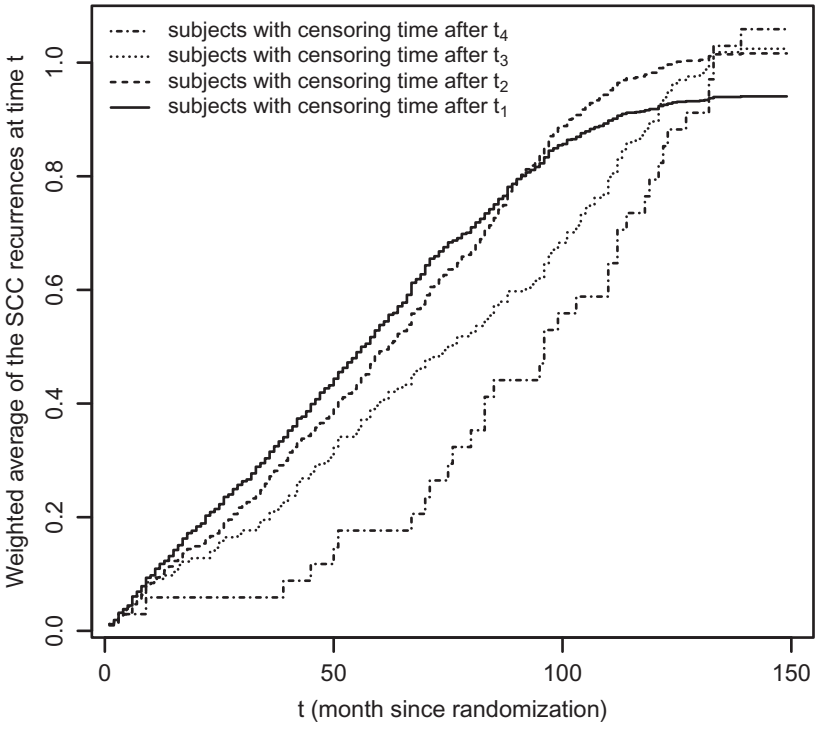

Figure 2: Weighted average of the SCC recurrences versus time (month since randomization) for subjects in the four selected risk sets ($t_1 = 54.9$, $t_2 = 86.3$, $t_3 = 115.5$,$t_4 = 135.2$), where the weighted average of the SCC recurrences for subjects in the rth risk set at time t is calculated by $\sum_{i=1}^{n} N_i(t \wedge C_i) I(C_i > t_r) / \sum_{i=1}^{n} I(C_i > t_r)$, $0 \le t \le \tau = 149.5$ where $r = 1, 2, 3, 4$.

After excluding 55 patients without any record of examination and SCC event and 2 without baseline plasma selenium measurements, 1,255 patients are included in the analysis to fit the semi-parametric model for the SCC recurrences. Among these patients, 473 had at least one SCC occurrence. The result of the fitted model is presented in Table 5. Since the RC and MC estimates are identical, only the RC estimates

Table 5: Regression analysis of the SCC recurrences in the NPC trial.

		Naive	RC	CPL
log (Selenium)	EST	−0.555	−1.502	−1.109
	SE	0.292	0.790	0.842
	Z-value	−1.897	−1.902	−1.317
Treatment	EST	0.185	0.223	0.125
	SE	0.140	0.141	0.125
	Z-value	1.317	1.581	1.002

Note: EST denotes the estimate, SE denotes the standard error which is estimated by the square root of the asymptotic variance estimator. The MC estimates are identical to the RC estimates.

are shown in the table. As illustrated, the treatment effect estimates of all approaches are positive but statistically non-significant. That is, the supplement of plasma selenium has no significant effect on preventing the recurrence of SCC. This result agrees with the previous studies [15, 23]. In Table 5, we can also observe the attenuation phenomenon exists in the naive estimation of the plasma selenium effect. Under the 95% confidence level, the adjusted estimates obtained from the RC and MC methods are significant with values equal to -1.502. The result implies that patients with higher plasma selenium level at baseline have fewer SCC recurrences.

6 Discussion

To identify the population risk factor in the recurrent event analysis, inference on the rate function is commonly preferred. The existing methods depend on the assumptions of either accurately measured covariates or independent censoring, which may not be always realistic. In this article, we consider statistical methods for recurrent event data with measurement error and informative censoring. Under the informative censoring and normal error assumption, our proposed estimators are consistent. In our estimating procedure, we do not need any additional assumptions on the frailty distribution or on the censoring time. The numerical results have shown that the naive method which ignores measurement errors in the covariates leads to a large biased estimator and that the CPL method strongly depends on the independence between the covariates and censoring time. Whereas, our proposed methods correct measurement errors effectively and give accurate confidence intervals under different scenarios.

The corrected methods considered in this paper are developed under a parametric distribution for the covariates and measurement errors, in which the distributions of the errors and covariates are specified. In the NPC data example, the distributional assumptions for the error model can be validated via adequate replicates. In practice, we may not have enough information to validate these distributional assumptions of the errors and covariates. To relax such assumptions, a non-parametric correction method, similar to Huang and Wang [30] for Cox regression with measurement error, might be further developed. However, the extension of nonparametric correction to the regression analysis of recurrent event data is not straightforward, and hence future research is warranted. The idea of measurement error correction can be applied not only to recurrent event data but also to panel count data, of which the number of events can only be observed at several random times.

Acknowledgements: We thank the editor and referees for their very helpful comments and suggestions that greatly improved the paper. This research was partially supported by Taiwan Ministry of Science and Technology MOST 104-2118-M-007-002 (Cheng and Yu), National Institutes of Health grants CA53996, ES017030, HL121347, and MH105857 (Wang), and a MOST travel award from the Mathematics Research Promotion Center (Wang).

Appendices

A Asymptotic properties

Let $\mathscr{B}_R$, $\mathscr{B}$ be any compact neighborhoods of $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}$ which are the roots of the limits of RC and MC estimating equations. Also, denote $\mathscr{W}_i = (1, \overline{\boldsymbol{W}}_i^{'}, \boldsymbol{Z}_i^{'})^{'}$ and $\mathscr{X}_i = (1, \mathscr{E}(\boldsymbol{X}_i|\overline{\boldsymbol{W}}_i, \boldsymbol{Z}_i, \boldsymbol{\gamma})^{'}, \boldsymbol{Z}_i^{'})^{'}$. To prove the asymptotic properties of the proposed estimators, we impose the following regularity conditions:

(a1) $\Lambda_0(\tau) > 0$;

(a2) $Pr(C \geq \tau, v > 0) > 0$;

(a3) $G(u) \equiv \mathscr{E}[vI(C \geq u)]$ is a continuous function for $u \in [0, \tau]$;

(a4) $\mathscr{E}\{\sup_{\boldsymbol{b}\in\mathscr{B}} \mathscr{W}\mathscr{W}^{'} \exp(D^{'}(\boldsymbol{b}, \eta)\mathscr{W})\}$ and $\mathscr{E}\{\sup_{\boldsymbol{b}\in\mathscr{B}_R} \mathscr{X}\mathscr{X}^{'} \exp(\boldsymbol{b}^{'}\mathscr{X})\}$ are bounded. Moreover, $\mathscr{E}\{\mathscr{W}\mathscr{W}^{'} \exp(D^{'}(\boldsymbol{\beta}, \eta)\mathscr{W})\}$ and $\mathscr{E}\{\mathscr{X}\mathscr{X}^{'} \exp(\boldsymbol{\beta}_R^{'}\mathscr{X})\}$ are non-singular.

Note that condition (a4) can be satisfied under the normality assumption imposed on the covariates.

Define $Q_1(t) \equiv G(t)\Lambda_0(t)$, $Q_2(t) \equiv \int_0^t G(u)d\Lambda_0(u)$. Under conditions (a1) through (a3), Wang et al. [13] had shown that

$$\widehat{\Phi}(t) - \Phi(t) = \frac{1}{n}\sum_{i=1}^{n} \Phi(t)d_i(t) + o_p(n^{-1/2}), \forall \inf\{s: \Lambda_0(s) > 0\} < t \leq \tau, \tag{6}$$

where $d_i(t) \equiv \sum_{j=1}^{m_i}\{\int_t^\tau I(T_{ij} \leq u \leq C_i)/Q_1^2(u)dQ_2(u) - I(t < T_{ij} \leq \tau)/Q_1(T_{ij})\}$ are iid terms with zero expectation. By the central limit theorem, $\sqrt{n}(\widehat{\Phi}(t) - \Phi(t))$ converges to a multivariate normal distribution with mean zero and variance $\Phi^2(t)\mathscr{E}[d_i^2(t)]$.

By the method of moments, the nuisance parameter estimator $\widehat{\boldsymbol{\gamma}}$ is obtained by solving

$$n^{-1}\sum_{i=1}^{n}\Psi_i(\boldsymbol{\gamma}) = n^{-1}\sum_{i=1}^{n}\begin{pmatrix} k_i(\overline{\boldsymbol{W}}_i - \boldsymbol{\mu}_X) \\ \boldsymbol{Z}_i - \boldsymbol{\mu}_Z \\ \sum_{j=1}^{k_i}(\boldsymbol{W}_{ij} - \bar{W}_i)(\boldsymbol{W}_{ij} - \bar{W}_i)^{'} - (k_i - 1)\Sigma_U \\ k_i(\overline{\boldsymbol{W}}_i - \boldsymbol{\mu}_x)(\overline{\boldsymbol{W}}_i - \boldsymbol{\mu}_x)^{'} - \Sigma_U - k_i\Sigma_X \\ (\boldsymbol{Z}_i - \boldsymbol{\mu}_Z)(\boldsymbol{Z}_i - \boldsymbol{\mu}_Z)^{'} - \Sigma_Z \\ (\overline{\boldsymbol{W}}_i - \boldsymbol{\mu}_X)(\boldsymbol{Z}_i - \boldsymbol{\mu}_Z)^{'} - \Sigma_{XZ} \end{pmatrix} = 0,$$

where $\Psi_i(\boldsymbol{\gamma})$ are iid terms. With the same techniques as these in M-estimators [31], it can be shown that $\widehat{\boldsymbol{\gamma}}$ converges in probability to $\boldsymbol{\gamma}$. Let $R \equiv \mathscr{E}\{-\partial\Psi_i(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^{'}\}$ where R is non-singular under condition (a4), and thus by a Taylor expansion,

$$\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = R^{-1}n^{-1}\sum_{i=1}^{n}\Psi_i(\boldsymbol{\gamma}) + o_p(n^{-1/2}). \tag{7}$$

By the central limit theorem, $\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ converges to a normal distribution with mean zero and a covariance-matrix $R^{-1}\mathscr{E}\{\Psi_i(\boldsymbol{\gamma})\Psi_i(\boldsymbol{\gamma})^{'}\}\{R^{-1}\}^{'}$.

With the consistencies of $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\Phi}(t)$, $\forall t \in [0, \tau]$, we can prove the following propositions of which the proofs are given in the Supplementary Information. Define V as the joint density of (W, Z, m, c), $\Pi \equiv \partial\mathscr{X}/\partial\boldsymbol{\gamma}^{'}$, and $\Gamma \equiv \partial D/\partial\boldsymbol{\gamma}^{'}$. Let

$$g_i = \mathscr{X}_i\left\{\frac{m_i}{\Phi(C_i)} - e^{\boldsymbol{\beta}_R^{'}\mathscr{X}_i}\right\} - \int\frac{\mathscr{X}md_i(C)}{\Phi(C)}d\mathrm{V}$$

$$+\int\left\{\left(\frac{m}{\Phi(C)}-e^{\boldsymbol{\beta}_R'\mathscr{X}}\right)\mathrm{I}_l-e^{\boldsymbol{\beta}_R'\mathscr{X}}\left(\boldsymbol{\beta}_R'\otimes\mathscr{X}\right)\right\}\Pi d\mathrm{V}R^{-1}\Psi_i(\boldsymbol{\gamma}),$$

and

$$h_i=\mathscr{W}_i\left\{\frac{m_i}{\Phi(C_i)}-e^{D'(\boldsymbol{\beta},\eta)\mathscr{W}_i}\right\}-\int\frac{m\mathscr{W}d_i(C)}{\Phi(C)}d\mathrm{V}-\left\{\int\mathscr{W}\mathscr{W}'e^{D(\boldsymbol{\beta},\eta)'\mathscr{W}}d\mathrm{V}\right\}\Gamma R^{-1}\Psi_i.$$

Proposition 1: *Under conditions (a1) through (a4),* $\widehat{\boldsymbol{\beta}}_R$ *converges in probability to* $\boldsymbol{\beta}_R$. Further, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_R)$ *asymptotically follows a normal distribution with mean zero and a covariance matrix* $A^{-1}\Sigma_g\{A^{-1}\}'$ *where* $A=\mathscr{E}(-\partial g_i/\partial\boldsymbol{\beta}_R')$, $\Sigma_g=\mathscr{E}(g_ig_i')$.

Proposition 2: *Under conditions (a1) through (a4),* $\widehat{\boldsymbol{\beta}}_M$ *converges in probability to* $\boldsymbol{\beta}$. *Further,* $\sqrt{n}(\widehat{\boldsymbol{\beta}}_M-\boldsymbol{\beta})$ *is asymptotically normally distributed with mean zero and a covariance matrix* $B^{-1}\Sigma_h\{B^{-1}\}'$ *where* $B=\mathscr{E}(-\partial h_i/\partial\boldsymbol{\beta}')$, $\Sigma_h=\mathscr{E}(h_ih_i')$.

B Covariance estimation of RC

To develop covariance estimation of the RC estimator, we first illustrate the covariance estimation of $\sqrt{n}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})$ and $\sqrt{n}(\widehat{\Phi}(t)-\Phi(t))$, $\forall t\in[0,\tau]$.

Let $R_n=n^{-1}\sum_{i=1}^n\partial\Psi_i(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}'|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}}$, and $\widehat{\Pi}_i=\partial\mathscr{X}_i/\partial\boldsymbol{\gamma}'|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}}$. The covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})$ can be estimated by $R_n^{-1}n^{-1}\sum_{i=1}^n\Psi_i(\widehat{\boldsymbol{\gamma}})\Psi_i(\widehat{\boldsymbol{\gamma}})'\{R_n^{-1}\}'$. Define that $\widehat{Q}_1(u)=n^{-1}\sum_{i=1}^n\sum_{j=1}^{m_i}I(T_{ij}\le u\le C_i)$, $d\widehat{Q}_2(u)=n^{-1}\sum_{i=1}^n\sum_{j=1}^{m_i}I(T_{ij}=u)$, and

$$\widehat{d}_i(t)=\sum_{j=1}^{m_i}\left[\sum_{T_{(l)}\in[t,\tau]}\frac{I(T_{ij}\le T_{(l)}\le C_i)d\widehat{Q}_2(T_{(l)})}{\widehat{Q}_1(T_{(l)})^2}-\frac{I(t<T_{ij}\le\tau)}{\widehat{Q}_1(T_{ij})}\right],$$

where $T_{(l)}$ are ordered and distinct values of $\{T_{ij}\}_{i=1,\ldots,n;\,j=1\ldots m_i}$. By Wang et al. [13], we can show that the covariance matrix of $\sqrt{n}(\widehat{\Phi}(t)-\Phi(t))$ can be consistently estimated by $\widehat{\Phi}^2(t)n^{-1}\sum_{i=1}^n\widehat{d}_i^{\,2}(t)$.

Denote $\otimes$ as a Kronecker product, and I_a as an identity matrix with size a. Let $\widehat{\mathscr{X}}_i=(1,\mathscr{E}(\boldsymbol{X}_i|\overline{\boldsymbol{W}}_i,\boldsymbol{Z}_i,\widehat{\boldsymbol{\gamma}})',\boldsymbol{Z}_i')'$. Finally, the covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_R)$ can be consistently estimated by $A_n^{-1}\widehat{\Sigma}_g\{A_n^{-1}\}'$ where $A_n=n^{-1}\sum_{i=1}^n\widehat{\mathscr{X}}_i\widehat{\mathscr{X}}_i'e^{\widehat{\boldsymbol{\beta}}_R\widehat{\mathscr{X}}_i}$, and $\widehat{\Sigma}_g=n^{-1}\sum_{i=1}^n\widehat{g}_i\widehat{g}_i'$ with

$$\widehat{g}_i=\widehat{\mathscr{X}}_i\left\{\frac{m_i}{\widehat{\Phi}(C_i)}-e^{\widehat{\boldsymbol{\beta}}_R\widehat{\mathscr{X}}_i}\right\}-\sum_{j=1}^n\frac{\widehat{\mathscr{X}}_jm_j\widehat{d}_i(C_j)}{\widehat{\Phi}(C_j)}$$
$$+\sum_{j=1}^n\left\{\left(\frac{m_j}{\widehat{\Phi}(C_j)}-e^{\widehat{\boldsymbol{\beta}}_R'\widehat{\mathscr{X}}_j}\right)\mathrm{I}_{1+p+q}-e^{\widehat{\boldsymbol{\beta}}_R'\widehat{\mathscr{X}}_j}\left(\widehat{\boldsymbol{\beta}}_R'\otimes\widehat{\mathscr{X}}_j\right)\widehat{\Pi}_j\right\}R_n^{-1}\Psi_i(\widehat{\boldsymbol{\gamma}}).$$

C Covariance estimation of MC

Let $\widehat{D}=D(\widehat{\boldsymbol{\beta}}_M,\widehat{\eta})$, $\widehat{\Gamma}=\partial D/\partial\boldsymbol{\gamma}'|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}}$. The covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_M-\boldsymbol{\beta}_M)$ can be consistently estimated by $B_n^{-1}\widehat{\Sigma}_h\{B_n^{-1}\}'$ where

$$B_n=n^{-1}\sum_{i=1}^n\mathscr{W}_i\left\{\mathscr{W}_i'\otimes\frac{\partial D(\boldsymbol{\beta},\widehat{\eta})}{\partial\boldsymbol{\beta}'}|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_M}\right\}e^{\widehat{D}'\mathscr{W}_i},$$

and $\widehat{\Sigma}_h=n^{-1}\sum_{i=1}^n\widehat{h}_i\widehat{h}_i'$ with

$$\widehat{h}_i = \mathscr{W}_i\left\{\frac{m_i}{\widehat{\Phi}(C_i)} - e^{\widehat{D}'\mathscr{W}_i}\right\} - \sum_{j=1}^{n}\frac{m_j\mathscr{W}_j\widehat{d}_i(C_j)}{\widehat{\Phi}(C_j)} - \left\{\sum_{j=1}^{n}\mathscr{W}_j\mathscr{W}_j'e^{\widehat{D}'\mathscr{W}_j}\right\}\widehat{\Gamma}R_n^{-1}\Psi_i(\widehat{\gamma}).$$

D Proof of RC = MC for regression parameters

Recall that $\mathscr{E}(\boldsymbol{X}_i|\overline{\boldsymbol{W}}_i, \boldsymbol{Z}_i, \boldsymbol{\gamma}) = \eta_0 + \eta_W\overline{\boldsymbol{W}}_i + \eta_Z\boldsymbol{Z}_i$, where η_0, η_W and η_Z are functions of $\boldsymbol{\gamma}$. Let $\mathbf{0}_{r\times s}$ be a $r\times s$ matrix of 0's. With simple algebra, we can write $\mathscr{X}_i = H\mathscr{W}_i, \forall i = 1, \ldots, n$ where

$$H = \begin{pmatrix} 1 & \mathbf{0}_{1\times p} & \mathbf{0}_{1\times q} \\ \eta_0 & \eta_W & \eta_Z \\ \mathbf{0}_{q\times 1} & \mathbf{0}_{q\times p} & \mathrm{I}_q \end{pmatrix}.$$

Since H remains the same for $i = 1, \ldots, n$, for any fixed $\boldsymbol{\gamma}$, eq. (4) can be written as

$$n^{-1}\sum_{i=1}^{n}\mathscr{W}_i\left\{m_i\widehat{\Phi}^{-1}(C_i) - e^{(H'\widehat{\boldsymbol{\beta}}_R)'\mathscr{W}_i}\right\} = 0. \tag{8}$$

Recall that $\widehat{\boldsymbol{b}}_N$ is the unique root of the equations with form

$$n^{-1}\sum_{i=1}^{n}\mathscr{W}_i\left\{m_i\widehat{\Phi}^{-1}(C_i) - e^{\boldsymbol{b}'\mathscr{W}_i}\right\} = 0.$$

It is easy to see that eq. (8) has the same form as the above equation. Thus, we have $H'\widehat{\boldsymbol{\beta}}_R = \widehat{\boldsymbol{b}}_N$. Besides, by definition $\widehat{\boldsymbol{b}}_N = D(\widehat{\boldsymbol{b}}_M, \boldsymbol{\gamma})$ for any fixed $\boldsymbol{\gamma}$. Therefore, we have

$$\begin{pmatrix} \widehat{\beta}_{R,0} + \eta_0\widehat{\boldsymbol{\beta}}_{R,X} \\ \eta_W'\widehat{\beta}_{R,X} \\ \eta_Z'\widehat{\boldsymbol{\beta}}_{R,X} + \widehat{\boldsymbol{\beta}}_{R,Z} \end{pmatrix} = H'\widehat{\boldsymbol{\beta}}_R = D(\widehat{\boldsymbol{b}}_M, \boldsymbol{\gamma}) = \begin{pmatrix} \widehat{\beta}_{M,0} + \eta_0\widehat{\beta}_{M,X} + \frac{1}{2}\widehat{\beta}_{M,X}'\Sigma\boldsymbol{\beta}_{M,X} \\ \eta_W'\widehat{\boldsymbol{\beta}}_{M,X} \\ \eta_Z'\widehat{\boldsymbol{\beta}}_{M,X} + \widehat{\boldsymbol{\beta}}_{M,Z} \end{pmatrix},$$

for any fixed $\boldsymbol{\gamma}$. The above equation implies that $\widehat{\beta}_{R,0} = \widehat{\beta}_{M,0} + \widehat{\boldsymbol{\beta}}_{M,X}'\Sigma\widehat{\boldsymbol{\beta}}_{M,X}/2$, $\widehat{\boldsymbol{\beta}}_{R,X} = \widehat{\boldsymbol{\beta}}_{M,X}$, and $\widehat{\boldsymbol{\beta}}_{R,Z} = \widehat{\boldsymbol{\beta}}_{M,Z}$. Hence the proof is complete.

References

1. Fleming TR, Harrington DP. Counting processes and survival analysis. New York: John Wiley & Sons, 1991.
2. Morgan WJ, Butler SM, Johnson CA, Colin AA, FitzSimmons SC, Geller DE, et al. Epidemiologic study of cystic fibrosis: design and implementation of a prospective, multicenter, observational study of patients with cystic fibrosis in the US and Canada. Pediatr Pulmonol 1999;28:231–41.
3. Hu XJ, Lagakos SW, Lockhart RA. Generalized least squares estimation of the mean function of a counting process based on panel counts. Stat Sinica 2009;19:561–80.
4. Hu XJ, Lawless JF. Estimation of rate and mean functions from truncated recurrent event data. J Am Stat Assoc 1996;91: 300–10.
5. Lawless JF, Hu J, Cao J. Methods for the estimation of failure distributions and rates from automobile warranty data. Lifetime Data Anal 1995;1:227–40.
6. Lin DY, Sun W, Ying Z. Nonparametric estimation of the gap time distribution for serial events with censored data. Biometrika 1999;86:59–70.
7. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. Biometrika 1981;68: 373–9.
8. Hu XJ, Lagakos SW. Nonparametric estimation of the mean function of a stochastic process with missing observations. Lifetime Data Analysis 2007;13:51–73.
9. Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. Technometrics 1995;37:158–68.
10. Lancaster T, Intrator O. Panel data with survival: hospitalization of HIV-positive patients. J Am Stat Assoc 1998;93:46–53.

11. Nielsen GG, Gill RD, Andersen PK, Sørensen TI. A counting process approach to maximum likelihood estimation in frailty models. Scand J Stat 1992;19:25–43.
12. Kalbfleisch JD, Schaubel DE, Ye Y, Gong Q. An estimating function approach to the analysis of recurrent and terminal events. Biometrics 2013;69:366–74.
13. Wang MC, Qin J, Chiang CT. Analyzing recurrent event data with informative censoring. J Am Stat Assoc 2001;96:1057–65.
14. Wang MC, Huang CY. Statistical inference methods for recurrent event processes with shape and size parameters. Biometrika 2014;101:553–66.
15. Clark LC, Combs GF, Turnbull BW, Slate EH, Chalker DK, Chow J, et al. Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin: a randomized controlled trial. J Am Med Assoc 1996;276:1957–63.
16. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. London: Chapman & Hall, 2006.
17. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika 1982;69:331–42.
18. Wang CY, Hsu L, Feng ZD, Prentice RL. Regression calibration in failure time regression. Biometrics 1997;53:131–45.
19. Nakamura T. Proportional hazards model with covariates subject to measurement error. Biometrics 1992;48:829–38.
20. Wu L. A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. J Am Stat Assoc 2002;97:955–64.
21. Liu W, Wu L. Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. Biometrics 2007;63:342–50.
22. Wu L, Liu W, Hu XJ. Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors. Biometrics 2010;66:327–35.
23. Jiang W, Turnbull BW, Clark LC. Semiparametric regression models for repeated events with random effects and measurement error. J Am Stat Assoc 1999;94:111–24.
24. Cook RJ, Lawless JF. The statistical analysis of recurrent events. New York: Springer, 2007.
25. Balakrishnan N, Peng Y. Generalized gamma frailty model. Stat Med 2006;25:2797–816.
26. Mazroui Y, Mathoulin-Pelissier S, Soubeyran P, Rondeau V. General joint frailty model for recurrent event data with a dependent terminal event: application to follicular lymphoma data. Stat Med 2012;31:1162–76.
27. Zeng D, Ibrahim J, Chen M, Hu K, Jia C. Multivariate recurrent events in the presence of multivariate informative censoring with applications to bleeding and transfusion events in myelodysplastic syndrome. J Biopharm Stat 2014;24:429–42.
28. Wang CY. Robust sandwich covariance estimation for regression calibration estimator in cox regression with measurement error. Stat Probab Lett 1999;45:371–8.
29. Stefanski LA. The effects of measurement error on parameter estimation. Biometrika 1985;72:583–92.
30. Huang Y, Wang CY. Cox regression with accurate covariates unascertainable: a nonparametric-correction approach. J Am Stat Assoc 2000;95:1209–19.
31. Huber PJ. Robust statistics. New Jersey: John Wiley & Sons, 2009.

Supplemental Material: The online version of this article (DOI: 10.1515/ijb-2016-0001) offers supplementary material, available to authorized users.