Contemporary Clinical Trials Communications 1 (2015) 5–16

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Trend-constrained corrected score for proportional hazards model with covariate measurement error

Ming Zhu [a, *], Yijian Huang [b]

[a] Abbvie Inc, 1 North Waukegan Rd, North Chicago, IL 60064, USA
[b] Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA

ARTICLE INFO

Article history:
Received 5 June 2015
Received in revised form
7 August 2015
Accepted 24 August 2015
Available online 23 October 2015

Keywords:
Cox proportional hazards model
Estimating equation
Functional modeling
Measurement error
Survival analysis

ABSTRACT

In many medical research studies, survival time is typically the primary outcome of interest. The Cox proportional hazards model is the most popular method to investigate the relationship between covariates and possibly right-censored survival time. However, in many clinical trials, the true covariates may not always be accurately measured due to natural biological fluctuation or instrument error. It is well know that for regression analysis in general, naively using mismeasured covariates in conventional inference procedures may incur substantial estimation bias. In the presence of covariate measurement error, several functional modeling methods have been proposed under the situation where the distribution of the measurement error is known. Among them are parametric corrected score and conditional score. Although both methods are consistent, each suffers from severe problem of multiple roots or absence of appropriate root when the measurement error is substantial. The problem persists even when the sample size is practically large. We conduct a detailed investigation on the pathological behaviors of parametric corrected score and propose an approach of incorporating additional estimating functions to remedy these pathological behaviors. The estimation and inference are then accomplished by means of quadratic inference function. Extensive simulation studies are conducted to evaluate the performance of proposed method.

1. Introduction

The proportional hazards model is one of the most popular models to investigate the relationship between time to failure and covariates. However, in many clinical trials, the true covariates may not always be accurately measured due to natural biological fluctuation or instrument error. In some studies, the magnitude of measurement error could be substantial to the extent that it is comparable to or even larger than that of the true underlying covariate. A typical example is the HIV viral load in HIV/AIDS studies.

For regression analysis in general, naively using mismeasured covariates in conventional inference procedures may incur substantial estimation bias and several statistical methods have been suggested to address covariate measurement error; see the monograph of Carroll et al. for a good summary [1]. Regression calibration is used frequently to yield approximate estimation. However, it is well known that the regression calibration estimator is inconsistent in general. For consistent estimation, methods have been developed under either structural or functional modeling, i.e., with or without parametric distributional assumptions imposed on the true covariates. By definition, functional modeling approach might be more appealing, particularly for its robustness. Available functional modeling methods for Cox proportional hazards model include the conditional score [2], the parametric corrected score [3–5], and the nonparametric corrected score [5–8]. The idea of the conditional score is to condition away the nuisance parameters based on certain sufficient statistics whereas the last two classes adopt a correction strategy by constructing a corrected estimating function with error-contaminated covariates that shares the same limit as a reference estimating function with true underlying covariates. If the reference estimating function admits consistent estimates only, the corrected estimating function shall inherit this property in a compact parameter space containing the true value. Conditional score and parametric corrected score are generally different. But in the case of the Cox proportional hazards model and normal measurement error, the conditional score and parametric

* Corresponding author.
E-mail addresses: ming.zhu@abbvie.com (M. Zhu), yhuang5@emory.edu (Y. Huang).

http://dx.doi.org/10.1016/j.conctc.2015.08.001

corrected score estimators are asymptotically equivalent.

Although all three aforementioned methods produce consistent estimators, they all suffer from finite-sample pathological behaviors especially when the measurement error is substantial, which limit the applicability of these methods in practice. Recently, Huang [9] proposed an approach to incorporate additional estimating functions which constrain the derivatives of the parametric corrected score for loglinear model. This approach effectively remedies those pathological behaviors and also considerably improves the estimation efficiency. Huang's approach provides a promising general strategy to handle similarly ill-behaved estimating functions.

In this paper, we first conduct a detailed investigation on pathological behaviors of parametric corrected score and conditional score. After that, we propose an augmented estimation procedure in which additional estimating functions are added to the parametric corrected score for the proportional hazards model. In Section 2, we briefly describe the parametric corrected score and conditional score for the proportional hazards model and present the investigation results on the pathological behaviors when covariate measurement error is substantial. The proposed approach of incorporating additional estimating functions for the parametric corrected score is presented in Section 3. Simulation studies with practical sample size are reported in Section 4 together with an application to the ACTG 175 clinical trial data. Further discussion is given in Section 5. Technical details is collected in the Appendix.

2. Parametric corrected score and conditional score for proportional hazards model and their pathological behaviors

The proportional hazards model postulates that the cumulative hazard function $\Lambda(\cdot)$ of survival time T of an individual with a p-vector of covariate $\mathbf{Z}$ has the form

$$\Lambda(dt|\mathbf{Z}) = \exp(\beta'\mathbf{Z})\Lambda_0(dt)$$

where β is a p-vector parameter of interest and $\Lambda_0(\cdot)$ is an unspecified underlying cumulative hazard function. Let C denotes the censoring time and adopt the usual independent censoring mechanism: given $\mathbf{Z}$,C is independent of T.

The observed data, $(X_i,\Delta_i,\mathbf{Z}_i,i=1,\ldots,n)$, consist of n independent and identically distributed (i.i.d) replicates of $\{X\equiv T\wedge C,\Delta\equiv I(T\le C),\mathbf{Z}\}$. The standard inference procedure for Cox proportional hazards model is then to maximize the partial likelihood or, equivalently, to solve estimating function

$$\xi^*(\mathrm{b},\tilde{\Lambda}_0(\cdot)) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\begin{pmatrix}1\\ \mathbf{Z}_i\end{pmatrix}\{dN_i(t) - Y_i(t)\exp(\mathrm{b}'\mathbf{Z}_i)d\tilde{\Lambda}_0(t)\}, \tag{1}$$

where $N_i(t)=I(X_i\le t,\Delta_i=1)$ is the counting process, $Y_i(t)=I(X_i\ge t)$ is the at-risk process and τ is a positive constant such that $\Pr(T\ge\tau)>0$. Profiling out $\tilde{\Lambda}_0(\cdot)$, the estimating function for β alone is

$$\xi(\mathrm{b}) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{Z}_i - \frac{\sum_{j=1}^{n}Y_j(t)\mathbf{Z}_j\exp(\mathrm{b}'\mathbf{Z}_j)}{\sum_{j=1}^{n}Y_j(t)\exp(\mathrm{b}'\mathbf{Z}_j)}\right\}dN_i(t). \tag{2}$$

Estimating function (2) is actually the usual partial score function.

2.1. Parametric corrected score and conditional score

Split covariates $\mathbf{Z}=(\mathbf{Z}_a^T,\mathbf{Z}_e^T)^T$ where $\mathbf{Z}_a$ are those covariates that can be accurately measured and $\mathbf{Z}_e$ are covariates prone to measurement error and cannot be accurately measured. Though $\mathbf{Z}_e$ cannot be measured directly, we can observe them through their surrogates $\mathbf{W}_e$. Under the classical additive measurement error model, $\mathbf{W}_e=\mathbf{Z}_e+\boldsymbol{\varepsilon}_e$, where $\boldsymbol{\varepsilon}_e$ is the error vector and $\boldsymbol{\varepsilon}_e$ is assumed to be independent of $(T,C,\mathbf{Z})$. In this paper, we assume that the distribution of $\boldsymbol{\varepsilon}_e$ is known; The situation where distribution of $\boldsymbol{\varepsilon}_e$ is unknown will be discussed in Section 5.

Let $\mathbf{W}=(\mathbf{Z}_a,\mathbf{W}_e)$ and $\boldsymbol{\varepsilon}=(\mathbf{0},\boldsymbol{\varepsilon}_e)$. The observed data now consist of $(X_i,\Delta_i,\mathbf{W}_i,i=1,\ldots,n)$ in the presence of covariate measurement error. It is well known that naively replacing $\mathbf{Z}$ by $\mathbf{W}$ in estimating functions (1) or (2) could incur substantial estimation bias. Denote the cumulant-generating function of $\boldsymbol{\varepsilon}$ as $\Omega(\mathrm{b})\equiv\log\mathscr{E}\{\exp(\mathrm{b}'\boldsymbol{\varepsilon})\}$ and its derivative $\dot{\Omega}(\mathrm{b})\equiv\partial\Omega(\mathrm{b})/\partial\mathrm{b}$. The parametric corrected score estimating function is given by

$$\begin{aligned}\boldsymbol{\eta}^*(\mathrm{b},\tilde{\Lambda}_0(\cdot)) = n^{-1}\sum_{i=1}^{n}&\int_0^{\tau}\begin{pmatrix}1\\ \mathbf{W}_i - \dot{\Omega}(0)\end{pmatrix}dN_i(t)\\ &-\int_0^{\tau}Y_i(t)\exp\{\mathrm{b}'\mathbf{W}_i-\Omega(\mathrm{b})\}\\ &\times\begin{pmatrix}1\\ \mathbf{W}_i - \dot{\Omega}(\mathrm{b})\end{pmatrix}d\tilde{\Lambda}_0(t).\end{aligned} \tag{3}$$

Further profiling out $\tilde{\Lambda}_0(\cdot)$, we obtain the corrected score for β [10]:

$$\begin{aligned}\boldsymbol{\eta}(\mathrm{b}) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau}&\left\{\mathbf{W}_i + \dot{\Omega}(\mathrm{b}) - \dot{\Omega}(0)\right.\\ &\left.-\frac{\sum_{j=1}^{n}Y_j(t)\mathbf{W}_j\exp(\mathrm{b}'\mathbf{W}_j)}{\sum_{j=1}^{n}Y_j(t)\exp(\mathrm{b}'\mathbf{W}_j)}\right\}dN_i(t)\end{aligned} \tag{4}$$

which has the same expectation as reference (2) asymptotically for each and every finite $\mathbf{b}$. The estimation is then to find the zero crossing of the above estimating function. The consistency and asymptotic normality of corrected score estimator are later established by Kong and Gu [11].

Huang and Wang [6] defined a root-consistent estimating function as such that every zero-crossing is consistent and showed that a normalized estimating function is root-consistent if its limit has a unique root at the estimand. By definition, reference (2) is a root-consistent estimating function and the new estimating function (4) shall inherit the root-consistency from (2). The root-consistency of (4) assures that in a compact parameter space containing the true parameter β, the parametric corrected score will admit a unique root asymptotically and the root is consistent and asymptotically normal.

When the measurement error is normally distributed and assume the variance matrix is $\boldsymbol{\Sigma}$, the conditional score estimating function for the Cox proportional hazards model [12] can be written as

$$\begin{aligned}\boldsymbol{\eta}_{con}(\mathrm{b}) =& n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{W}_i + \boldsymbol{\Sigma}\mathrm{b}\right.\\ &\left.-\frac{\sum_{j=1}^{n}Y_j(t)[\mathbf{W}_j+\boldsymbol{\Sigma}\mathrm{b}dN_j(t)]\exp(\mathrm{b}'[\mathbf{W}_j+\boldsymbol{\Sigma}\mathrm{b}dN_j(t)])}{\sum_{j=1}^{n}Y_j(t)\exp(\mathrm{b}'[\mathbf{W}_j+\boldsymbol{\Sigma}\mathrm{b}dN_j(t)])}\right\}\\ &dN_i(t)\end{aligned} \tag{5}$$

In fact, it can be shown that estimators from parametric

corrected score and conditional score are asymptotically equivalent in the case of normal measurement error.

2.2. Pathological behaviors

When the magnitude of measurement error is small, the asymptotic results of parametric corrected score and conditional score provide a good approximation for practical purposes. However, when the measurement error increases, the pathological behaviors may start to arise [13]. These pathological behaviors include multiple zero-crossings or a single wrong zero-crossing that is inappropriate. These pathological behaviors may cause serious concerns when the measurement error is substantial and limit the applicability of parametric corrected score and conditional score in practice. In this section, we will conduct a detailed investigation of pathological behaviors for these two methods.

We first consider the parametric corrected score. Consider a single-covariate model with normal measurement error. In the absence of measurement error, the partial likelihood score function $\xi(b)$ is monotonically decreasing and has a unique root. The asymptotic result suggests that the parametric corrected score $\eta(b)$ should be monotonically decreasing as well in a compact parameter space containing the true parameter when the sample size is large. One may speculate the parametric corrected score to have an overall decreasing trend over the entire parameter space. But surprisingly, the overall trend of parametric corrected score in an unbounded parameter space is increasing. Function $\eta(b)$ takes a value of $-\infty$ when $b=-\infty$ and a value of $+\infty$ when $b=+\infty$. This observation suggests that the parametric corrected score $\eta(b)$ has an odd number of roots. In our numerical studies, only single- and triple-root patterns have been observed and two typical plots of parametric corrected score are illustrated in Fig. 1. The same root patterns were observed in Huang's investigation of loglinear model [9].

If we characterize a root by increase or decrease of $\eta(\cdot)$ around it, the increasing and decreasing roots correspond to local minimizers and maximizers of corresponding objective function, respectively. Therefore, an increasing root is considered as an inappropriate one. In the case of single-root pattern, the only root is increasing and thus inappropriate. For the triple-root pattern, an appropriate root exists since there is only one decreasing root. The single-root pattern is considered as root-finding failure.

We conduct a simulation study to exam the prevalence of single-root pattern. We consider a single covariate Z and generate it from various distributions: A) standard normal distribution, B) modified chi-square distribution, and C) uniform distribution with mean 0 and variance 1. To generate the modified chi-square distribution, the chi-square distribution with 1 degree of freedom was first truncated at 5 and then location-shifted to mean 0 and rescaled to variance 1. The measurement error follows standard normal distribution. The true coefficient was taken to be $\beta=-1$ and the baseline hazard is constant 1. Censoring was generated from a uniform distribution on $[0,\mu]$, where μ is chosen so that the censoring rate is ranged from 20% to 60%. These set-ups represent a practical scenario with substantial error contamination on the covariate. The results based on 1000 iterations are reported in Table 1. The prevalence of single-root patten is similar across different scenarios with various censoring rates. When the sample size is 100, the percentage of single-root pattern under all three distributions are close to or over 60%. Even when the sample size increases to 800, the prevalence of single-root pattern is still quite

Table 1
Prevalence (%) of single-root pattern for the parametric corrected score with $E(X)=0$, $\mathrm{Var}(X)=1$, $\beta=-1$, and $\varepsilon \sim \mathrm{Normal}(0,1)$.

Censoring rate	Distribution of X	Size			
		100	200	400	800
20%	Normal	68.0	52.1	38.0	20.3
	Modified chi-square	65.8	57.9	52.3	40.1
	Uniform	64.5	51.9	37.6	19.6
40%	Normal	62.4	49.0	36.3	17.8
	Modified chi-square	64.4	58.0	50.9	43.1
	Uniform	60.4	49.5	37.1	19.3
60%	Normal	61.0	46.5	32.1	16.0
	Modified chi-square	62.7	57.3	50.9	42.6
	Uniform	58.5	47.9	36.8	21.2

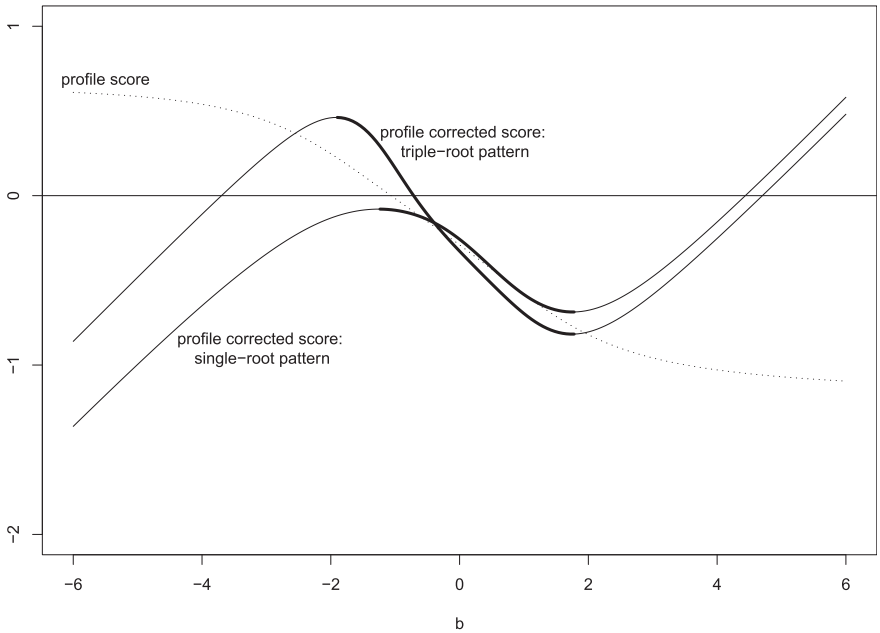

Fig. 1. Observed root patterns of the parametric corrected score $\eta(b)$. The true β is -1 and these two corrected curves correspond to the same profile score (with true covariates). Portion of a corrected curve is thickened to indicate negative derivative.

high.

For the conditional score, the patterns are more complicated. When the absolute value of β gets large, the estimating function fluctuates around zero and finally approaches zero as β goes to infinity. Therefore the conditional score may have many zero-crossings. When the β is not so extreme, two general patterns for the conditional scores are observed and plots from two simulation datasets are shown in Fig. 2. In the first pattern, the conditional score has one zero-crossing close to the true parameter. This single zero-crossing is decreasing and thus an appropriate one. In the second patten, the conditional score appears to have no proper zero-crossing near the truth though it may have multiple zero-crossings at extreme values of β. For conditional score, we define root-finding failure as following: we start with the naive estimator and search locally in the direction of the derivative of estimating function for a new estimator such that the ℓ^2-norm of the estimating functions decreases after each step. Repeat this step and we will eventually achieve a root or a local minimizer for the ℓ^2-norm of the estimating functions. We consider the latter case as root finding failure. Table 2 summarizes the prevalence of root-finding failure for the conditional score. The same simulation set up as in the corrected score is used. With a sample size of 100, the root-finding failure rate for conditional score varies from 3% to 5% for normal covariate and from 14% to 25% for modified chi-square covariate. As the sample size increases to 800, the failure rate drops to 0.2% for normal covariate but remains at least 14% for modified chi-square covariate.

The above investigation results show that, in the presence of substantial measurement error, both parametric corrected score and conditional score suffer from severe finite-sample pathological behaviors. Therefore improvements are required for these methods to have practical applicability. We observe that an appropriate zero-crossing of an estimating function should be a decreasing one. This observation suggests that the trend of estimating function is also informative and could be taken into account in the estimate determination. By Taylor expansion, the trend of estimating function may be quantified by its derivative. Recognizing this feature, Huang [9] proposed an approach to incorporate additional estimating functions which constrain the derivatives of the corrected score for the loglinear model. The estimation and inference are then accomplished by means of empirical likelihood. This approach

Table 2
Prevalence (%) of root-finding failure for the conditional score with $E(X)=0$, $\mathrm{Var}(X)=1$, $\beta=-1$, and $\varepsilon\sim$ Normal(0,1).

Censoring rate	Distribution of X	Size			
		100	200	400	800
20%	Normal	2.7	1.5	.6	.3
	Modified chi-square	14.4	12.2	15.1	13.9
	Uniform	3.1	1.6	.6	.1
40%	Normal	4.6	2.3	.8	.2
	Modified chi-square	20.0	16.7	20.9	18.7
	Uniform	3.1	2.1	.3	.5
60%	Normal	4.7	2.5	.7	.3
	Modified chi-square	24.6	21.7	25.7	25.0
	Uniform	5.7	2.7	1.4	1.0

effectively remedies the pathological behaviors of corrected score for loglinear model and also considerably improves the estimation efficiency. However, in the case of the Cox proportional hazards model, we are unable to construct additional estimating functions that effectively constrain the derivative of the parametric corrected score or conditional score because of the very nature of these two estimating functions. Nevertheless, Huang's approach provides an insight into a new approach to address pathological behaviors of estimating functions. If we could identify additional estimating functions that do not share the same wrong roots as the original estimating function, then by combining the original and additional estimating functions, pathological behaviors could be reduced or eliminated. But if either the original estimating function or the additional estimating functions vanish to zero, then wrong root sharing would easily arise. We have shown in the simulation that the conditional score would vanish to zero when the absolute value of parameter becomes large. Thus, the trend pattern of the parametric corrected score is more desirable than that of the conditional score and our method will be developed for the parametric corrected score.

3. Improving corrected score

3.1. Augmented estimation method

Motivated by Huang's trend-constrained corrected score [9], we

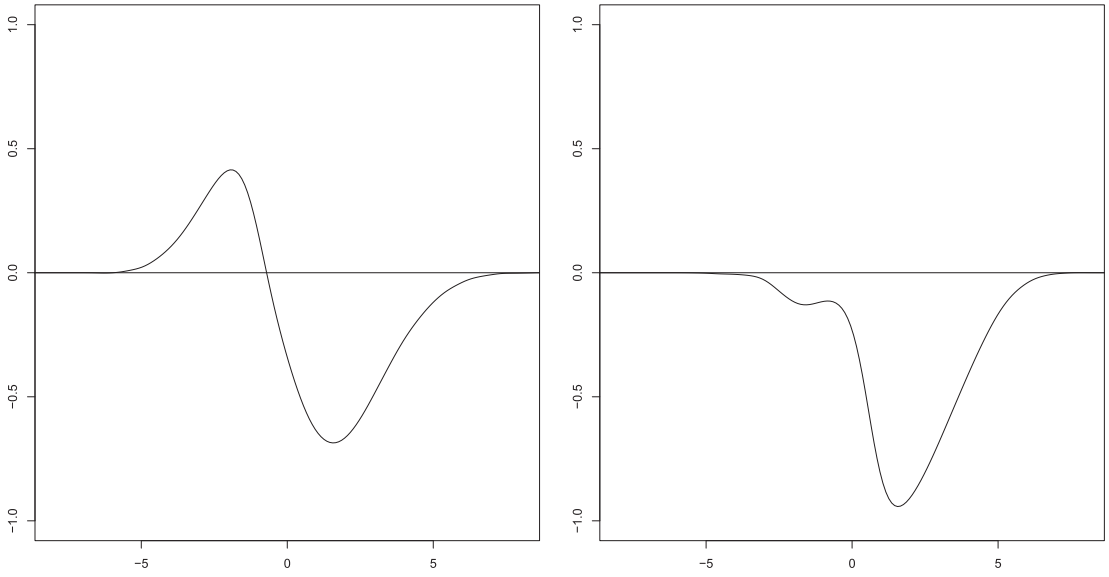

Fig. 2. Observed root patterns of the conditional score estimating function. The true β is -1.

first derive the following result:

Theorem 1. *Under the proportional hazards model and the classical additive measurement error model,*

$$\varepsilon\left[\int_0^\tau \frac{\partial^{k_1+\cdots+k_p}\exp\{\mathrm{b}'\mathbf{W}-\Omega(\mathrm{b})\}}{\partial b_1^{k_1}\cdots b_p^{k_p}}dN_i(t)\Big|_{\mathrm{b}=\mathbf{0}}\int_0^\tau \frac{\partial^{k_1+\cdots+k_p}Y(t)\exp\{\mathrm{b}'\mathbf{W}-\Omega(\mathrm{b})\}}{\partial b_1^{k_1}\cdots b_p^{k_p}}d\Lambda_0(t)\Big|_{\mathrm{b}=\beta}\right]=0, \tag{6}$$

for $k_l \geq 0, l=1,\ldots,p$, *where* $b_l, l=1,\ldots,p$, *is the lth element of* $\mathbf{b}$.

Equation (6) is useful in constructing additional estimating equations for β. When $\sum_{l=1}^p k_l = 0$ and 1, one may obtain the usual parametric corrected score. When $\sum_{l=1}^p k_l = 2$, the additional estimating functions are the upper triangular elements of the following symmetric matrix:

$$n^{-1}\sum_{i=1}^n\int_0^\tau\left[\left\{\mathbf{W}_i-\dot{\Omega}(0)\right\}^{\otimes 2}-\ddot{\Omega}(0)\right]dN_i(t)-\int_0^\tau Y_i(t)\exp\{\mathrm{b}'\mathbf{W}_i-\Omega(\mathrm{b})\}\left[\left\{\mathbf{W}_i-\dot{\Omega}(\mathrm{b})\right\}^{\otimes 2}-\ddot{\Omega}(\mathrm{b})\right]d\tilde{\Lambda}_0(t).$$

By profiling out $\tilde{\Lambda}_0(\cdot)$, we obtain

$$n^{-1}\sum_{i=1}^n\int_0^\tau\left[\left\{\mathbf{W}_i-\dot{\Omega}(0)\right\}^{\otimes 2}-\ddot{\Omega}(0)-\frac{\sum_{j=1}^n Y_j(t)\exp\{\mathrm{b}'\mathbf{W}_j-\Omega(\mathrm{b})\}\left[\left\{\mathbf{W}_j-\dot{\Omega}(\mathrm{b})\right\}^{\otimes 2}-\ddot{\Omega}(\mathrm{b})\right]}{\sum_{j=1}^n Y_j(t)\exp\{\mathrm{b}'\mathbf{W}_j-\Omega(\mathrm{b})\}}\right]dN_i(t) \tag{7}$$

The additional estimating functions would be helpful if both parametric corrected score and additional estimating function are close to 0 around the truth (not necessarily having roots) and the additional estimating function is not close to 0 when the parametric corrected score is close to 0 at any point far away from the truth. Fig. 3 shows four typical patterns of parametric corrected score and corresponding additional estimating function based on Equation (7) for a single-covariate model with true parameter $\beta=-1$. Plot (a) is the ideal scenario. The parametric corrected score have three zero-crossings. The additional estimating function shares the same decreasing zero-crossing as the parametric corrected score. Moreover, they do not share any wrong zero-crossings. In plot (b), the parametric corrected score have three zero-crossings and two of them are close to each other. In this case, discriminating the two roots around the truth is not very important since they are close to each other anyway. In plots (c) and (d), the parametric corrected score has a single wrong zero-crossing but the additional estimating function is not close to 0 near this wrong zero-crossing. Both estimating functions are close to 0 around the truth.

3.2. Estimation and inference

With the additional estimating functions, we have more estimating functions than the number of parameters. Available methods to synthesize estimating functions that exceed the number of parameters include empirical likelihood [14] and quadratic inference function (QIF) method [15]. In this research, we shall use the quadratic inference function method to determine the estimate since the estimating functions are not sums of iid terms, thus the empirical likelihood would be computational difficult.

Let $\boldsymbol{\varphi}(\mathbf{b})$ denotes the estimating functions. $\boldsymbol{\varphi}(\mathbf{b})$ is comprised of the original parametric corrected score and additional estimating functions. The quadratic inference function takes the form

$$Q\left(\mathrm{b};\widehat{\mathbf{C}}\right)=\boldsymbol{\varphi}'(\mathrm{b})\widehat{\mathbf{C}}^{-1}(\mathrm{b})\boldsymbol{\varphi}(\mathrm{b}), \tag{8}$$

where $\widehat{\mathbf{C}}(\mathbf{b})$ is any consistent estimator for the asymptotic variance of $\boldsymbol{\varphi}(\mathbf{b})$. Then the estimator is defined as the minimizer of (8) and is consistent for the true value of $\mathbf{b}$. Furthermore, the estimator is asymptotically efficient in the class of consistent estimators based on linear combination of parametric corrected score and additional estimating functions. The construction of a quadratic inference function helps to solve both aspects of pathological behaviors of the parametric corrected score. Firstly, in the case of multiple zero-crossings, the introduction of additional estimating functions helps to pick up the right zero-crossing out from multiple ones if the additional estimating functions do not share the same wrong roots as the original parametric corrected score. Secondly, the problem of no appropriate zero-crossing could be solved by minimizing the quadratic inference function.

We name the method to incorporate additional estimating functions based on (6) as *the augmented parametric corrected score*. One important special case is the method incorporating the upper triangular elements of matrix (7) and was termed *the second order augmented parametric corrected score*. Augmented parametric corrected scores with higher order are also available, with additional estimating functions corresponding to k_l such that $\sum_{l=1}^p k_l > 2$. In Section 4, we will conduct extensive simulation studies to evaluate the performance of the augmented parametric corrected score and its applicability in practice.

Fig. 4 plots quadratic inference functions corresponding to the two datasets in Fig. 1. The second order augmented parametric corrected score is adopted. The formula for a consistent estimator of the asymptotic variance of $\boldsymbol{\varphi}(\mathbf{b})$ is given in Appendix.

The interval estimation can be easily achieved by inverting the hypothesis testing statistics. As an inference function, $Q(\mathbf{b};\widehat{\mathbf{C}})$ has similar properties as the log-likelihood function [15].

(a) $Q(\beta_0)-Q(\widehat{\beta})$ is asymptotically chi-squared with degree of freedom p;
(b) the profile test statistics $Q(\psi_0,\widehat{\lambda}_0)-Q(\widehat{\psi},\widehat{\lambda})$, where (ψ,λ) is a partitioning of the parameter of β, is asymptotically chi-squared as a test of $H{:}\psi{=}\psi_0$, with degree of freedom equal to the dimension of ψ.

4. Numerical studies

4.1. Simulations

Extensive simulations studies were conducted to evaluate the performance of augmented parametric corrected score. For reference and comparison, the ideal, naive, regression calibration, conditional score, and parametric corrected score were also studied.

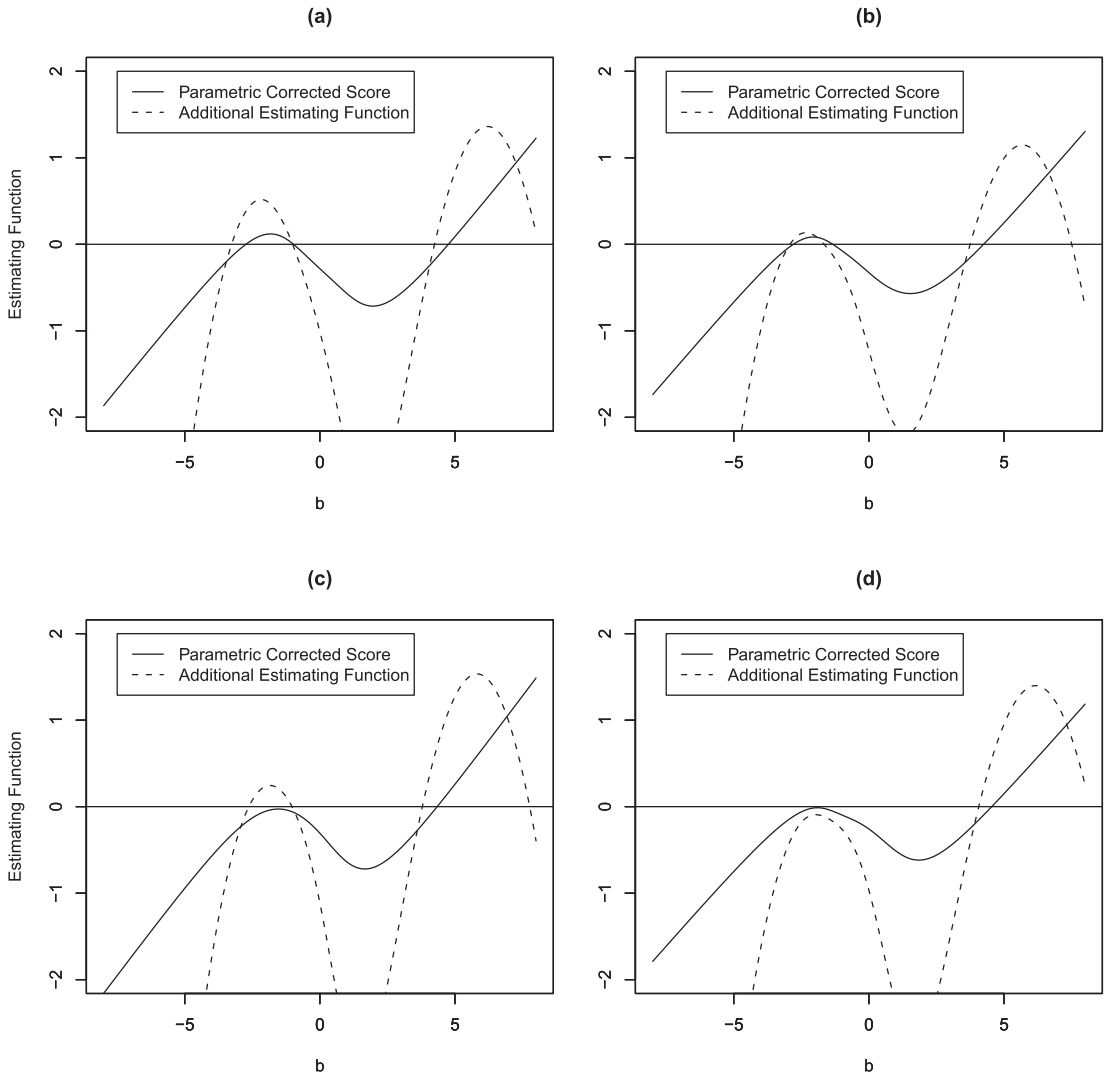

Fig. 3. Parametric corrected score and additional estimation function based on Equation (7) for a single-covariate model with $\beta = -1$.

The ideal estimator used the ordinary partial score function with true covariates and, of course, it is not a realistic estimator. The naive approach uses the mismeasured surrogates in place of true covariates in the partial score function. For the regression calibration method, $\mathbf{Z}_e$ is replaced by $E(\mathbf{Z}_e|\mathbf{Z}_a,\mathbf{W}_e)$ in the partial score function. For the proposed approach, the optimization algorithm of Nelder and Mead [16] will be used.

As shown in previous section, both conditional score and parametric corrected score have high prevalence of root-finding failure. To utilize all simulated data and conduct a fair comparison, we propose the following re-defined conditional score and re-defined parametric corrected score. If an appropriate zero-crossing could be found, the estimators will take the value of zero-crossing. But if root-finding failure occurs, the estimators will be defined as the local minimizer of the ℓ^2-norm of the estimating functions closest to the naive estimator. Operationally, we will use the following modified Newton–Raphson algorithm. We start with the naive estimator and calculate the Newton–Raphson step size. Since the goal is to find a root or local minimizer, we need prevent overshooting. In our simulation, we cap the step size at .2. During each iteration, we compare the ℓ^2-norm evaluated at new estimator to that evaluated at current estimator. If the ℓ^2-norm evaluated at new estimator is smaller, then the new estimator will be accepted and the algorithm continues to the next iteration. Otherwise, we will halve the step size and calculate the new estimator again. Iterations will be repeated until that i) the absolute value of estimating function is less than 10^{-6} or ii) the step size has been halved for more than 10 time during any single iteration. If criteria i) is satisfied, the algorithm converges to a zero-crossing and a root is identified in this case. If criteria ii) is satisfied, the algorithm converges to a local minimizer and root-finding failure occurs. Comparing to the original definition that estimators take the values of zero-crossings of estimating function, this new definition actually benefits the conditional score and corrected score. As shown in Figs. 1 and 2, in the case of root-finding failure, zero-crossings of conditional score and corrected score are at extreme and far away from the true value.

In the simulation, we consider both single- and double-covariate models. In the single-covariate models, the true covariate X is of mean 0 and variance 1. The true regression coefficient was set to be -1 and the baseline hazard is constant 1. The measurement error follows the standard normal distribution. Two different distributions of X were studied: A) standard normal distribution and B) modified chi-square distribution. To generate the modified chi-square distribution, the chi-square distribution with 1 degree of freedom was first truncated at 5 and then location-shifted

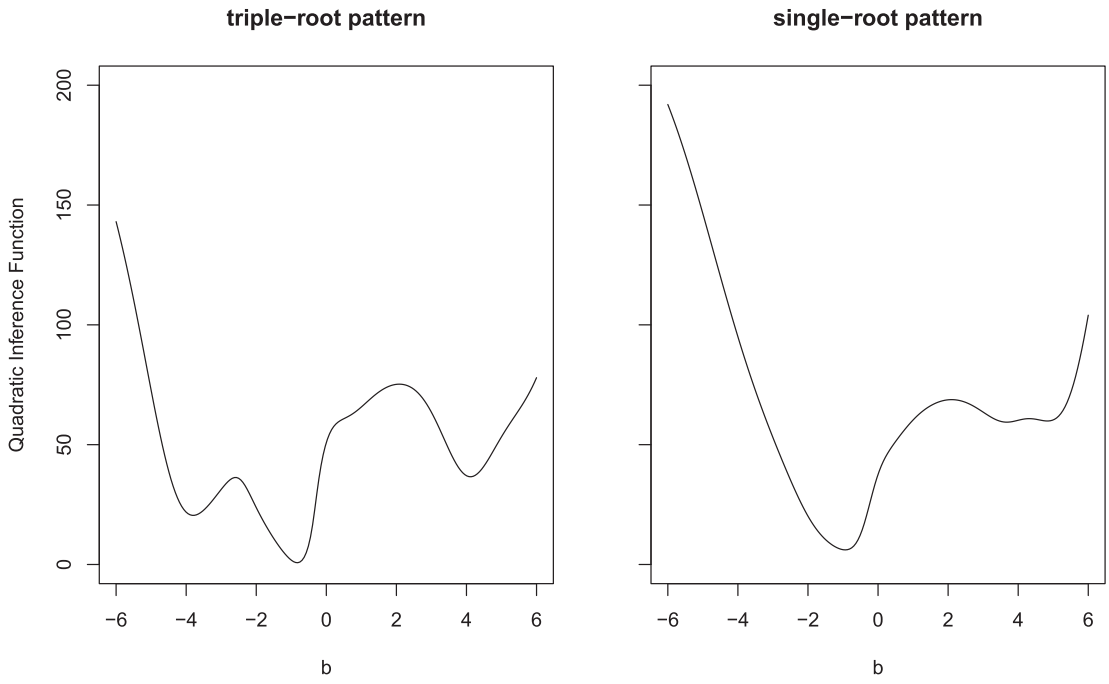

Fig. 4. Quadratic inference functions for the two datasets in Fig. 1.

to mean 0 and rescaled to variance 1. Censoring time was generated from a uniform distribution on $[0,\mu]$ and we will consider two different settings of censoring rate at 20% and 60%.

In the double-covariate models, true covariate X follows bivariate normal distribution with mean (0,0), variance (1,1) and correlation coefficient of .5. The first covariate was subject to a standard normal measurement error, whereas the second covariate was accurately measured. The regression coefficients were set to $(-1,1)$ and the baseline hazard is constant 1. Censoring time was also generated from a uniform distribution on $[0,\mu]$, where μ is chosen so that the censoring rate is 20% or 60%.

Sample sizes 100, 200, 400, 800, and 1600 were investigated. For each scenario, 1000 samples were simulated. We report the results on point and interval estimation separately.

Tables 3 and 4 summarize the simulation results on the estimators in the single-covariate models with censoring rates of 20% and 60% respectively. The quantile–quantile plots are shown in Figs. 5 and 6. For each scenario, the mean bias, and standard deviation were calculated. For augmented parametric corrected score, three sets of additional estimating functions were considered where $k = 2,3,4$. As expected, the naive estimator has substantial bias under both scenarios. The regression calibration estimator shows moderate bias, with larger bias in modified chi-square covariate case than normal covariate case. The re-defined conditional score shows slight bias under both scenarios, probably due to its left skewness. The quantile–quantile plots show that the re-defined conditional score deviates from normality considerably even when the sample size is 1600. It also has a much larger standard deviation comparing to the re-defined parametric corrected score and the augmented parametric corrected score. The re-defined parametric corrected score is unbiased for both scenarios and the standard deviation is small. All three augmented corrected scores are consistent under both scenarios, and they become less biased as sample size increases. In comparison with higher-order augmented corrected scores, the second order augmented corrected score seems more favorable overall in terms of bias and standard error. Compared to the re-defined parametric corrected score, the second order augmented corrected score has a larger

Table 3
Simulation summary statistics for the single-covariate models with 20% censoring rate: Ideal, naive(NV), regression calibration (RC), re-defined conditional score (ConS), re-defined parametric corrected score (CS), and first k augmented parametric corrected score (ACS: k), k = 2, 3, 4.

Size	Ideal		NV		RC		ConS			CS			ACS: 2		ACS: 3		ACS: 4	
	B	SD	B	SD	B	SD	F	B	SD	F	B	SD	B	SD	B	SD	B	SD
Normal covariate																		
100	−15	154	593	99	148	254	2.7	−376	920	68.0	−11	287	46	431	−43	414	−103	381
200	−8	104	602	65	189	163	1.5	−251	719	52.1	−89	273	−60	330	−130	348	−172	351
400	−5	73	601	48	195	113	.6	−202	575	38.0	−146	282	−99	287	−153	313	−184	327
800	1	49	604	32	206	76	.3	−99	351	20.0	−129	267	−87	246	−83	226	−123	256
1600	1	36	605	23	208	53	0	−43	200	5.6	−75	206	−45	163	−60	178	−62	166
χ^2 Covariate																		
100	−25	204	639	91	242	247	14.4	−485	1221	65.8	145	241	110	387	41	409	9	409
200	−5	136	640	63	267	155	12.2	−396	1022	57.9	47	241	−10	351	−81	381	−128	381
400	−6	98	642	43	280	101	15.1	−329	895	52.3	−21	237	−10	285	−76	306	−122	323
800	−3	70	641	32	281	72	13.9	−249	740	40.1	−70	245	−34	241	−81	237	−95	241
1600	0	47	644	21	289	47	12.5	−139	486	30.3	−82	242	−9	152	−29	133	−42	139

Note: F: root-finding failure (%); B: mean bias ($\times 10^3$); SD: standard deviation($\times 10^3$).

Table 4
Simulation summary statistics for the single-covariate models with 60% censoring rate: Ideal, naive(NV), regression calibration (RC), re-defined conditional score (ConS), re-defined parametric corrected score (CS), and first k augmented parametric corrected score (ACS: k), k = 2, 3, 4.

Size	Ideal		NV		RC		ConS			CS			ACS: 2		ACS: 3		ACS: 4	
	B	SD	B	SD	B	SD	F	B	SD	F	B	SD	B	SD	B	SD	B	SD
Normal covariate																		
100	−24	206	550	134	57	335	4.7	−429	1014	61.0	−27	301	94	478	−24	502	−90	551
200	−15	142	562	88	109	206	2.5	−272	738	46.5	−101	303	10	298	−99	347	−188	395
400	−8	97	563	63	119	142	.7	−197	578	32.1	−139	305	−27	235	−96	257	−152	294
800	−2	65	567	42	132	96	.3	−100	379	16.0	−113	271	−33	189	−72	199	−103	223
1600	1	46	567	30	133	69	0	−51	238	4.9	−68	205	−17	128	−38	125	−57	148
χ^2 covaraite																		
100	−66	351	687	124	351	288	24.6	−259	1154	62.7	247	297	196	459	219	531	260	568
200	−24	239	691	84	370	190	21.7	−291	1042	57.3	133	287	81	359	44	429	54	496
400	−16	159	693	57	382	124	25.7	−202	841	50.9	57	268	37	322	−22	368	−52	423
800	−9	112	693	40	385	86	25.0	−172	684	42.6	−8	263	1	280	−55	276	−94	306
1600	0	78	697	28	394	60	21.5	−77	412	34.0	−29	258	−3	232	−29	194	−57	207

Note: Same as in that of Table 3.

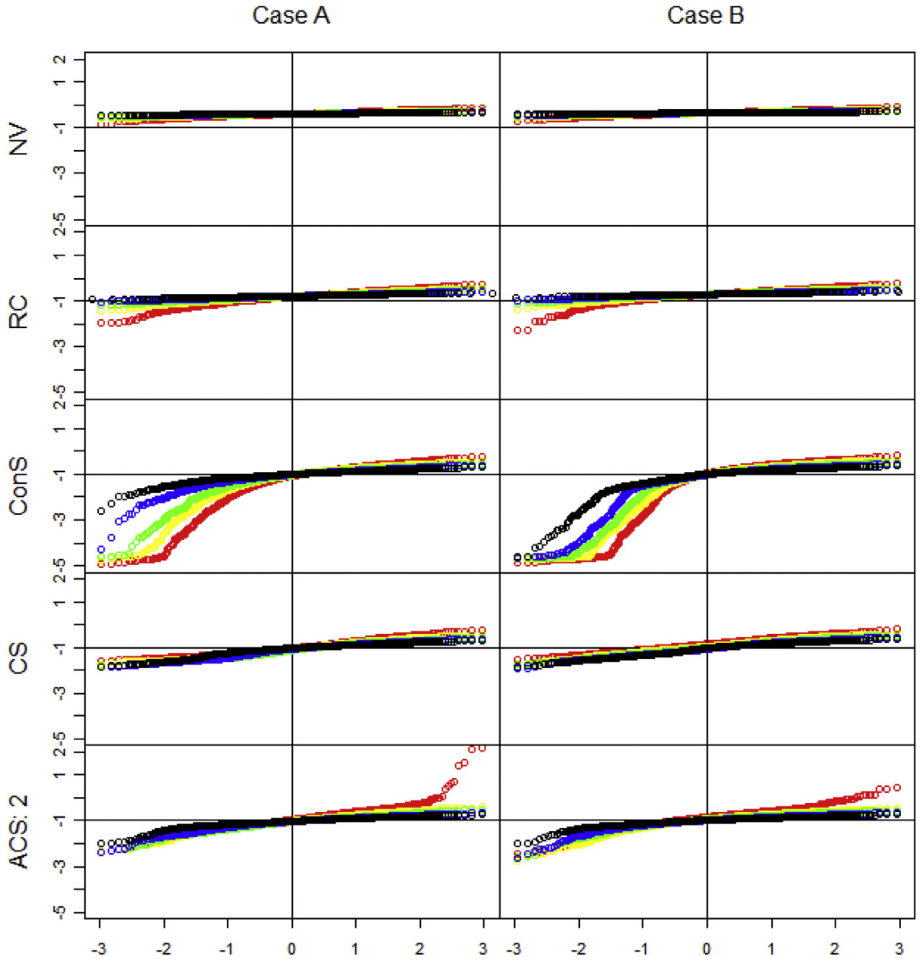

Fig. 5. Quantile–quantile plots for β in the single-covariate models with 20% censoring rate, where $\beta = -1$. Red, yellow, green, blue, and black correspond to sample sizes 100, 200, 400, 800, and 1600.

standard deviation when the sample size is small. But the standard deviation of augmented corrected score decreases rapidly as the sample size increases and is smaller than that of the re-defined corrected score when then sample size is 1600.

Tables 5 and 6 show the simulation results for the double-covariate models. As expected, when two covariates are correlated, the measurement error generally has impact not only on the mismeasured covariate but also on that of the accurately measured one. The relative performance of all estimators is similar to what was observed in the single-covariate models. For multiple-covariate models, each set of additional estimating functions contains more than one element. For example, when $k = 2$, the additional estimating functions includes three elements in the upper triangle of matrix (3.7). Various augmented corrected score could be constructed depending on the additional estimating functions chosen. Since the second order augmented corrected score shows a

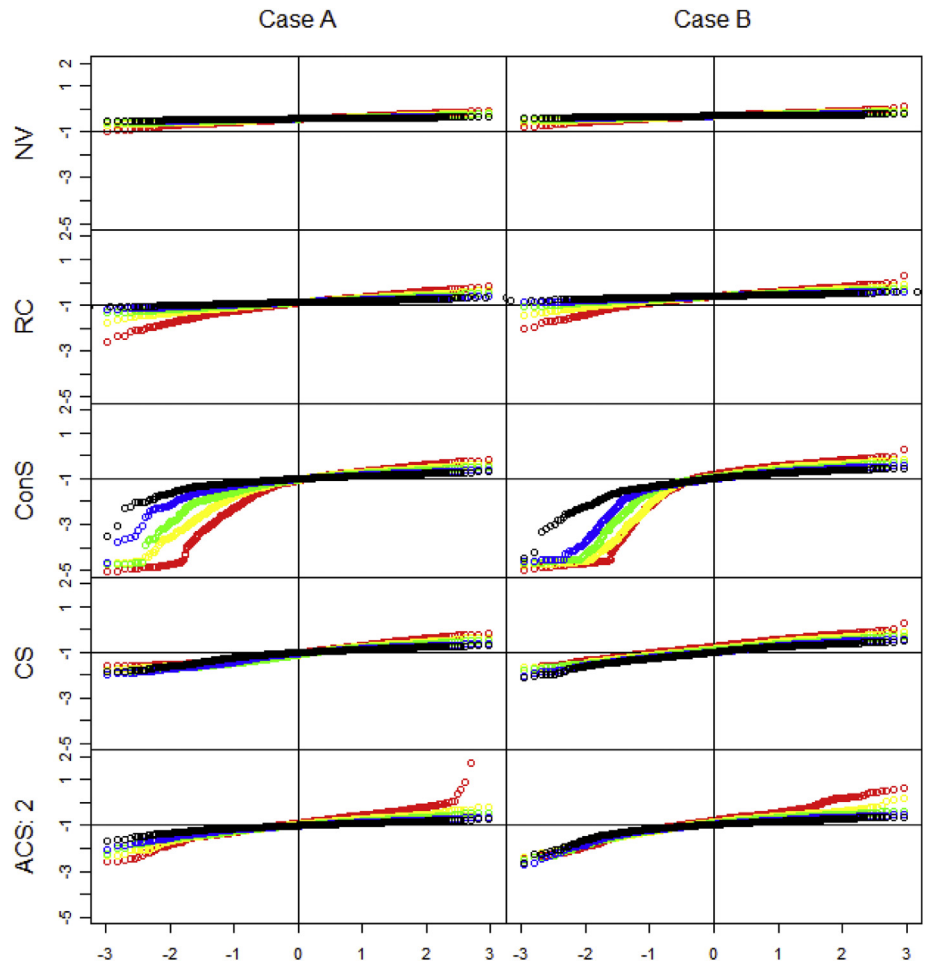

Fig. 6. Quantile–quantile plots for β in the single-covariate models with 60% censoring rate, where $\beta = -1$. Red, yellow, green, blue, and black correspond to sample sizes 100, 200, 400, 800, and 1600.

Table 5
Simulation summary statistics for the double-covariate models with 20% censoring rate: Ideal, naive(NV), regression calibration (RC), re-defined conditional score (ConS), re-defined parametric corrected score (CS), and augmented parametric corrected score (ACS: 1/3, ACS: 2/3, ACS: 2).

Size		Ideal		NV		RC		ConS			CS			ACS: 1/3		ACS: 2/3		ACS: 2	
		B	SD	B	SD	B	SD	F	B	SD	F	B	SD	B	SD	B	SD	B	SD
100	β_1	−23	175	637	101	66	381	57.5	215	348	72.0	267	322	170	532	245	514	308	538
	β_2	23	171	−400	142	−113	246		−67	336		−100	327	−60	456	−97	404	−142	399
200	β_1	−13	114	648	66	136	224	46.9	155	315	62.8	186	294	−18	382	42	375	92	339
	β_2	16	117	−412	101	−154	161		−35	310		−47	286	54	349	18	327	−20	276
400	β_1	−5	82	649	48	168	144	33.4	65	297	46.2	71	280	−89	301	−56	277	−28	255
	β_2	4	80	−416	68	−177	103		5	257		15	243	85	269	60	233	40	210
800	β_1	−2	55	651	33	181	93	20.0	22	224	30.5	−3	252	−99	282	−80	262	−56	220
	β_2	1	56	−420	52	−185	74		16	186		44	203	88	255	72	230	52	186
1600	β_1	−2	40	652	23	185	66	8.0	−13	178	10.5	−43	207	−71	219	−67	204	−61	190
	β_2	1	39	−419	34	−186	49		21	134		45	155	59	180	57	163	51	145

F: root-finding failure (%); B: mean bias ($\times 10^3$); SD: standard deviation($\times 10^3$).
ACS: 1/3 and ACS: 2/3 correspond to the additional estimating functions containing the (1,1) element and the first row elements of matrix (7), respectively. ACS: 2 is the second order augmented parametric corrected score.

more favorable overall performance in the single-covariates models, we consider only the second order augmented corrected score and augmented corrected scores using a subset of the three elements in this simulation. Simulation results show that the second order augmented corrected score performs better than two other augmented corrected scores. The bias of the second order augmented corrected score reduced quickly as sample size increased. Meanwhile, it also has the smallest standard deviation.

Tables 7 and 8 report the coverage of three types of 95% confidence intervals in the single- and double-covariate models. All three scenarios use the second order augmented corrected score. In constructing the confidence interval, we use two different approaches: inverting the hypothese testing statistics as introduced in Section 2.2, and the Wald-type confidence interval. For the former, we use two critical values based on the asymptotic chi-square distribution and the bootstrap calibration [17]. Bootstrap size of

Table 6
Simulation summary statistics for the double-covariate models with 60% censoring rate: Ideal, naive(NV), regression calibration (RC), re-defined conditional score (ConS), re-defined parametric corrected score (CS), and augmented parametric corrected score (ACS: 1/3, ACS: 2/3, ACS: 2).

Size		Ideal		NV		RC		ConS			CS			ACS: 1/3		ACS: 2/3		ACS: 2	
		B	SD	B	SD	B	SD	F	B	SD	F	B	SD	B	SD	B	SD	B	SD
100	β_1	−32	242	610	140	−13	513	43.3	96	428	64.8	239	320	225	668	326	666	392	660
	β_2	33	245	−352	206	−40	336		−27	408		−72	362	−50	536	−104	500	147	465
200	β_1	−20	160	617	95	59	286	36.1	29	393	57.8	145	300	14	387	93	360	140	344
	β_2	21	161	−362	138	−81	209		24	344		−8	301	48	369	−7	310	−41	264
400	β_1	−6	109	619	65	97	182	21.5	−2	293	40.4	45	279	−65	321	−9	273	30	241
	β_2	6	111	−366	97	−106	137		31	225		36	254	79	301	35	246	6	197
800	β_1	−1	72	621	45	112	120	10.8	−23	249	25.7	−16	257	−73	267	−32	213	−9	192
	β_2	2	76	−369	69	−115	94		34	198		55	214	72	236	40	189	23	160
1600	β_1	−2	53	622	33	114	87	7.6	−24	186	11.3	−46	212	−55	193	−40	171	−29	162
	β_2	0	53	−371	46	−117	65		28	142		47	161	46	162	36	140	26	124

Note: Same as in that of Table 5.

Table 7
Coverage of 95% confidence interval for the second order augmented parametric corrected score with 20% censoring rate. C (chi-square distribution), BC (bootstrap calibration) and W (Wald-type) indicate the type of confidence interval.

Size	100			200			400			800			1600		
	C	BC	W	C	BC	W	C	BC	W	C	BC	W	C	BC	W
Single-covariate: normal covariate															
β	90.2	96.8	87.0	91.7	96.6	91.1	87.1	96.6	93.8	87.6	97.1	94.8	91.8	95.6	94.9
Single-covariate: χ^2 covariate															
β	82.0	92.9	76.8	88.8	92.2	84.3	87.6	94.3	85.6	90.0	94.6	89.2	91.6	94.8	91.8
Double-covariate															
β_1	75.8	94.5	78.4	86.3	95.4	87.3	88.3	96.3	94.0	89.7	96.7	95.8	89.4	96.2	97.2
β_2	79.2	93.0	83.9	86.8	93.8	92.3	88.1	92.3	95.6	86.7	93.5	96.2	89.4	95.5	97.6

Table 8
Coverage of 95% confidence interval for the second order augmented parametric corrected score with 60% censoring rate. C (chi-square distribution), BC (bootstrap calibration) and W (Wald-type) indicate the type of confidence interval.

Size	100			200			400			800			1600		
	C	BC	W	C	BC	W	C	BC	W	C	BC	W	C	BC	W
Single-covariate: normal covariate															
β	89.5	97.6	86.8	94.2	96.7	92.0	91.5	96.8	94.7	93.6	96.2	96.3	98.5	95.5	97.5
Single-covariate: χ^2 covariate															
β	87.5	93.7	73.9	89.7	92.4	79.1	88.4	93.2	80.4	90.4	93.9	85.8	91.8	94.3	88.6
Double-covariate															
β_1	92.0	92.6	80.7	93.8	96.6	89.6	92.6	94.9	94.1	93.2	96.9	95.8	92.6	96.2	98.0
β_2	92.0	93.9	87.3	93.5	94.5	93.7	92.3	92.7	96.0	91.5	93.4	96.3	92.7	95.8	97.1

500 is used for the bootstrap calibration. The Wald-type confidence interval and test based one with chi-square distribution critical values have poor coverage when the sample size is small, but improve with larger sample size. The coverage probability of test based confidence interval using bootstrap-calibrated critical value is close to the nominal level of 95% for all sample sizes, but this method is much more computational intensive.

4.2. Application to ACTG 175 data

We apply the proposed approach to the AIDS Clinical Trial Group (ACTG) 175 study, a randomized clinical trial to evaluate four treatments in HIV-infected patients with an initial screening CD4 counts of between 200 and 500 per cubic millimeter. A total of 2467 patients were enrolled and an almost equal number of patients were randomized into each of the four treatment groups: zidovudine alone (ZDV), zidovudine plus didanosine (ZDV + ddI), zidovudine plus zalcitabine (ZDV + ddC), and zalcitabine alone (ddC). We are interested in assessing the effect of baseline CD4 count on time to AIDS or death in antiretroviral-naive patients. Among all study patients, 1067 had no prior antiretroviral therapy at enrollment, among which 1036 patients had two CD4 measurements prior to the start of treatment and within 3 weeks of randomization. For this analysis, we will consider the subset of 1036 patients. The median length of follow-up was 32 months, and 85 events were observed.

Table 9
Comparison of regression coefficient estimators in the ACTG 175 data.

	log(CD4)		ZDV + ddI		ZDV + ddC		ddI	
	Est	Var	Est	Var	Est	Var	Est	Var
NV	−1.838	.1183	−.652	.0881	−.895	.1006	−.598	.0802
ConS	−2.172	.1698	−.659	.0916	−.892	.1024	−.604	.0835
CS	−2.177	.1678	−.659	.0900	−.892	.1012	−.604	.0819
ACS	−2.177	.1422	−.657	.0991	−.876	.1028	−.596	.0827
		.1678		.0905		.0964		.0811

Note: For proposed estimator, the first row of variance estimator is obtained by inverting hypothese testing statistics with bootstrap critical value; the second row of values is from sandwich variance estimator. Est: Estimated coefficient; Var: Variance.

We consider a Cox regression mode with 4 covariates: the true baseline log(CD4) and three indicators for the four treatments with ZDV group as the reference. We define the baseline log(CD4) as the average of the two log(CD4) measurements. From the duplicated measurements, we estimated the variance for error and true underlying log(CD4) to be .033 and .076 respectively. Note that the variance of measurement error is estimated using two replicated measurements of baseline log(CD). Therefore there is an additional estimating function for the variance of measurement error. Table 9 shows the estimators based on the naive, conditional score, parametric corrected score, and the proposed augmented corrected score. In comparison, the naive approach gives an coefficient estimator of log(CD4) with substantially smaller magnitude. All the other approaches have similar estimates for all coefficients.

5. Discussion

Measurement error is a common issue for many clinical trials. The main reason for the presence of measurement error is study participants' natural biological fluctuation. The imprecision of measurement tool is another contributor. One example of measurement error is from the Nutritional Prevention of Cancer (NPC) trial, which involved multiple recurrences of skin cancers, basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) [18]. Baseline plasma selenium as an important prognostic risk factor for BCC and SCC is subject to substantial measurement error. The aforementioned AIDS Clinical Trial Group (ACTG) 175 study is another example of measurement error in clinical trial.

For proportional hazards model with covariate measurement error, several consistent methods have been proposed under the functional modeling framework, including the conditional score and the parametric corrected score. However, when the measurement error is substantial to the extent that the errors are comparable to the true covariates in variance, both methods might experience pathological behaviors and root finding failure. Recently, Huang [9] developed a novel approach to incorporate additional estimating functions which constrain the derivatives of the parametric corrected score. That approach proves effective and eliminates finite sample pathological behaviors of parametric corrected score for the loglinear model. Motivated by Huang's approach, we conduct an investigation on the pathological behaviors of parametric corrected score and conditional score and propose an augmented parametric corrected score for the proportional hazards model by incorporating additional estimating functions to the original parametric corrected score. Results of simulation studies show the proposed approach is effective in eliminating pathological behaviors even with small sample size and substantial measurement error. The variance of proposed estimator appears to be larger than the parametric corrected score when sample size is smaller than 400, but it decreases rapidly and become smaller than the parametric corrected score as the sample size increases to 400. With the ability of eliminating finite sample pathological behaviors of conditional score and the parametric corrected score, this proposed augmented parametric corrected score could be widely used for time-to-event data in clinical trials where covariate measurement error issue is of concern.

In this paper, we have only considered the situation where the distribution of the measurement error is known. With additional data available on the measurement error, the parametric distribution imposed on the measurement error may be spared. With the availability of replicated mismeasured covariates, Huang and Wang [7] developed a nonparametric corrected score method for the proportional hazards model. Extension of the approach of incorporating additional estimating functions to the nonparametric corrected score is currently under study.

Appendix

A Asymptotic variance of estimating function

For simplicity, we consider the single-covariate model given in Section 2.2. The second order augmented corrected score has the form

$$\varphi(b) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left\{\begin{pmatrix} W_i + b\sigma^2 \\ W_i^2 - \sigma^2 \end{pmatrix} - \frac{\sum_{j=1}^{n} Y_j(t)\exp(bW_j)\begin{pmatrix} W_j \\ \left(W_j - b\sigma^2\right)^2 - \sigma^2 \end{pmatrix}}{\sum_{j=1}^{n} Y_j(t)\exp(bW_j)}\right\} dN_i(t).$$

With functional delta method [7], straightforward algebra gives

$$n^{1/2}\varphi(b) = n^{-1/2}\sum_{i=1}^{n}(B_{i1} - B_{i2}) + o_p(1)$$

where

$$B_{i1} = \int_0^{\tau}\left\{\begin{pmatrix} W_i + b\sigma^2 \\ W_i^2 - \sigma^2 \end{pmatrix} - \frac{\varepsilon\left\{Y(t)\exp(bW)\begin{pmatrix} W \\ \left(W - b\sigma^2\right)^2 - \sigma^2 \end{pmatrix}\right\}}{\varepsilon\{Y(t)\exp(bW)\}}\right\} dN_i(t), \text{ and}$$

$$B_{i2} = \int_0^{\tau}\left\{\frac{\exp(bW_i)\begin{pmatrix} W_i \\ \left(W_i - b\sigma^2\right)^2 - \sigma^2 \end{pmatrix}}{\varepsilon\{Y(t)\exp(bW)\}} - \frac{\varepsilon\left\{Y(t)\exp(bW)\begin{pmatrix} W \\ \left(W - b\sigma^2\right)^2 - \sigma^2 \end{pmatrix}\right\}\exp(bW_i)}{[\varepsilon\{Y(t)\exp(bW)\}]^2}\right\} Y_i(t)d\mathscr{E}N(t).$$

Thus, $n^{1/2}\varphi(b)$ is asymptotically a sum of iid random variables.

For fixed b, $n^{1/2}\varphi(b)$ is asymptotically normal with a covariance matrix $\Sigma(b)$ that can be consistently estimated by

$$\widehat{\Sigma}(b) = n\sum_{i=1}^{n}\{\omega_i(b) - \overline{\omega}(b)\}\{\omega_i(b) - \overline{\omega}(b)\}'$$

where $\omega_i(b) = n^{-1}(B_{i3} - B_{i4})$, $\overline{\omega}(b) = n^{-1}\sum_{i=1}^{n}\omega_i(b)$, and B_{i3} and B_{i4} are defined as

$$B_{i3} = \int_0^\tau \left\{ \begin{pmatrix} W_i + b\sigma^2 \\ W_i^2 - \sigma^2 \end{pmatrix} - \frac{\widehat{\varepsilon}\left\{ Y(t)\exp(bW) \begin{pmatrix} W \\ \left(W - b\sigma^2\right)^2 - \sigma^2 \end{pmatrix} \right\}}{\widehat{\varepsilon}\{Y(t)\exp(bW)\}} \right\} dN_i(t), \text{ and}$$

$$B_{i4} = \int_0^\tau \left\{ \frac{\exp(bW_i) \begin{pmatrix} W_i \\ \left(W_i - b\sigma^2\right)^2 - \sigma^2 \end{pmatrix}}{\widehat{\varepsilon}\{Y(t)\exp(bW)\}} - \frac{\widehat{\varepsilon}\left\{ Y(t)\exp(bW) \begin{pmatrix} W \\ \left(W - b\sigma^2\right)^2 - \sigma^2 \end{pmatrix} \right\} \exp(bW_i)}{\left[\widehat{\varepsilon}\{Y(t)\exp(bW)\}\right]^2} \right\} Y_i(t) d\widehat{\varepsilon} N(t).$$

Asymptotic variance of other augmented corrected scores could be derived similarly.

B Proof of Theorem 1

Given

$$\varepsilon[\exp\{\mathrm{b}'\mathbf{W} - \Omega(\mathrm{b})\}|\mathbf{Z}] = \exp(\mathrm{b}'\mathbf{Z})$$

under additive measurement error model, one may obtain

$$\varepsilon\left[\int_0^\tau \frac{Y \partial^{k_1+\cdots+k_p} \exp\{\mathrm{b}'\mathbf{W} - \Omega(\mathrm{b})\}}{\partial b_1^{k_1} \cdots b_p^{k_p}} d\Lambda_0(t) \middle| \mathbf{Z} \right] = \prod_{l=1}^p Z_l^{k_l} \int_0^\tau Y(t)\exp(\mathrm{b}'\mathbf{Z}) d\Lambda_0(t)$$

and

$$\varepsilon\left[\int_0^\tau \frac{\partial^{k_1+\cdots+k_p} \exp\{\mathrm{b}'\mathbf{W} - \Omega(\mathrm{b})\}}{\partial b_1^{k_1} \cdots b_p^{k_p}} dN(t)\big|_{\mathbf{b}=0} \middle| \mathbf{Z} \right] = \prod_{l=1}^p Z_l^{k_l} N(\tau).$$

Then given the fact that $M(t) = N(t) - \int_0^t Y(u)\exp(\boldsymbol{\beta}'\mathbf{Z}) d\Lambda_0(u)$ is a mean zero martingale, Equation (6) is implied by the above two equations.

References

[1] R.J. Carroll, D. Ruppert, L.A. Stefanski, C. Crainiceanu, Measurement Error in Nonlinear Models: A Modern Perspective, Chapman & Hall, London, 2006.
[2] L.A. Stefanski, R.J. Carroll, Conditional scores and optimal scores for generalized linear measurement-error models, Biometrika 74 (4) (1987) 703–716.
[3] T. Nakamura, Corrected score function for errors in variables models: methodology and application to generalized linear models, Biometrika 77 (1990) 127–137.
[4] L.A. Stefanski, Unbiased estimation of a nonlinear function of a noral mean with application to measurement error models, Commun. Statistics Theory Methods 18 (1989) 4335–4358.
[5] Y.J. Huang, C.Y. Wang, Consistent functional methods for logistic regression with errors in covariates, J. Am. Statistical Assoc. 96 (2001) 1469–1482.
[6] Y.J. Huang, C.Y. Wang, Nonparametric Correction to Errors in Covariates, Fred Hutchinson Cancer Research Center, Seattle, 1999.
[7] Y.J. Huang, C.Y. Wang, Cox regression with accurate covariates unascertainable: A nonparametric-correction approach, J. Am. Statistical Assoc. 95 (2000) 1209–1219.
[8] Y.J. Huang, C.Y. Wang, Errors-in-covariates effect on estimating functions: additivity in limit and non-parametric correction, Stat. Sin. 16 (2006) 861–881.
[9] Y.J. Huang, Corrected score with sizable covariate measurement error: pathology and remedy, Stat. Sin. 24 (2014) 357–374.
[10] T. Nakamura, Proportional hazards models with covariates subject to measurement error, Biometrics 48 (1992) 829–838.
[11] F.H. Kong, M. Gu, Consistent estimation in Cox proportional hazards model with covariate measurement erros, Stat. Sin. 9 (1999) 953–969.
[12] A.A. Tsiatis, M. Davidian, A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error, Biometrika 88 (2001) 447–458.
[13] X. Song, Y.J. Huang, On corrected score approach for proportional hazards model with covariate measurement error, Biometrics 61 (2005) 702–714.
[14] J. Qin, J. Lawless, Empirical likelihood and general estimating equations, Ann. Statistics 22 (1994) 300–325.
[15] B.G. Lindsay, A. Qu, Inference functions and quadratic score tests, Stat. Sci. 18 (2003) 394–410.
[16] J.A. Nelder, R. Mead, A simplex algorithm for function minimization, Comput. J. 7 (1965) 308–313.
[17] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, New York, 1993.
[18] L.C. Clark, G.F. Combs, B.W. Turnbull, E.H. Slate, D.K. Chalker, J. Chow, et al., Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin: a randomized controlled tria, J. Am. Med. Assoc. 276 (24) (1996) 1957–1963.