Published in final edited form as:

Biometrics. 2018 September ; 74(3): 966–976. doi:10.1111/biom.12857.

Methods for multivariate recurrent event data with measurement error and informative censoring

Hsiang Yu¹, Yu-Jen Cheng¹, and Ching-Yun Wang²

¹Institute of Statistics, National Tsing-Hua University, General Building III, Hsin-Chu 300, Taiwan

²Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, U.S.A.

Summary

In multivariate recurrent event data regression, observation of recurrent events is usually terminated by other events that are associated with the recurrent event processes, resulting in informative censoring. Additionally, some covariates could be measured with errors. In some applications, an instrumental variable is observed in a subsample, namely a calibration sample, which can be applied for bias correction. In this article, we develop two non-parametric correction approaches to simultaneously correct for the informative censoring and measurement errors in the analysis of multivariate recurrent event data. A shared frailty model is adopted to characterize the informative censoring and dependence among different types of recurrent events. To adjust for measurement errors, a nonparametric correction method using the calibration sample only is proposed. In the second approach, the information from the whole cohort is incorporated by the generalized method of moments. The proposed methods do not require the Poisson-type assumption for the multivariate recurrent event process and the distributional assumption for the frailty. Moreover, we do not need to impose any distributional assumption on the underlying covariates and measurement error. Both methods perform well, but the second approach improves efficiency. The proposed methods are applied to the Nutritional Prevention of Cancer trial to assess the effect of selenium treatment on the recurrences of basal cell carcinoma and squamous cell carcinoma.

Keywords

Generalized method of moments; Informative censoring; Instrumental variable; Measurement error; Multivariate recurrent event data; Surrogate covariate

1. Introduction

Recurrent event data arise when a particular event of interest repeatedly occurs in a subject. In practice, one subject may experience more than one type of event, resulting in

Correspondence to: Yu-Jen Cheng.

Supplementary Materials

Web appendices referenced in Sections 3, 4, and 5, and the R script to obtain the proposed estimators are available with this paper at the *Biometrics* website on Wiley Online Library.

multivariate recurrent event data. For example, children with asthma may return to the hospitals due to physician office visit or hospitalization (Schaubel et al., 1996), patients after receiving bone marrow transplantation may experience bacterial, fungal or viral infections (Zhu et al., 2010), and different types of febrile nonhemolytic transfusion reaction may occur in hematology/oncology patients (Zhao et al., 2012).

Analysis for multivariate recurrent event data may focus on the estimation of an intensity function (Abu-Libdeh et al., 1990) or a rate function (Cai and Schaubel, 2004). Compared with the estimation of the intensity function, estimation on the rate function is more appealing recently because it provides a direct interpretation when the interest lies on a population aspect. Methods for estimating the rate function (Cai and Schaubel, 2004) have been extensively studied, yet most of them require the assumption of independent censoring, which may be invalid practically. Indeed, informative censoring occurs when the multivariate recurrent event processes are interrupted by other events that are correlated with the recurrent event processes. For example, in our motivating data set collected from the nutritional prevention of Cancer (NPC) study (Clark et al., 1996), the patients might repeatedly experience two types of skin cancer: basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). According to Clark et al. (1996), the SCC and BCC events could be terminated by several reasons, including death. Furthermore, we have found in our data analysis that the patients in different risk sets had distinct occurrence rates of SCC and BCC. Hence, the assumption of independent censoring may not hold in the NPC data. To incorporate informative censoring in multivariate recurrent event data analysis, some authors consider shared frailty models (Ning et al., 2017; Bedair et al., 2016). To be specific, a shared frailty model allows each type of recurrent event process to be correlated with the censoring time through a type-specific frailty variable. The existing methods based on the shared frailty model assume either a specific distribution for the frailties (Bedair et al., 2016) or the Poisson model for the multivariate recurrent event processes conditioning on the frailties (Ning et al., 2017). Although these two assumptions can be relaxed simultaneously by a recent work (Xu et al., 2017), all of these methods require that the covariates are accurately measured, which may not be valid in practice.

When covariates are measured with errors, neglecting measurement errors usually yields inconsistent estimation for the regression parameters in the failure time models (Prentice, 1982). There have been numerous correction methods dealing with measurement error problems in the literature and can roughly be classified as a parametric or nonparametric correction. The parametric correction approaches, which impose a distributional assumption on the error term or true covariate, include corrected score (Nakamura, 1992), conditional score (Tsiatis and Davidian, 2001), and regression calibration (Prentice, 1982). On the other hand, the non-parametric correction approaches relax these distributional assumptions with the assistance of additional information about the measurement errors, such as replicated surrogates (Huang and Wang, 2000) and instrumental variables (Song and Wang, 2014). An instrumental variable is a variable that is related to the error-prone covariate but independent of the measurement error and outcome given the true covariate (Carroll et al., 2006, Chapter 6). In some practical cases, replicates may not be available, but only an instrumental variable is accessible. For instance, in the NPC trial study, the major goal is to evaluate the efficacy of the selenium (Se) supplement in preventing the recurrences of SCC and BCC, and

baseline Se level, age, body mass index (BMI), gender, and smoking status are included to adjust for the effect of Se treatment. Nevertheless, the true value of Se level can never be measured because of imprecise instrument and biological fluctuation. Thus, the observed measurements of Se level are subject to measurement errors. Replicates did not exist since each patient in the study had only one Se measurement at baseline. However, a subset of individuals had their Se measurements after baseline, which would be used as the instrumental variable in our application. To the best of our knowledge, only a few methods consider the measurement error problem in recurrent event data analysis. Under the assumption of independent censoring, Jiang et al. (1999) corrected the naive estimator by imposing a conditional normal distribution on the true covariate given the observed one; Hu and Lin (2004) exploited the use of replicates to correct the partial score function with the assumption that the error terms are symmetrically distributed; Yi and Lawless (2012) applied a simulation extrapolation method to the naive estimating equation, but their method leads to an inconsistent estimator if the extrapolation function is invalid. These methods may not be appropriate for the NPC data due to the violation of the independent censoring assumption. To account for informative censoring with measurement error, Yu et al. (2016) proposed a parametric correction approach in which both the covariates and the error terms are assumed to be normally distributed.

In this paper, we consider a semiparametric estimation procedure for multivariate recurrent event data subject to informative censoring and measurement error. We propose a shared frailty model to formulate the association between the multivariate recurrent event processes and the informative censoring. The shared frailty model is also used to formulate the dependence among distinct types of recurrent events. Additionally, we consider a general nonparametric model for the instrumental variable. Since the instrumental variable is only observed in a subsample as in the NPC study, the assumption of missing at random is imposed. Based on this assumption with our modelling framework, we can construct a set of estimating equations to estimate the regression coefficients. However, this approach only uses the calibration sample, which may lead to some efficiency loss when the proportion of calibration sample is small. To improve the efficiency, we incorporate the information from the non-calibration sample in the second approach. Because the number of estimating equations exceeds the number of parameters in this approach, we exploit the generalized method of moments (GMM) with optimal weight to enhance the efficiency of the proposed estimator. The proposed methods have the following advantages over the existing works. First, we extend Xu et al. (2017)'s result from univariate recurrent event data to multivariate recurrent event data so that both the Poisson-type assumption for the multivariate recurrent events and the distributional assumption for the frailty are not required. Thus, the association between the informative censoring and the multivariate recurrent event processes is unspecified. In addition, the dependence among different types of recurrent events is also allowed to be unspecified. Second, no distributional assumptions of the underlying covariates and the error terms are assumed in the estimating procedure.

The article is organized as follows. Section 2 describes the statistical models for multivariate recurrent events and measurement errors. In section 3, we develop the estimating procedures of the proposed methods. Section 4 investigates the relative efficiency and finite sample performance of the proposed methods through a simulation study. The proposed methods are

applied to the NPC trial data to evaluate the effect of Se supplement on the recurrences of SCC and BCC in Section 5. Some concluding remarks are made in Section 6. Supplementary Materials provide the regularity conditions and technical proofs.

2. Statistical Modelling

2.1 Recurrent event model

Suppose that each subject may experience K types of recurrent events. Let τ be a prespecified constant that the recurrent events could be potentially observed up to τ , $N_k(t)$ be the number of the k th type recurrent event occurring up to time t within $[0, \tau]$, and ν_k be a non-negative frailty variable for $k = 1, \dots, K$. Let $\mathbf{X}$ be a $p \times 1$ vector of covariates that can be measured with errors and $\mathbf{Z}$ be a $q \times 1$ vector of covariates that are precisely measured. Suppose that all types of recurrent events are censored by the same time $Y \equiv \min(Y^*, \tau)$, where Y^* denotes the informative censoring. We consider the following model assumptions for the k th type recurrent event process:

(M1) $N_k(\cdot)$ and Y are conditionally independent given $(\nu_k, \mathbf{X}, \mathbf{Z})$.

(M2) Conditioning on $(\nu_k, \mathbf{X}, \mathbf{Z})$, the rate function of $N_k(t)$ is defined as

$$\frac{d}{dt} \mathcal{E}(N_k(t) | \nu_k, \mathbf{X}, \mathbf{Z}) = \nu_k \lambda_{0k}(t) e^{\beta'_X \mathbf{X} + \beta'_Z \mathbf{Z}}, t \in [0, \tau], \text{ where } \mathcal{E} \text{ denotes expectation and } \lambda_{0k}(t) \text{ is an unspecified baseline rate function.}$$

Model (M1) implies that $N_k(t)$ and Y are mutually dependent where the dependence is characterized by the frailty variable ν_k . In addition, the K types of recurrent event processes are allowed to be correlated with each other. The association between each type of recurrent event process and the censoring time can be arbitrary because the frailty distribution is unspecified throughout the estimating procedure. The dependence among the multivariate recurrent event processes is arbitrary as well. Furthermore, unlike the recurrent event model illustrated in Wang et al. (2001), we do not impose the assumption of a Poisson model on the recurrent event processes. Hence, the k th type events within a subject are allowed to be dependent even conditioning on the frailty ν_k for $k = 1, \dots, K$. We assume that $\mathcal{E}(\boldsymbol{\nu} | \mathbf{X}, \mathbf{Z}) = \mathcal{E}(\boldsymbol{\nu}) = \boldsymbol{\mu}$, where $\boldsymbol{\nu} = (\nu_1, \dots, \nu_K)'$ and $\boldsymbol{\mu} = (\mu_{\nu_1}, \dots, \mu_{\nu_K})'$ are the $K \times 1$ vectors of frailty variables and unknown parameters. To ensure identifiability, we assume that $\Lambda_{0k}(\tau) = 1$. Under (M2), the mean function of the k th type of recurrent events is given by

$$\mathcal{E}\{N_k(t) | \mathbf{X}, \mathbf{Z}\} = \mu_{\nu_k} \Lambda_{0k}(t) e^{\beta'_X \mathbf{X} + \beta'_Z \mathbf{Z}}, t \in [0, \tau], k = 1, \dots, K, \quad (1)$$

where $\Lambda_{0k}(t) \equiv \int_0^t \lambda_{0k}(u) du$. For simplicity, assumption (M2) assumes the covariates and corresponding effects remain the same for different types of recurrent events. In fact, the proposed estimating procedure developed below can be applied straightforwardly if there exist type-specific covariates and effects. In that case, for instance, one can replace $(\mathbf{X}', \mathbf{Z}')'$ and $(\beta'_X, \beta'_Z)'$ by $(\mathbf{X}'_k, \mathbf{Z}'_k)'$ and $(\beta'_{Xk}, \beta'_{Zk})'$ in assumption (M2), respectively.

2.2 Measurement error model and instrumental variable

Suppose that the true value of $\mathbf{X}$ is unobservable and one error-contaminated measurement $\mathbf{W}$ is available for each subject. In addition, an instrumental variable (IV) $\mathbf{V}$ is observed in the calibration sample. Let $I(V)$ be the indicator of observing $\mathbf{V}$. Suppose there are n independent individuals in the cohort and let subscript i be the index of a subject. For the i th subject, let m_{ik} be the total number of events of type k that occurred before Y_i and T_{ikj} be the j th event time of type k . Then, the observed data consist of $\{T_{ik1}, \dots, T_{ikm_{ik}}, m_{ik}, Y_i, \mathbf{W}_i, \mathbf{V}_i, i=1, \dots, n, k=1, \dots, K\}$. The model assumptions imposed on the surrogate and IV are given as follows:

- (E1) The surrogate $\mathbf{W}$ follows $\mathbf{W} = \mathbf{X} + \mathbf{U}$, in which $\mathbf{U}$ is a vector of the random errors with zero means and are independent of $(\mathbf{X}, \mathbf{Z})$. More importantly, the distributions of $\mathbf{X}$ and $\mathbf{U}$ are unspecified.
- (E2) The IV $\mathbf{V}$ satisfies that $\mathbf{V} = g(\mathbf{X}, \mathbf{Z}, \mathbf{e})$, where $g(\cdot)$ is an unknown function and $\mathbf{e}$ is a vector of random variables which are independent of $\mathbf{U}$ and conditionally independent of $\{Y, N_1(t), \dots, N_K(t)\}$ given $(\mathbf{X}, \mathbf{Z})$.
- (E3) $\mathbf{V}$ is missing at random (MAR) such that $Pr(I(V)=1 | T_{11}, \dots, T_{Km_K}, m_1, \dots, m_K, Y, \mathbf{V}, \mathbf{W}, \mathbf{Z}) = Pr(I(V)=1 | m_1, \dots, m_K, Y, \mathbf{W}, \mathbf{Z})$.

Assumption (E1) is generally referred to as the classical measurement error model, which is commonly used in the measurement error literatures (Carroll et al., 2006). No distributional assumptions on $\mathbf{X}$ and $\mathbf{U}$ are needed in our framework, and thus the proposed methods are referred to as nonparametric correction approaches. We further assume that the measurement error is non-differential, which means that $\mathbf{W}$ provides no additional information about the recurrent event processes when the true covariate $\mathbf{X}$ is given (Carroll et al., 2006, Chapter 2). In assumption (E2), we consider a general nonparametric model for $\mathbf{V}$. It implies that $\mathbf{V}$ can be dependent on both of $\mathbf{X}$ and $\mathbf{Z}$ but is conditionally independent of $\mathbf{W}$ and $\{Y, N_1(t), \dots, N_K(t)\}$ when $(\mathbf{X}, \mathbf{Z})$ is given. This model includes many special cases such as the repeated measurements of $\mathbf{W}$, non-linear instruments, and non-parametric instruments. Note that the dimension of $\mathbf{V}$ can be larger than $\mathbf{X}$. For the sake of convenience, we assume that $\mathbf{V}$ has the same dimension as $\mathbf{X}$ in the inference. Assumption (E3) indicates that the missing mechanism of $\mathbf{V}$ only depends on the variables that are completely observed. In this article, we focus on the estimation of the regression parameters $(\beta'_X, \beta'_Z)'$ in model (1) based on the observed data.

3. Estimating methods

In this section, we begin with an estimating procedure for the regression parameters $(\beta'_X, \beta'_Z)'$ when $\mathbf{X}$ is observed. Then, we extend the estimating procedure to the situation when $\mathbf{X}$ is measured with errors.

First, we consider the estimation of $\Lambda_{0k}(t)$. For the univariate recurrent event data ($K=1$), under the assumption of a Poisson process, Wang et al. (2001) proposed a conditional likelihood approach to estimate $\Lambda_{0k}(t)$ and the estimator is $\hat{\Lambda}_{0k}(t) = \prod_{s_{k(l)} > t} (1 - n_{k(l)}/N_{k(l)})$, where $\{s_{k(l)}\}$ are the ordered and distinct values of $\{T_{ik1}, \dots, T_{ikm_{ik}}\}_{i=1, \dots, n}$.

$n_{k(l)} = \sum_{i=1}^n \sum_{j=1}^{m_{ik}} I(T_{ikj} = s_{(l)})$, and $N_{k(l)} = \sum_{i=1}^n \sum_{j=1}^{m_{ik}} I(T_{ikj} \leq s_{(l)} \leq Y_i)$, $k = 1, \dots, K$. In fact, even if the assumption of a Poisson process is invalid, $\hat{\Lambda}_{0k}(t)$ is still consistent to $\Lambda_{0k}(t)$ (Xu et al., 2017). In this work, we extend this result to multivariate recurrent event data. Specifically, by the approximation of the product limit estimator, it can be shown that $\hat{\Lambda}_{0k}(t)$ is asymptotically equivalent to the estimator $\exp\{-\int_t^\tau \hat{R}_k^{-1}(u) d\hat{Q}_k(u)\}$, where

$\hat{R}_k(t) = n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_{ik}} I(T_{ikj} \leq t \leq Y_i)$ and $\hat{Q}_k(t) = n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_{ik}} I(T_{ikj} \leq t)$. By the strong law of large number with assumptions (M1) and (M2), $\hat{R}_k(t)$ and $\hat{Q}_k(t)$ converge almost surely to the limiting functions

$$R_k(t) = \mathcal{E}\{\sum_{j=1}^{m_k} I(T_{kj} \leq t \leq Y)\} = \mathcal{E}[\mathcal{E}\{I(Y \geq t)N_k(t)|\nu_k, Y, X, Z\}] = G_k(t)\Lambda_{0k}(t) \text{ and}$$

$$Q_k(t) = \mathcal{E}\{\sum_{j=1}^{m_k} I(T_{kj} \leq t)\} = \mathcal{E}[\mathcal{E}\{N_k(t \wedge Y)|\nu_k, Y, X, Z\}] = \int_0^t G_k(u) d\Lambda_{0k}(u), \forall t \in [0, \tau], \text{ where}$$

$$G_k(u) = \mathcal{E}\{I(Y \geq u)\nu_k e^{\beta'_X X + \beta'_Z Z}\} \text{ and } \wedge \text{ denotes the minimum. Thus, } \exp\{-\int_t^\tau \hat{R}_k^{-1}(u) d\hat{Q}_k(u)\}$$

converges almost surely to $\Lambda_{0k}(t)$ with the fact $\Lambda_{0k}(\tau) = 1$. It is worth noting that the estimation of $\Lambda_{0k}(t)$ does not involve ν_k and the covariates, so it is invariant even if the covariates are subject to measurement errors.

With the consistent estimator of $\Lambda_{0k}(t)$, a set of estimating equations for the regression parameters $(\beta'_X, \beta'_Z)'$ can be defined as

$$n^{-1} \sum_{i=1}^n \sum_{k=1}^K \begin{pmatrix} e_k \\ X_i \\ Z_i \end{pmatrix} \left[m_{ik} \hat{\Lambda}_{0k}^{-1}(Y_i) - \exp(b'_0 e_k + b'_X X_i + b'_Z Z_i) \right] = 0,$$

where $b_0 = (b_{01}, \dots, b_{0K})$ and e_k is a $K \times 1$ vector of zeros except the k th component equals to one. The estimating equation given above is unbiased because

$$\mathcal{E}[\mathcal{E}\{N_k(Y)\Lambda_{0k}^{-1}(Y)|\nu_k, Y, X, Z\}|X, Z] = \exp(\beta_{0k} + \beta'_X X + \beta'_Z Z), k = 1, \dots, K, \quad (2)$$

where $\beta_{0k} = \log(\mu_{\nu_k})$, and the second equality holds under assumption (M1) and model (1).

However, the above estimating equation is not applicable since X is unobserved. A naive approach is replacing X with the surrogate W in the estimating equation, but it is known that the resulting estimator is usually inconsistent (Carroll et al., 2006). In the following sections, we introduce two estimating methods to correct the measurement errors by using the surrogate W and the IV V . In Section 3.1, we propose a simple nonparametric correction method which uses the calibration sample only. Then, in Section 3.2, we propose a method which improves the efficiency of the simple nonparametric correction estimator by incorporating the information from the non-calibration sample. In contrast to the naive estimator, both proposed estimators are shown to be consistent for the regression parameters $(\beta'_X, \beta'_Z)'$.

3.1 Simple non-parametric correction method

When the IVs are available for all subjects, Huang and Wang (2000) proposed a nonparametric correction method under a Cox regression model. Based on assumptions (E1) and (E2), their proposed estimators can be shown to be consistent. We extend their works to recurrent event data and propose a new correction approach when $\mathbf{V}$ is only available in the calibration sample.

Under assumption (E1), we have

$\mathcal{E}\{\exp(\beta_{0k} + \beta'_X \mathbf{W} + \beta'_Z \mathbf{Z}) | \mathbf{X}, \mathbf{Z}\} = \exp(\beta_{0k} + \beta'_X \mathbf{X} + \beta'_Z \mathbf{Z}) \mathcal{E}\{\exp(\beta'_X \mathbf{U})\}$, where the value of $\mathcal{E}\{\exp(\beta'_X \mathbf{U})\}$ is only dependent on β_X and the distributional assumption of $\mathbf{U}$. On the other hand, it can be found from equation (2) that this result is almost equivalent to the conditional expectation of $m_k \Lambda_{0k}^{-1}(Y)$ given $(\mathbf{X}, \mathbf{Z})$ except for the intercept term. Thus, together with assumption (E2), it can be shown that

$$\begin{aligned} & \mathcal{E}\left[\left(\begin{array}{c} 1 \\ \mathcal{E}(\mathbf{V} | \mathbf{X}, \mathbf{Z}) \\ \mathbf{Z} \end{array}\right) \mathcal{E}\left[m_k \Lambda_{0k}^{-1}(Y) - \exp(\beta_{0k}^* + \beta'_X \mathbf{W} + \beta'_Z \mathbf{Z}) | \mathbf{X}, \mathbf{Z}\right]\right] \\ &= 0, k = 1, \dots, K, \end{aligned}$$

where $\beta_{0k}^* = \beta_{0k} - \log \mathcal{E}\{\exp(\beta'_X \mathbf{U})\}$. From the above equation, it implies that the estimation of $(\beta'_X, \beta'_Z)'$ can be obtained if both of $\mathbf{V}$ and $\mathbf{W}$ are available. However, this result can not be applied to our case since $\mathbf{V}$ is only observed in the calibration sample. To deal with the missingness, we consider the assumption of missing at random (E3). Denote $\beta^* = (\beta_{01}^*, \dots, \beta_{0K}^*, \beta'_X, \beta'_Z)'$, $\Lambda_0 = (\Lambda_{10}, \dots, \Lambda_{K0})$ and $\mathcal{W}_{ik} = (e_k', \mathbf{W}_i', \mathbf{Z}_i')'$. By using the inverse probability weighting, we have $\mathcal{E}\{\Phi_S(\beta^*; \pi, \Lambda_0)\} = 0$, where

$$\Phi_S(\mathbf{b}; \pi, \Lambda_0) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K \frac{\Delta_i}{\pi_i} \begin{pmatrix} e_k \\ \mathbf{V}_i \\ \mathbf{Z}_i \end{pmatrix} \left[\frac{m_{ik}}{\Lambda_{0k}(Y_i)} - \exp(\mathbf{b}' \mathcal{W}_{ik}) \right],$$

$\mathbf{b} = (b_{01}, \dots, b_{0K}, \mathbf{b}'_X, \mathbf{b}'_Z)'$ and $\pi_i = Pr(i = 1 | m_{i1}, \dots, m_{iK}, Y_i, \mathbf{W}_i, \mathbf{Z}_i)$. When π is known, an estimator of $(\beta'_X, \beta'_Z)'$ can be obtained by solving $\Phi_S(\mathbf{b}; \pi, \Lambda_0) = 0$ with replacing the unknown Λ_0 by its consistent estimator $\hat{\Lambda}_0$. In practice, π is usually unknown and needs to be estimated. In this article, we impose a parametric model on π . Specifically, we assume follows a logistic model with mean $\pi(\theta; \mathbf{a}) = H(\mathbf{a}' \theta)$, where $H(u) = \{1 + \exp(-u)\}^{-1}$, $\theta = (1, m_1, \dots, m_K, Y, \mathbf{W}, \mathbf{Z})$ and $\mathbf{a}$ is the corresponding parameter. Then, we can estimate π by its maximum likelihood estimator $\hat{\pi} = \pi(\hat{\theta}; \hat{\mathbf{a}})$, where $\hat{\mathbf{a}}$ is solved from $n^{-1} \sum_{i=1}^n \mathcal{O}_i \{\Delta_i - H(\mathbf{a}' \mathcal{O}_i)\} = 0$. Finally, a simple non-parametric correction (SNC) estimator, $\hat{\beta}_S$, is obtained by solving $\Phi_S(\mathbf{b}; \hat{\pi}, \hat{\Lambda}_0) = 0$. The asymptotic properties of $\hat{\beta}_S$ are given in Theorem 1.

Theorem 1—Under the regularity conditions (a1)–(a6) in Web Appendix A, the SNC estimator $\hat{\beta}_S$ exists and converges almost surely to the limit β^* . In addition, $\sqrt{n}(\hat{\beta}_S - \beta^*)$ is asymptotically normal with mean zero and covariance matrix $J^{-1}(\beta^*)\Omega_S(\beta^*)J^{-1}(\beta^*)$, where $J(b) \equiv \mathcal{E}\{-\Phi_S(b; \pi, \Lambda_0)/b'\}$, $\Omega_S(b) \equiv \mathcal{E}\{\varphi(b; \pi, \Lambda_0)\varphi'(b; \pi, \Lambda_0)\}$ and φ is defined in Web Appendix A.

The covariance matrix of $\sqrt{n}(\hat{\beta}_S - \beta^*)$ can be estimated by $b \hat{J}^{-1}(\hat{\beta}_S) \hat{\Omega}_S(\hat{\beta}_S) \hat{J}^{-1}(\hat{\beta}_S)$, where $\hat{J}(b) = -n^{-1} \sum_{i=1}^n \varphi_i(b; \hat{\pi}, \hat{\Lambda}_0)/b'$ and $\hat{\Omega}_S(b) = n^{-1} \sum_{i=1}^n \{\varphi_i(b; \hat{\pi}, \hat{\Lambda}_0)\varphi_i'(b; \hat{\pi}, \hat{\Lambda}_0)\}$. Note that the SNC estimator $\hat{\beta}_S$ is consistent for the regression parameter (β_X', β_Z') , except for the intercepts $(\beta_{01}, \dots, \beta_{0K})$.

3.2 GMM non-parametric correction method

The SNC method only uses the calibration sample, and thus may be inefficient. To improve efficiency, we incorporate the information from both the calibration and non-calibration samples by constructing the estimating procedure based only on $\mathbf{W}$. Further, we combine the proposed estimating procedure with the SNC approach by using the GMM method (Song and Wang, 2014, in Cox regression).

More precisely, under model (1) and assumption (E1), we have

$$\mathcal{E}\{W m_k \Lambda_{0k}^{-1}(Y) | X, Z\} = X \exp(\beta_{0k} + \beta_X' X + \beta_Z' Z) \text{ and}$$

$$\mathcal{E}\{W \exp(\beta_{0k}^* + \beta_X' W + \beta_Z' Z) | X, Z\} = X \exp(\beta_{0k} + \beta_X' X + \beta_Z' Z) + \mathcal{E}\{U \exp(\beta_X' U)\} \exp(\beta_{0k}^* + \beta_X' X + \beta_Z' Z)$$

Therefore, by simple linear algebra, it can be shown that

$$\mathcal{E}\left\{\left(\frac{1}{W} \frac{m_k}{\Lambda_{0k}(Y)} - \left(\frac{1}{W - c} \right) \exp(\beta_{0k}^* + \beta_X' W + \beta_Z' Z)\right)\right\} = 0, k = 1, \dots, K, \quad (3)$$

where $c \equiv \mathcal{E}\{U \exp(\beta_X' U)\} / \mathcal{E}\{\exp(\beta_X' U)\}$. If the distribution of U is specified, for example, U follows a normal distribution with mean zero and variance Σ_U , then $c = \Sigma_U \beta_X$. Thus, equation (3) is reduced to

$\mathcal{E}\{(1, W', Z')' m_k \Lambda_{0k}^{-1}(Y) - (1, W' - \beta_X' \Sigma_U, Z')' \exp(\beta_{0k}^* + \beta_X' W + \beta_Z' Z)\} = 0$, which can be referred to as a parametric correction method. In this work, we consider a non-parametric estimation for c to avoid specifying the distribution of U . To do so, we start from equation (3) and c can be solved as $c = -\mathcal{E}\left[W\{m_k \Lambda_{0k}^{-1}(Y) - \exp(\beta_X' W)\}\right] / \mathcal{E}\{\exp(\beta_X' W)\}$. It suggests that c can be estimated based on $(W', Z')'$ and the result from the SNC approach.

Specifically, the correction term c is estimated by

$$\hat{c} = -\sum_{i=1}^n \sum_{k=1}^K \frac{\Delta_i}{\pi_i} W_i \left\{ m_{ik} \hat{\Lambda}_{0k}^{-1}(Y_i) - \exp(\hat{\beta}_S' W_{ik}) \right\} / \sum_{i=1}^n \sum_{k=1}^K \frac{\Delta_i}{\pi_i} \exp(\hat{\beta}_S' W_{ik}), \text{ and thus an estimating equation for } (\beta_X', \beta_Z')' \text{ can be constructed based on equation (3) with } \Lambda_0 \text{ and } c$$

replaced by $\hat{\Lambda}_0$ and $\hat{c}$, respectively. Note that the estimating equation involves both calibration and non-calibration samples, although only the calibration sample is used to estimate the correction term c . Further, we can combine this result with the SNC approach to improve the efficiency of the SNC estimator as follows. Define that

$$\Phi_G(\mathbf{b}; c, \pi, \Lambda_0) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K \begin{pmatrix} \frac{\Delta_i}{\pi_i} \mathbf{V}_i \left\{ m_{ik} \Lambda_{0k}^{-1}(Y_i) - \exp(\mathbf{b}' \mathcal{W}_{ik}) \right\} \\ \mathbf{W}_i m_{ik} \Lambda_{0k}^{-1}(Y_i) - \{ \mathbf{W}_i - c \} \exp(\mathbf{b}' \mathcal{W}_{ik}) \\ \begin{pmatrix} \mathbf{e}_k \\ \mathbf{Z}_i \end{pmatrix} \left\{ m_{ik} \Lambda_{0k}^{-1}(Y_i) - \exp(\mathbf{b}' \mathcal{W}_{ik}) \right\} \end{pmatrix}.$$

We can not directly solve $\Phi_G(\mathbf{b}; \hat{c}, \hat{\pi}, \hat{\Lambda}_0) = 0$ since the number of equations in $\Phi_G(\mathbf{b}; \hat{c}, \hat{\pi}, \hat{\Lambda}_0)$ is larger than the number of the unknown parameters. By adopting the GMM technique, $\hat{\beta}_G(A)$ is obtained by minimizing the quadratic form $\{ \Phi_G(\mathbf{b}; \hat{c}, \hat{\pi}, \hat{\Lambda}_0) \}' A \Phi_G(\mathbf{b}; \hat{c}, \hat{\pi}, \hat{\Lambda}_0)$, where A is a $(K+2p+q) \times (K+2p+q)$ positive-definite matrix. The asymptotic properties of the estimator $\hat{\beta}_G(A)$ are given in Theorem 2.

Theorem 2—Under the regularity conditions (a1)–(a7) in Web Appendix A, $\hat{\beta}_G(A)$ exists and converges almost surely to the limit β^* . Furthermore, $\sqrt{n}(\hat{\beta}_G(A) - \beta^*)$ is asymptotically normally distributed with mean zero and covariance matrix $V(A) = \{ D'(\beta^*) A D(\beta^*) \}^{-1} D(\beta^*)' A \Omega_G(\beta^*) A' D(\beta^*) \{ D'(\beta^*) A' D(\beta^*) \}^{-1}$, where $D(\mathbf{b}) \equiv \mathcal{E} \{ -\Phi_G(\mathbf{b}; c, \pi, \Lambda_0) / \mathbf{b}' \}$, $\Omega_G(\mathbf{b}) \equiv \mathcal{E} \{ \psi(\mathbf{b}; c, \pi, \Lambda_0) \psi'(\mathbf{b}; c, \pi, \Lambda_0) \}$ and ψ is defined in Web Appendix A.

As shown in Theorem 2, the asymptotic variance of $\sqrt{n}(\hat{\beta}_G(A) - \beta^*)$ depends on the weighting matrix A . The most efficient estimator is achieved at $A^{opt} = \Omega_G^{-1}$ with the minimum variance $V(A^{opt}) = \{ D'(\beta^*) \Omega_G^{-1}(\beta^*) D(\beta^*) \}^{-1}$ and A^{opt} can be estimated based on the preliminary parameter estimate $\hat{\Omega}_G^{-1}(\hat{\beta}_S)$. Finally, $\hat{\beta}_G^{opt} = \hat{\beta}_G(\hat{A}^{opt})$ is referred to as the GMM nonparametric (GNC) estimator and the covariance matrix of $\sqrt{n}(\hat{\beta}_G^{opt} - \beta^*)$ can be estimated by $\hat{V}(\hat{A}^{opt}) = \{ \hat{D}'(\hat{\beta}_G^{opt}) \hat{\Omega}_G^{-1}(\hat{\beta}_G^{opt}) \hat{D}(\hat{\beta}_G^{opt}) \}^{-1}$, where $\hat{D}(\mathbf{b}) = -n^{-1} \sum_{i=1}^n \psi(\mathbf{b}; \hat{c}, \hat{\pi}, \hat{\Lambda}_0) / \mathbf{b}'$ and $\hat{\Omega}_G(\mathbf{b}) = n^{-1} \sum_{i=1}^n \{ \psi_i(\mathbf{b}; \hat{c}, \hat{\pi}, \hat{\Lambda}_0) \psi_i'(\mathbf{b}; \hat{c}, \hat{\pi}, \hat{\Lambda}_0) \}$.

Notice that if there exist type-specific effects $(\beta_{X_k}, \beta_{Z_k})$ for (X_k, Z_k) , $k = 1, \dots, K$, the correction terms for different types of recurrent events should be estimated separately. For example, the correction term of the k th type of recurrent events can be estimated by

$$\begin{aligned} \tilde{c}_k = & - \sum_{i=1}^n \frac{\Delta_i}{\pi_i} \mathbf{W}_i \left\{ m_{ik} \hat{\Lambda}_{0k}^{-1}(Y_i) - \exp(\hat{\beta}_{S,0k} + \hat{\beta}_{S,Xk}' \mathbf{W}_i + \hat{\beta}_{S,Zk}' \mathbf{Z}_i) \right\} / \sum_{i=1}^n \frac{\Delta_i}{\pi_i} \exp(\hat{\beta}_{S,0k}), \quad k = 1, \\ & + \hat{\beta}_{S,Xk}' \mathbf{W}_i + \hat{\beta}_{S,Zk}' \mathbf{Z}_i \\ & \dots, K, \text{ where } (\hat{\beta}_{S,0k}, \hat{\beta}_{S,Xk}, \hat{\beta}_{S,Zk}) \text{ is solved from the SNC method. Similarly, the GNC} \end{aligned}$$

estimator can be obtained by minimizing $\Phi'_G(\mathbf{b}; \tilde{c}, \hat{\pi}, \hat{\Lambda}_0) \tilde{A}^{opt} \Phi_G(\mathbf{b}; \tilde{c}, \hat{\pi}, \hat{\Lambda}_0)$, where $\tilde{c} = (\tilde{c}_1, \dots, \tilde{c}_K)$ and $\tilde{A}^{opt}$ is the estimate of the corresponding optimal matrix.

4. Simulation study

We conduct a simulation study to compare the performance of the following four methods: naive, ideal, SNC and GNC, where the ideal method pretends that the true X can be obtained. A bivariate recurrent event processes ($K = 2$) is considered under these four simulation scenarios. In each scenario, the frailty variables (ν_1, ν_2) are generated from a bivariate gamma distribution with mean $(10, 10\sqrt{3})$, variance $(20, 20)$ and correlation coefficient 0.5. The error-prone covariate X follows $N(0, 0.5)$ and is contaminated by the error term U , and the error-free covariate Z follows $Bin(0.5)$. The informative censoring time Y^* has a hazard function $h_{Y^*}(y | \nu_1, \nu_2, X, Z) = (\nu_1 + \nu_2) \exp(X + 0.5Z)y/400$ and the IV V is generated from model (E2) with $g(X, Z, e) = -0.5X^2Z + X + e$, where $e \sim N(0, 0.2)$. The indicator is generated from the logistic model in Section 3.1 to satisfy the proportion of calibration sample as 0.7, 0.4, and 0.2. The recurrent events are generated from model (M2) with the baseline functions $\lambda_{01}(t) = (1 + t)^{-1}/\log(11)$ and $\lambda_{02}(t) = 1/10$, for $t \in [0, \tau = 10]$. Specifically,

- i. given ν_k , $N_k(t)$ follows a Poisson process with the intensity function $\nu_k \lambda_{0k}(t) \exp(\beta_{Xk}X + \beta_{Zk}Z)$ and $U \sim N(0, \sigma_U^2)$;
- ii. given ν_k , $N_k(t)$ follows a Poisson process with the intensity function $\nu_k \lambda_{0k}(t) \exp(\beta_{Xk}X + \beta_{Zk}Z)$ and $U \sim Unif(-\sqrt{3\sigma_U^2}, \sqrt{3\sigma_U^2})$;
- iii. given (ρ, ν_k) , $N_k(t)$ follows a Poisson process with the intensity function $\rho \nu_k \lambda_{0k}(t) \exp(\beta_{Xk}X + \beta_{Zk}Z)$ and $U \sim Unif(-\sqrt{3\sigma_U^2}, \sqrt{3\sigma_U^2})$, where $\rho \sim \Gamma(1, 1)$;
- iv. given ν_k , $N_k(t)$ follows a Poisson process with the intensity function $\nu_k \lambda_{0k}(t) \exp(\beta_{Xk}X + \beta_{Zk}Z)$ and $U \sim SN(0, \sigma_U^2, \gamma = 2)$, where SN denotes a skew normal distribution and γ denotes a skewness parameter.

In all scenarios, we let $\beta_{X1} = \beta_{Z1} = 0.5$ and $\beta_{X2} = \beta_{Z2} = 0.8$. In Scenario (iii), the assumption of a Poisson process imposed in Wang et al. (2001) is relaxed since given ν_k the events within a subject are still correlated due to the additional frailty variable ρ . Such a simulation setting is also used in Kalbeisch et al. (2013). The results of Scenarios (i) to (iv) are given in Tables 1 through 4, respectively. In the tables, ESD denotes the empirical sample standard deviation, ASE denotes the average standard error and CP denotes the coverage probability of the 95% confidence interval based on 300 replicates with $n = 600$.

From Tables 1–4, it can be seen that the naive estimator for the error-prone covariate X is biased and the performance gets worse with increasing variance of measurement error (σ_U^2).

The naive estimator of the error-free covariate Z is not affected by the measurement errors because in our estimation Z is generated to be independent of X . On the other hand, the SNC and GNC estimators for X and Z have negligible biases and the Wald-type 95% confidence intervals based on the estimated asymptotic standard errors have proper coverage

probabilities under different error distributions. In addition, the variances of the proposed estimators decrease when the variance of measurement errors decreases or proportion of calibration sample increases. It is worth noting that the proposed estimators in Scenario (iii) still perform well in terms of bias and coverage probability (see Table 3) even if the assumption of a Poisson process is not satisfied. Comparing these two methods, the GNC estimator has a smaller variance than the SNC estimator, especially when the proportion of calibration sample decreases. In general, the averages of the estimated standard errors for the proposed estimators are close to their empirical standard deviations. More simulation results (sample size $n = 1000$) with similar conclusions are also presented in Web Appendix B.

5. Data application

We apply the proposed methods to the bivariate recurrent event data collected from the NPC study. The purpose of the investigation is to access the effects of Se supplement on the recurrence rates of SCC and BCC adjusted for the baseline Se level, age, body mass index (BMI), gender, and smoking status. This randomized, double-blinded clinical trial lasted up to twelve years and recruited a total of 1312 patients with histories of skin cancers. All patients in the study had potential risks of developing SCC and BCC. For each patient, the Se level was measured one time at the baseline. As mentioned earlier, the baseline Se measurement was contaminated with errors caused by the measuring instrument or temporary biological fluctuation and is treated as a surrogate of the true Se level. A fraction of patients in the trial had received the periodical exams and had the Se measurements after randomization. Because the subsequent Se measurement depends on both the baseline measurement and treatment assignment, it is reasonable to be used as an IV.

We consider six covariates in the models. Let X denote the logarithm of the underlying unobserved baseline Se value (abbreviated as $\log(\text{Se})$) and $\mathbf{Z}$ denote the other five covariates including the treatment assignment, age, body mass index (BMI), gender, and smoking status, where smoking status was classified as never smoker, former smoker and current smoker with never smoker as the reference group. In addition, let W denote the logarithm of the observed Se measurement at the baseline and V denote the logarithm of the second observed Se measurement measured within 3.5 months since randomization. About 50% of the patients are included in the calibration sample. Based on the calibration sample, the relationship between the IV (V) and the observed covariate (W) is displayed in Web Figure 1 of Web Appendix C, where the left and right panels of Figure 1 are the scatter plots of V versus W for placebo and treatment groups, respectively. Basically, it can be seen that there exists a positive association between V and W for both groups although the magnitude of association seems to be different. This result indicates that the IV (V) is associated with $(X, \mathbf{Z})$. Moreover, the proposed approach is still valid even when the relationship between V and $(X, \mathbf{Z})$ is nonlinear.

We define that the censoring time Y as the last examination time since randomization and $\tau = 149.5$ (*months*) as the maximum time among the Y 's. The informative censoring is examined by calculating the weighted averages of the SCC and BCC recurrences versus time for the subjects in the four selected risk sets ($t_1 = 54.9$, $t_2 = 86.3$, $t_3 = 115.5$, $t_4 = 135.2$). More specifically, let $N_{i,S}(t)$ and $N_{i,B}(t)$ denote the numbers of SCC and BCC recurrences

before time t for the i th subject. The weighted averages of SCC and BCC recurrences before time t are calculated as $\sum_{i=1}^n N_{i,S}(t \wedge Y_i)I(Y_i > t_r) / \sum_{i=1}^n I(Y_i > t_r)$ and $\sum_{i=1}^n N_{i,B}(t \wedge Y_i)I(Y_i > t_r) / \sum_{i=1}^n I(Y_i > t_r)$ for the subjects in the r th risk set, $r = 1, \dots, 4$. The results can be found in Web Figures 2 and 3 of Web Appendix C. For example, in Figure 2, it suggests that the assumption of independent censoring is not satisfied since the patients who stayed in the trial longer (censoring time after 135.2 months) tended to have fewer SCC recurrences in the early time period.

A total of 1271 patients were included in the data set to fit the regression models, where 640 and 631 patients were in the treatment and placebo group, respectively. Among these patients, 473 and 516 patients developed at least one SCC and BCC event, and 231 patients experienced both of SCC and BCC occurrences. The results obtained from the naive, SNC and GNC approach are presented in Table 5. The estimates of the log(se) effect obtained from all approaches are statistically nonsignificant, and the estimated standard errors using GNC approach are smaller than SNC approach, revealing the efficiency of GNC approach. It should be noted that the naive estimates may not be attenuated since the magnitude of bias is also related to the conditional distribution of X given (W, Z) (Carroll et al., 2006, Chapter 4). For error-free covariates, gender is the only covariate which is significantly associated with the recurrences of SCC and BCC (almost for all approaches), while age is another significant covariate but is only associated with the recurrence of SCC. The result with three covariates is presented in Web Appendix C.

6. Discussion

In this article, we propose non-parametric correction approaches for correcting measurement errors in the regression analysis of multivariate recurrent event data subject to informative censoring. We consider a shared frailty model to accommodate the dependences between the censoring time and recurrent event processes, whereas the frailty distribution is allowed to be unspecified. The shared frailty model also accounts for the association among different types of recurrent events. In the absence of measurement errors, this model can be considered as an extension of the univariate recurrent event model suggested by Wang et al. (2001) and Xu et al. (2017). To deal with measurement errors, we assume that the surrogate W follows a classical measurement error model without specifying the distributions of X and U . In addition to a surrogate variable W , an IV V is needed for bias-correction. Our methods do not require replicates of W (Huang and Wang, 2000; Hu and Lin, 2004). In fact, in the NPC data, the second Se measurement depends on both baseline Se level and treatment assignment (Clark et al., 1996) and the dependence is unknown. Hence, it should be treated as an IV rather than a replicated measurement. In addition, the proposed methods allow the second Se measurement to be partially observed because only a subset of subjects had the subsequent Se measurements after the baseline.

One restriction of the proposed approaches is that our methods can not be applied to time-dependent covariates or time-by-covariate interactions even in the absence of measurement errors. In particular, the truncation product-limit estimator of $\Lambda_{0k}(t)$ is not available anymore since the time-dependent covariates or time-by-covariate interactions can not be eliminated

in the estimation of $\Lambda_{0k}(t)$. As a result, the proposed estimating procedures in Section 3.1 and 3.2 are no longer valid. Hence future research is warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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 (Wang), GM100573 (Wang), HL121347 (Yu and Wang), HL130483 (Wang), MH105857 (Wang), and a travel award from the Mathematics Research Promotion Center of National Science Council of Taiwan (Wang).

References

- Abu-Libdeh H, Turnbull BW, Clark LC. Analysis of multi-type recurrent events in longitudinal studies: application to a skin cancer prevention trial. *Biometrics*. 1990; 46:1017–1034. [PubMed: 2085623]
- Bedair K, Hong Y, Li J, Al-Khalidi HR. Multivariate frailty models for multi-type recurrent event data and its application to cancer prevention trial. *Computational Statistics and Data Analysis*. 2016; 101:161–173.
- Cai J, Schaubel DE. Marginal means/rates models for multiple type recurrent event data. *Lifetime data analysis*. 2004; 10:121–138. [PubMed: 15293628]
- Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement error in nonlinear models: a modern perspective*. Chapman & Hall; London: 2006.
- Clark LC, Combs GF, Turnbull BW, Slate EH, Chalker DK, Chow J, Davis LS, Glover RA, Graham GF, Gross EG, et al. Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin: a randomized controlled trial. *Journal of the American Medical Association*. 1996; 276:1957–1963. [PubMed: 8971064]
- Hu C, Lin DY. Semiparametric failure time regression with replicates of mismeasured covariates. *Journal of the American Statistical Association*. 2004; 99:105–118.
- Huang Y, Wang CY. Cox regression with accurate covariates unascertainable: a nonparametric-correction approach. *Journal of the American Statistical Association*. 2000; 95:1209–1219.
- Jiang W, Turnbull BW, Clark LC. Semiparametric regression models for repeated events with random effects and measurement error. *Journal of the American Statistical Association*. 1999; 94:111–124.
- Kalbeisch JD, Schaubel DE, Ye Y, Gong Q. An estimating function approach to the analysis of recurrent and terminal events. *Biometrics*. 2013; 69:366–374. [PubMed: 23651362]
- Nakamura T. Proportional hazards model with covariates subject to measurement error. *Biometrics*. 1992; 48:829–838. [PubMed: 1420844]
- Ning J, Rahbar MH, Choi S, Piao J, Hong C, del Junco DJ, Rahbar E, Fox EE, Holcomb JB, Wang MC. Estimating the ratio of multivariate recurrent event rates with application to a blood transfusion study. *Statistical Methods in Medical Research*. 2017; 26:1969–1981. [PubMed: 26160825]
- Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*. 1982; 69:331–342.
- Schaubel D, Johansen H, Dutta M, Desmeules M, Becker A, Mao Y. Neonatal characteristics as risk factors for preschool asthma. *Journal of Asthma*. 1996; 33:255–264. [PubMed: 8707780]
- Song X, Wang CY. Proportional hazards model with covariate measurement error and instrumental variables. *Journal of the American Statistical Association*. 2014; 109:1636–1646. [PubMed: 25663724]
- Tsiatis AA, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*. 2001; 88:447–458.

- Wang MC, Qin J, Chiang CT. Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association*. 2001; 96:1057–1065.
- Xu G, Chiou SH, Huang CY, Wang MC, Yan J. Joint scale-change models for recurrent events and failure time. *Journal of the American Statistical Association*. 2017; 112:794–805. [PubMed: 28943684]
- Yi GY, Lawless JF. Likelihood-based and marginal inference methods for recurrent event data with covariate measurement error. *Canadian Journal of Statistics*. 2012; 10:530–549.
- Yu H, Cheng YJ, Wang CY. Semiparametric regression estimation for recurrent event data with errors in covariates under informative censoring. *The International Journal of Biostatistics*. 2016; 12(2)
- Zhao X, Liu L, Liu Y, Xu W. Analysis of multivariate recurrent event data with time-dependent covariates and informative censoring. *Biometrical Journal*. 2012; 54:585–599. [PubMed: 22886587]
- Zhu L, Sun J, Tong X, Srivastava DK. Regression analysis of multivariate recurrent event data with a dependent terminal event. *Lifetime Data Analysis*. 2010; 16:478–490. [PubMed: 20217474]

Table 1

Estimation for the effects of (X, Z) based on 300 replicates with $n = 600$, where X is contaminated by a normal error U and the recurrent event processes satisfy the assumption of a Poisson process. (β_{X1}, β_{Z1}) and (β_{X2}, β_{Z2}) denote the effects of (X, Z) on type 1 and 2 recurrent event processes, respectively, Cal denotes proportion of calibration sample, BIAS denotes the average of $\hat{\beta} - \beta$ from 300 samplings, ASE denotes the average standard error from 300 samplings, ESD denotes the empirical standard deviation from 300 samplings, CP denotes the coverage probability of Wald 95% confidence interval from 300 samplings.

σ_u^2	Cal	β_{X1}					β_{Z1}				
		BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP	BIAS	ESD
0.1	0.7	Ideal	-0.003	0.042	0.038	0.92	0.000	0.052	0.052	0.95	
		naive	-0.086	0.038	0.036	0.32	0.001	0.053	0.053	0.96	
		SNC	0.001	0.060	0.058	0.94	-0.005	0.072	0.070	0.94	
	0.4	GNC	0.000	0.053	0.053	0.94	0.000	0.053	0.054	0.96	
		SNC	0.008	0.078	0.075	0.93	-0.007	0.099	0.093	0.93	
		GNC	0.002	0.063	0.063	0.93	0.001	0.053	0.054	0.96	
0.2	0.7	Ideal	0.003	0.133	0.111	0.91	-0.006	0.156	0.142	0.92	
		naive	-0.145	0.036	0.034	0.04	0.001	0.054	0.055	0.96	
		SNC	0.002	0.069	0.066	0.94	-0.004	0.072	0.073	0.95	
	0.4	GNC	0.001	0.063	0.062	0.94	0.000	0.054	0.056	0.96	
		SNC	0.008	0.088	0.086	0.94	-0.005	0.103	0.096	0.93	
		GNC	0.004	0.079	0.077	0.93	0.001	0.055	0.056	0.97	
0.1	0.7	Ideal	0.007	0.152	0.128	0.91	-0.004	0.163	0.148	0.92	
		naive	-0.002	0.122	0.106	0.92	-0.000	0.055	0.057	0.96	
		SNC	0.007	0.152	0.128	0.91	-0.004	0.163	0.148	0.92	
	0.4	GNC	0.003	0.070	0.064	0.95	0.005	0.056	0.058	0.95	
		SNC	0.010	0.096	0.086	0.93	0.010	0.110	0.106	0.94	
		GNC	0.003	0.070	0.064	0.95	0.005	0.056	0.058	0.95	

σ_u^2	Cal	BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP
0.2	GNC	0.010	0.080	0.078	0.93	0.006	0.056	0.059	0.97
	SNC	0.002	0.158	0.129	0.91	-0.007	0.187	0.161	0.89
	GNC	-0.006	0.113	0.101	0.94	0.005	0.058	0.060	0.95
	naive	-0.229	0.036	0.038	0.04	0.006	0.056	0.060	0.95
0.7	SNC	0.005	0.094	0.085	0.96	0.006	0.082	0.087	0.97
	GNC	0.007	0.092	0.080	0.95	0.004	0.060	0.064	0.97
0.4	SNC	0.016	0.120	0.110	0.93	0.009	0.117	0.116	0.95
	GNC	0.017	0.112	0.103	0.94	0.006	0.061	0.065	0.96
0.2	SNC	0.010	0.200	0.166	0.92	-0.008	0.202	0.177	0.93
	GNC	-0.008	0.155	0.133	0.91	0.004	0.063	0.066	0.95

Table 2

Estimation for the effects of (X, Z) based on 300 replicates with $n = 600$, where X is contaminated by a uniform error U and the recurrent event processes satisfy the assumption of a Poisson process. (β_{X1}, β_{Z1}) and (β_{X2}, β_{Z2}) denote the effects of (X, Z) on type 1 and 2 recurrent event processes, respectively, Cal denotes proportion of calibration sample, BIAS denotes the average of $\hat{\beta} - \beta$ from 300 samplings, ASE denotes the average standard error from 300 samplings, ESD denotes the empirical standard deviation from 300 samplings, CP denotes the coverage probability of Wald 95% confidence interval from 300 samplings.

σ_u^2	Cal	β_{X1}					β_{Z1}				
		BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP	BIAS	ESD
0.1	Ideal	0.003	0.037	0.039	0.95	0.000	0.055	0.052	0.95		
	naive	-0.080	0.037	0.037	0.42	0.001	0.056	0.053	0.94		
	0.7 SNC	0.007	0.056	0.057	0.95	0.003	0.078	0.071	0.94		
	GNC	0.006	0.050	0.053	0.96	0.001	0.056	0.054	0.94		
0.2	0.4 SNC	0.002	0.071	0.075	0.96	0.001	0.094	0.094	0.96		
	GNC	0.002	0.059	0.063	0.96	0.001	0.056	0.054	0.95		
	0.2 SNC	0.000	0.113	0.110	0.93	0.004	0.153	0.142	0.91		
	GNC	-0.005	0.081	0.084	0.96	0.000	0.056	0.054	0.95		
0.2	naive	-0.139	0.035	0.034	0.03	0.001	0.057	0.054	0.94		
	0.7 SNC	0.009	0.064	0.065	0.95	0.003	0.080	0.073	0.93		
	GNC	0.007	0.060	0.061	0.96	0.001	0.058	0.056	0.94		
	0.4 SNC	0.003	0.082	0.085	0.96	0.000	0.097	0.097	0.95		
0.2	GNC	0.003	0.074	0.075	0.95	0.001	0.057	0.056	0.95		
	0.2 SNC	0.006	0.135	0.128	0.94	0.002	0.160	0.149	0.91		
	GNC	-0.006	0.101	0.102	0.95	0.002	0.057	0.056	0.96		
0.1	Ideal	0.002	0.044	0.040	0.92	0.004	0.051	0.053	0.96		
	naive	-0.132	0.044	0.040	0.13	0.007	0.056	0.058	0.95		
	0.7 SNC	0.001	0.077	0.068	0.93	0.010	0.078	0.081	0.97		
	GNC	-0.001	0.067	0.065	0.95	0.006	0.056	0.059	0.96		
0.4	0.4 SNC	0.005	0.093	0.088	0.95	0.014	0.112	0.105	0.92		

σ_u^2	Cal		BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP
0.2		GNC	0.001	0.082	0.081	0.96	0.007	0.057	0.059	0.95
	0.2	SNC	0.005	0.153	0.127	0.89	0.006	0.179	0.159	0.91
		GNC	-0.004	0.120	0.107	0.95	0.005	0.058	0.060	0.97
	0.2	naive	-0.227	0.042	0.038	0.01	0.007	0.060	0.060	0.96
		SNC	0.002	0.094	0.084	0.94	0.010	0.083	0.087	0.97
0.4		GNC	-0.001	0.083	0.080	0.94	0.007	0.061	0.064	0.97
	0.4	SNC	0.008	0.116	0.110	0.95	0.012	0.119	0.114	0.92
		GNC	0.005	0.110	0.102	0.94	0.007	0.062	0.065	0.96
	0.2	SNC	0.017	0.195	0.167	0.91	0.001	0.195	0.178	0.92
		GNC	-0.008	0.150	0.134	0.93	0.005	0.062	0.066	0.96

Table 3

Estimation for the effects of (X, Z) based on 300 replicates with $n = 600$, where X is contaminated by a uniform error U and the recurrent event processes violate the assumption of a Poisson process. (β_{X1}, β_{Z1}) and (β_{X2}, β_{Z2}) denote the effects of (X, Z) on type 1 and 2 recurrent event processes, respectively, Cal denotes proportion of calibration sample, BIAS denotes the average of $\hat{\beta} - \beta$ from 300 samplings, ASE denotes the average standard error from 300 samplings, ESD denotes the empirical standard deviation from 300 samplings, CP denotes the coverage probability of Wald 95% confidence interval from 300 samplings.

σ_u^2	Cal	β_{X1}					β_{Z1}				
		BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP		
0.1	Ideal	-0.003	0.086	0.081	0.93	-0.005	0.107	0.108	0.96		
		-0.083	0.078	0.074	0.74	-0.002	0.108	0.109	0.96		
	naive	-0.003	0.105	0.101	0.95	0.003	0.123	0.122	0.95		
		GNC	-0.002	0.101	0.097	0.94	-0.002	0.107	0.109	0.96	
	0.7	SNC	0.004	0.121	0.123	0.95	-0.002	0.146	0.149	0.95	
		GNC	-0.001	0.110	0.111	0.94	-0.003	0.108	0.109	0.96	
0.2	Ideal	0.011	0.216	0.193	0.93	-0.013	0.251	0.242	0.94		
		GNC	-0.007	0.154	0.150	0.94	-0.004	0.110	0.110	0.96	
	naive	-0.143	0.071	0.069	0.43	-0.002	0.109	0.109	0.96		
		SNC	-0.001	0.120	0.114	0.95	0.004	0.124	0.124	0.96	
	0.7	GNC	0.000	0.117	0.110	0.94	-0.002	0.108	0.110	0.96	
		SNC	0.009	0.138	0.139	0.95	-0.004	0.149	0.152	0.95	
0.4	Ideal	0.001	0.128	0.128	0.95	-0.003	0.108	0.111	0.96		
		SNC	0.015	0.236	0.220	0.93	-0.013	0.260	0.248	0.94	
	naive	-0.008	0.183	0.178	0.93	-0.001	0.111	0.112	0.96		
		GNC	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94	
	0.7	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
0.7	Ideal	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95		
		GNC	0.003	0.149	0.144	0.93	0.009	0.172	0.167	0.94	
	naive	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	
0.4	Ideal	0.000	0.093	0.088	0.94	0.006	0.116	0.112	0.94		
		naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94	
	naive	-0.134	0.088	0.081	0.57	0.007	0.119	0.114	0.94		
		SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
	0.7	SNC	0.000	0.120	0.115	0.94	0.008	0.145	0.134	0.94	
		GNC	-0.002	0.117	0.114	0.94	0.007	0.119	0.115	0.95	

σ_u^2	Cal	BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP
0.2	GNC	0.002	0.139	0.134	0.95	0.010	0.119	0.115	0.96
	SNC	0.034	0.337	0.264	0.91	-0.018	0.362	0.305	0.90
	GNC	-0.015	0.234	0.183	0.91	0.005	0.122	0.118	0.94
	naive	-0.227	0.082	0.076	0.16	0.007	0.121	0.115	0.95
0.7	SNC	0.003	0.140	0.138	0.95	0.009	0.144	0.138	0.95
	GNC	-0.001	0.139	0.134	0.95	0.008	0.122	0.118	0.94
0.4	SNC	0.013	0.176	0.173	0.95	0.008	0.178	0.172	0.93
	GNC	0.003	0.157	0.156	0.96	0.008	0.123	0.118	0.95
0.2	SNC	0.029	0.315	0.287	0.91	-0.008	0.348	0.289	0.92
	GNC	-0.022	0.262	0.226	0.89	0.003	0.126	0.124	0.94

Table 4

Estimation for the effects of (X, Z) based on 300 replicates with $n = 600$, where X is contaminated by a skew normal error U and the recurrent event processes satisfy the assumption of a Poisson process. (β_{X1}, β_{Z1}) and (β_{X2}, β_{Z2}) denote the effects of (X, Z) on type 1 and 2 recurrent event processes, respectively, Cal denotes proportion of calibration sample, BIAS denotes the average of $\hat{\beta} - \beta$ from 300 samplings, ASE denotes the average standard error from 300 samplings, ESD denotes the empirical standard deviation from 300 samplings, CP denotes the coverage probability of Wald 95% confidence interval from 300 samplings.

σ_u^2	Cal	β_{X1}					β_{Z1}				
		BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP		
0.1	Ideal	0.003	0.038	0.038	0.95	-0.003	0.052	0.052	0.95		
		naive	-0.084	0.037	0.036	0.33	-0.003	0.056	0.054	0.95	
	0.7	SNC	0.004	0.062	0.058	0.94	0.001	0.069	0.071	0.94	
		GNC	0.005	0.056	0.053	0.95	-0.003	0.056	0.054	0.95	
	0.4	SNC	0.008	0.085	0.075	0.91	0.001	0.096	0.095	0.93	
		GNC	0.007	0.067	0.063	0.95	-0.004	0.056	0.054	0.96	
0.2	0.2	SNC	0.011	0.127	0.113	0.92	0.003	0.160	0.146	0.91	
		GNC	0.002	0.088	0.087	0.93	-0.003	0.056	0.055	0.95	
	naive	-0.147	0.033	0.033	0.02	-0.004	0.056	0.055	0.95		
		0.7	SNC	0.007	0.069	0.067	0.96	0.001	0.073	0.074	0.95
	GNC	0.008	0.065	0.063	0.96	-0.004	0.058	0.057	0.96		
		0.4	SNC	0.003	0.091	0.086	0.95	0.003	0.100	0.099	0.93
0.4	GNC	0.005	0.082	0.076	0.93	-0.003	0.059	0.057	0.94		
		0.2	SNC	0.019	0.153	0.133	0.92	0.003	0.170	0.153	0.92
	GNC	0.001	0.114	0.108	0.91	-0.002	0.058	0.058	0.95		
		β_{X2}					β_{Z2}				
	0.1	Ideal	0.005	0.049	0.042	0.94	0.005	0.050	0.054	0.96	
			naive	-0.137	0.045	0.041	0.13	0.004	0.057	0.058	0.94
0.7		SNC	0.003	0.071	0.068	0.94	0.011	0.082	0.081	0.94	
		GNC	0.004	0.068	0.065	0.95	0.005	0.058	0.060	0.94	
0.4		SNC	0.005	0.090	0.087	0.94	0.015	0.117	0.108	0.92	
		GNC	0.005	0.049	0.042	0.94	0.005	0.050	0.054	0.96	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

σ_u^2	Cal	BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP
0.2	GNC	0.003	0.080	0.079	0.95	0.006	0.060	0.060	0.95
	SNC	0.019	0.154	0.134	0.92	0.003	0.190	0.169	0.93
	GNC	0.003	0.114	0.106	0.94	0.004	0.060	0.061	0.95
	naive	-0.240	0.040	0.039	0.00	0.004	0.060	0.061	0.94
0.7	SNC	0.005	0.089	0.086	0.95	0.014	0.092	0.091	0.95
	GNC	0.007	0.085	0.083	0.95	0.007	0.066	0.068	0.95
	SNC	0.006	0.112	0.110	0.93	0.017	0.128	0.119	0.92
	GNC	0.001	0.102	0.100	0.95	0.007	0.067	0.068	0.94
0.2	SNC	0.025	0.191	0.167	0.94	0.003	0.206	0.184	0.94
	GNC	-0.002	0.149	0.137	0.92	0.006	0.068	0.070	0.95

Table 5

Regression analysis of the SCC and BCC recurrences in the NPC trial.

		SCC				BCC			
		Est	SE	Z-value	P-value	Est	SE	Z-value	P-value
log(se)	naive	-0.442	0.275	-1.609	0.108	0.496	0.289	1.714	0.087
	SNC	-0.791	1.625	-0.487	0.627	0.923	0.756	1.221	0.222
	GNC	-0.272	1.463	-0.186	0.853	0.826	0.737	1.121	0.262
treatment	naive	0.157	0.157	0.999	0.318	0.023	0.087	0.262	0.793
	SNC	0.189	0.190	0.996	0.319	-0.056	0.094	-0.596	0.551
	GNC	0.132	0.149	0.888	0.374	0.006	0.085	0.065	0.948
age	naive	0.380	0.077	4.899	0.000 *	-0.004	0.047	-0.091	0.928
	SNC	0.356	0.091	3.904	0.000 *	0.021	0.051	0.406	0.685
	GNC	0.379	0.073	5.186	0.000 *	0.001	0.047	0.013	0.990
BMI	naive	0.053	0.072	0.735	0.462	-0.035	0.043	-0.806	0.420
	SNC	-0.051	0.111	-0.462	0.644	-0.042	0.049	-0.868	0.386
	GNC	0.058	0.069	0.840	0.401	-0.033	0.043	-0.781	0.435
gender	naive	0.506	0.248	2.037	0.042 *	0.448	0.123	3.643	0.000 *
	SNC	0.878	0.297	2.957	0.003 *	0.439	0.148	2.974	0.003 *
	GNC	0.468	0.249	1.875	0.061	0.433	0.126	3.432	0.001 *
former smoker	naive	0.178	0.236	0.753	0.451	0.016	0.099	0.165	0.869
	SNC	0.335	0.235	1.423	0.155	0.091	0.120	0.759	0.448
	GNC	0.172	0.234	0.735	0.462	-0.002	0.095	-0.025	0.980
current smoker	naive	0.293	0.230	1.273	0.203	-0.006	0.127	-0.050	0.960
	SNC	0.240	0.236	1.013	0.311	-0.096	0.138	-0.698	0.485
	GNC	0.307	0.226	1.362	0.173	-0.004	0.126	-0.029	0.977

EST denotes the estimate, SE denotes the standard error which is calculated by the square root of the asymptotic variance estimator, and * denotes the estimate reaches the significant level.