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
2451-8654/

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]:

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

which has the same expectation as reference (2) asymptotically for each and every finite **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

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

In fact, it can be shown that estimators from parametric