NIH Public Access
Author Manuscript
J Am Stat Assoc. Author manuscript; available in PMC 2015 December 01.

Published in final edited form as:
J Am Stat Assoc. 2014 December 1; 109(504): 1636–1646. doi:10.1080/01621459.2014.896805.

Proportional Hazards Model with Covariate Measurement Error and Instrumental Variables

Xiao Song and
Associate Professor, Department of Epidemiology and Biostatistics, University of Georgia, Athens, GA 30602

Ching-Yun Wang
Member, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109

Xiao Song: xsong@u.uga.edu; Ching-Yun Wang: cywang@fhcrc.org

Abstract

In biomedical studies, covariates with measurement error may occur in survival data. Existing approaches mostly require certain replications on the error-contaminated covariates, which may not be available in the data. In this paper, we develop a simple nonparametric correction approach for estimation of the regression parameters in the proportional hazards model using a subset of the sample where instrumental variables are observed. The instrumental variables are related to the covariates through a general nonparametric model, and no distributional assumptions are placed on the error and the underlying true covariates. We further propose a novel generalized methods of moments nonparametric correction estimator to improve the efficiency over the simple correction approach. The efficiency gain can be substantial when the calibration subsample is small compared to the whole sample. The estimators are shown to be consistent and asymptotically normal. Performance of the estimators is evaluated via simulation studies and by an application to data from an HIV clinical trial. Estimation of the baseline hazard function is not addressed.

Keywords

Generalized methods of moments; Nonparametric correction; Survival

1. INTRODUCTION

Survival data often arise in biomedical studies where the outcome of interest is time to an event of interest (failure). The proportional hazards model is the most widely used survival model to characterize the relationship between survival time and covariates. However, some covariates, say X, could be measured with error in practice. For example, important covariates such as CD4 counts in HIV studies are subjected to substantial measurement error due to both imperfect instruments and biological fluctuation.

It is well-known that the naive approach that ignores measurement error can lead to biased estimation and erroneous inference (e.g. Prentice, 1982). Various approaches have been proposed to deal with measurement error. The regression calibration (Prentice, 1982; Wang et al. 1997; Dafni and Tsiatis, 1998; Liao et al., 2011) approximates the hazard function conditional on the observed covariates. It can reduce estimation bias but is still inconsistent. Likelihood based approaches are usually computationally intensive (e.g. Wulfson and Tsiatis, 1997; Hu, Tsiatis and Davidian, 1998, Song et al., 2002a; Wang, 2008). Consistent estimation based on corrected scores (parametric correction) was first proposed by Nakamura (1992) which required no distributional assumption on the underlying true covariates, but the standard deviation was assumed known. Huang and Wang (2000) developed a nonparametric correction approach that further relaxed the distributional assumption on the measurement error, but required repeated measurements. The correction approaches were extended to more general measurement error models (Hu and Lin, 2002; Wang, 2006; Tapsoba et al., 2011) and more general error assessment sets (Huang and Wang, 2006). A related approach is the conditional score (Tsiatis and Davidian, 2001; Song et al., 2002b), which is asymptotic equivalent to the corrected score. The conditional score approach has better finite sample performance (Song and Huang, 2005), but still depends on the normality assumption of the error. Motivated by the difference in the corrected score and conditional score estimating functions, Song and Huang (2005) proposed refined parametric correction and nonparametric correction approaches. The refined parametric correction estimator has comparable finite sample performance as the conditional score estimator. While the literature of proportional hazards regression with covariate measurement error is rich, to our knowledge, existing approaches require either knowledge of the measurement error standard deviation, repeated error-prone measurements, longitudinal error-prone measurements, or a validation set. Such information may not be available in practice.

Instead, instrumental variables may be observed in a subset of the sample. Instrumental variables are variables correlated with X, independent of the measurement error, and independent of the outcome given the covariates (Carroll et al., 2006, chapter 6; Stock and Watson, 2010, chapter 12). They are widely used in econometrics when the covariates are correlated with disturbance due to omitted variables, errors-in-variables, or simultaneous causality (Stock and Watson, 2010, chapter 12). Standard approaches that ignore the correlation between the covariates and disturbance usually lead to inconsistent estimators. Instrumental variables are used to obtain consistent estimators of the regression coefficients under this situation. Here we consider the case when the instrumental variables are observed in a subset of the sample. An example is AIDS clinical Trials Group (ACTG) 175, a randomized trial to compare zidovudine alone, ziduvudine plus didanosine, zidovudine plus zalcitabine, or didanosine alone in HIV-infected subjects on the basis of time to progression to AIDS or death (Hammer et al., 1996). It is of interested to assess the effect of treatments on survival time adjusted for baseline CD4 counts X. The closest CD4 measurement within one week before randomization was taken as the baseline CD4 measurements. It is well known that observed CD4 counts are contaminated by substantial measurement error. Among the 2174 randomized patients with at least one baseline CD4 measurements, there were no replicated baseline CD4 measurements on the same day. However, 989 patients had at least one CD4 measurement between one to three weeks prior to randomization. Since the underlying true CD4 counts might change over time, these CD4 measurements were not simple replication of baseline CD4 counts. But they may be used as instrumental variables. Figure 1 shows the scatter plot and a Loess smooth of log CD4 counts within one to three weeks versus one week before randomization. The logarithmic transformation was applied to CD4 counts to achieve approximate constant variance. The Loess curve indicates a possible nonlinear relationship between log CD4 counts during these two time periods.

Instrumental variables have been used in literature to deal with measurement error when there are no replicates or validation datasets (Carroll et al., 2006, chapter 6), mostly based on a parametric model between the instrumental variables and the covariates. But it may not be easy to identify the relationship between the instrumental variables and the covariates when the covariates are measured with error. Carroll et al. (2004) relaxed the parametric model assumption and adopted a varying coefficient model that is linear in *X*. The linearity assumption may still be too restricted as indicated in Figure 1. In this paper, we adopt a more general nonparametric model for the instrumental variables. The instrumental variables may be observed only in a subsample as in the ACTG 175 study. As in Huang and Wang (2001, 2006), we assume a functional measurement error model with no specification on the error distribution. We develop novel nonparametric correction methods under this general framework. The methods will have broader applications than those described by Huang and Wang (2006) due to the flexibility of the instrumental variable model.

The paper is organized as follows. In Section 2, we give the model definition. We develop a simple nonparametric correction estimator in Section 3 and propose an improved generalized methods of moments nonparametric correction estimator in Section 4. The asymptotic properties are derived with the proofs given in the Appendix. The performance of the estimators is assessed by simulations in Section 5 and illustrated by an application in Section 6. The paper concludes with a discussion in Section 7.

2. MODEL DEFINITION

Let T denote the survival time and C the censoring time. The observed survival data are $V = \min(T, C)$ and $\Delta = I(T \le C)$, where $I(\cdot)$ is the indicator function. Let X denote a vector of p covariates that can be measured with error and Z denote a vector of q accurately measured covariates. The hazard of failure depends on covariates X and Z through the proportional hazard model

$$\lambda(t)=\lambda_0(t)\exp(\beta_0^T X+\gamma_0^T Z),$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, and (β_0, γ_0) are the regression parameters. We assume that the survival time T is independent of the censoring time C given (X, Z).

Suppose that the true value of X is not observable. Only an error contaminated measurement W is available, which satisfies the classical measurement error model

$$W=X+e,$$

where e denotes the additive measurement error with $E(e) = 0$, and X and e are independent. In addition, measurements are available on an instrumental variable R in a subset of subjects such that

$$R = g(X, Z, \varepsilon), \quad (1)$$

where $g(\cdot)$ is an unknown function, and ε is a set of unspecified random variables that are independent of (T, C) given (X, Z) and independent of e. This includes as special cases the replicates $R = X + \varepsilon$, linear instrument $R = a_0 + a_1X + \varepsilon$, varying coefficient instrument $R = a_0(Z)+a_1(Z)+\varepsilon$ (Carroll et al., 2004), and nonparametric instrument $R = g_*(X, Z) + \varepsilon$, where $g_*(X, Z)$ is an unspecified function of (X, Z). The instrumental variable R may depend on both X and Z. It may also depend on other variables as included in ε, but R and (T, C) are independent given (X, Z). The dimension s of R should satisfy $s \geq p$ to ensure identifiability. For simplicity, we assume $s = p$. An extension to $s > p$ is discussed in Section 7. Assume that the errors e and ε are independent of (T, C, X, Z) and each other. Note that no other assumptions are placed on X, e, ε and the function $g(\cdot)$. Let $\eta = I(R$ is observed) be the indicator of whether the instrument variable is observed. Assume η is independent of $\{T, C, X, Z, e, \varepsilon\}$.

Suppose $\{T_i, C_i, V_i, \Delta_i, X_i, W_i, Z_i, e_i, \varepsilon_i, \eta_i\}$ are independent and identically distributed samples of $\{T, C, V, \Delta, X, W, Z, e, \varepsilon, \eta\}$ and the observed data set is $\{(V_i, \Delta_i, \eta_i, W_i, \eta_iR_i, Z_i) : i = 1, \dots, n\}$. For brevity of notations, we may drop the subscript i throughout the paper when there is no confusion. We focus on estimating the regression parameters $\theta_0=(\beta_0^T, \gamma_0^T)^T$.

3. SIMPLE NONPARAMETRIC CORRECTION

Huang and Wang (2000, 2006) proposed nonparametric correction estimation based on (V, Δ, W). The essential idea is to correct the naive estimating function such that the bias is removed. However, their approach requires replicated measurements on W or a linear instrument variable, and thus cannot be used directly in our case. Alternatively, using the instrumental variable R, we may develop a nonparametric correction estimator in the same spirit.

Let $\theta = (\beta^T, \gamma^T)^T$, $N_i(t) = I(V_i \leq t, \Delta_i = 1)$ be the counting process of failures, and $Y_i(t) = I(V_i \geq t)$ the at risk process. For any scalar, vector or matrix H_i, let $F_i(t, \theta; H, X) = Y_i(t)H_i \exp(\beta^TX_i + \gamma^TZ_i)$. Here H_i can be either fixed or random. Note that F_i also depends on (Z_i, V_i, Δ_i), which are dropped in the notation for simplicity. Let $\hat{G}(t, \theta; H, X) = n^{-1}\sum_{i=1}^{n} F_i(t, \theta; H, X)$ and $G(t, \theta; H, X) = E\{F_i(t, \theta; H, X)\}$. Note that $G(t, \theta; H, X)$ is a fixed function of t and θ.

The naive estimating function replaces the true covariates X by W in the partial likelihood function and can be written as

$$\hat{U}_{NR}(\theta;(W^T,Z^T)^T,W)=n^{-1}\sum_{i=1}^{n}\int_0^L\left\{(W_i^T,Z_i^T)^T-\frac{\hat{G}(t,\theta;(W^T,Z^T)^T,W)}{\hat{G}(t,\theta;1,W)}\right\}dN_i(t)$$

at a given time L. This estimating function is biased (Prentice, 1982), which is essentially due to the "bias" of the ratio term $\hat{G}(t, \theta; (W^T, Z^T)^T, W)/\hat{G}(t, \theta; 1, W)$ from $\hat{G}(t, \theta; (X^T, Z^T)^T, X)/\hat{G}(t, \theta; 1, X)$ when replacing X by W. When the measurement error is normal with known variance, Nakamura (1992) proposed a corrected score approach which added a correction term to compensate the bias. In the same spirit, Huang and Wang (2000) took an nonparametric correction when replications of W_i were available. The key idea of Huang and Wang (2000) was to substitute different replicates for W in the ratio term. Due to the independence of the errors in the replicates, the bias of the ratio term is corrected. The estimating function can be represented by $\hat{U}_{NR}(\theta; (\hat{U}_{NR}(\theta;(W_*^T,Z^T)^T,W), Z^T)^T, W)$ with W_* being another replicate of W and averaging over all possible combinations of replicates. We do not have a replicated W_*, but we may consider replacing W_* by R in $\hat{U}_{NR}(\theta; (\hat{U}_{NR}(\theta;(W_*^T,Z^T)^T,W), Z^T)^T, W)$.

As R is only observed on a subset of the subjects, we consider

$$\hat{U}_C(\theta)=n^{-1}\sum_{i=1}^{n}\int_0^L\eta_i\left\{(R_i^T,Z_i^T)^T-\frac{\hat{G}(t,\theta;\eta(R^T,Z^T)^T,W)}{\hat{G}(t,\theta;\eta,W)}\right\}dN_i(t)=0. \quad (2)$$

Note that $\hat{U}_C(\theta)$ converges to

$$U_0(\theta)=E\left[\int_0^L\left\{\eta(R^T,Z^T)^T-\frac{G(t,\theta;\eta(R^T,Z^T)^T,W)}{G(t,\theta;\eta,W)}\right\}dN(t)\right]. \quad (3)$$

When $R = X$, $U_0(\theta)$ is the limit of the standard partial likelihood estimating function. Let $\mathscr{F}_i(t) = \{N_i(u), Y_i(u), X_i, W_i, R_i, Z_i : u \le t\}$. By lemma 1 and the independence of η from (V, Δ, W, R, Z), with iterated expectations,

$$\begin{aligned}E\left\{\int_0^L\eta(R^T,Z^T)^TdN(t)\right\}&=E(\eta)E\left[E\left\{\int_0^L(R^T,Z^T)^TdN(t)|\mathscr{F}_i(t)\right\}\right]\\&=E(\eta)E\left\{\int_0^LG(t,\theta;(R^T,Z^T)^T,X)dt\right\}.\end{aligned}$$

In addition, we have $E\left\{\int_0^L\eta dN(t)\right\}=E(\eta)E\left\{\int_0^LG(t,\theta;1,X)dt\right\}$, $G(t, \theta; \eta, W) = E(\eta)E\{\exp(\beta^Te)\}G(t, \theta; 1, X)$,and

$$G(t,\theta;\eta(R^T,Z^T)^T,W)=E(\eta)E\{\exp\beta^Te)\}G(t,\theta;(R^T,Z^T)^T,X).$$

Hence it can be easily seen that $U_0(\theta) = 0$. Therefore $\hat{U}_C(\theta; (R^T, Z^T)^T, W)$ is asymptotically unbiased and (2) is a simple nonparametric correction equation.

Let

$$\Gamma(\theta_0)=\int_0^L \left[\frac{G\left(t,\theta_0;(R^T,Z^T)^T(W^T,Z^T),W\right)}{G(t,\theta_0;1,W)}-\frac{G\left(t,\theta_0;(R^T,Z^T)^T,W\right)G\left(t,\theta_0;(W^T,Z^T),W\right)}{G^2(t,\theta_0;1,W)}\right]dE\{N(t)\}. \quad (4)$$

We derive the asymptotic properties of the simple nonparametric correction estimator using empirical process theory.

Theorem 1

Under conditions A–E given in the Appendix, a solution $\tilde{\theta} = (\tilde{\beta}^T, \tilde{\gamma}^T)^T$ of (2) exists and converges to θ_0 almost surely. Further, $n^{1/2}(\tilde{\theta} - \theta_0)$ is asymptotically normal with mean zero and variance $V_C=\{\Gamma_\eta^{-1}(\theta_0)\}^T \, var\,\{\omega_{\eta i}(\theta_0)\}\,\Gamma_\eta^{-1}(\theta_0)$, where

$$\Gamma_\eta(\theta_0)=\int_0^L \left[\frac{G\left(t,\theta_0;\eta(R^T,Z^T)^T(W^T,Z^T),W\right)}{G(t,\theta_0;\eta,W)}-\frac{G\left(t,\theta_0;\eta(R^T,Z^T)^T,W\right)G\left(t,\theta_0;\eta(W^T,Z^T),W\right)}{G^2(t,\theta_0;\eta,W)}\right]dE\{\eta N(t)\} = E(\eta)\Gamma(\theta_0),$$

and

$$\omega_{\eta i}(\theta_0)=\eta_i\int_0^L \left((R_i^T,Z^T)^T-\frac{G\left(t,\theta_0;\eta(R^T,Z^T)^T,W\right)}{G(t,\theta_0;\eta,W)}\right)\times\left\{dN_i(t)-\frac{F_i(\theta;1,W)dE\{\eta N(t)\}}{G(t,\theta_0;\eta,W)}\right\}.$$

A consistent estimator of the variance can be obtained by substituting $\tilde{\theta}$ for θ_0 and the empirical means for the population means in the variance formula.

Remark

Condition E requires $\Gamma(\theta_0)$ to be nonsingular. It can be easily shown that $\Gamma(\theta_0) = 0$ when R is independent of X. Thus R and X should be dependent. But R and X may have a nonlinear association, for example, $R = X^2 + \varepsilon$ with X having a distribution symmetric around zero and ε independent of X, although the linear correlation is zero. This is due to the nonlinearity of the model, which is different from the linear instrumental model in econometrics (Stock and Watson, 2010, chapter 12).

To better understand what factors affect the variance of $\tilde{\theta}$, we expand V_C in Theorem 1 although it is not needed for estimation of V_C. With some algebra, it can be shown that $\Gamma(\theta_0) = \Gamma_*(\theta_0)$, where $\Gamma_*(\theta_0)$ is $\Gamma(\theta_0)$ with W replaced by X, that is,

$$\Gamma_*(\theta_0)=\int_0^L \left[\frac{G\left(t,\theta_0;(R^T,Z^T,X)^T(X^T,Z^T)\right)}{G(t,\theta_0;1,X)} - \frac{G\left(t,\theta_0;(R^T,Z^T)^T,X\right)G\left(t,\theta_0;(X^T,Z^T),X\right)}{G^2(\theta_0;1,X)} dE\{N(t)\}.\right.$$

Further,

$$V_C=E^{-1}(\eta)\,(V_I+V_A), \quad (5)$$

where

$$V_I=\{\Gamma_*^{-1}(\theta_0)\}^T S_1(\theta_0;(R^T,Z^T)^T,X)\Gamma_*^{-1}(\theta_0),$$
$$V_A=\{\Gamma_*^{-1}(\theta_0)\}^T S_2(\theta_0;(R^T,Z^T)^T,X)\Gamma_*^{-1}(\theta_0),$$

with

$$S_1(\theta_0;(R^T,Z^T)^T,X)=E\left\{\int_0^L \left((R^T,Z^T)^T-\frac{G\left(t,\theta_0;(R^T,Z^T)^T,X\right)}{G(T,\theta_0;1,X)}\right)\times\left[dN_i(t)-\frac{F_i(t,\theta;1,X)dE\{N(t)\}}{G(t,\theta_0;1,X)}\right]\right\}^{\otimes 2},$$
$$S_2(\theta_0;(R^T,Z^T)^T,X)=\left\{\frac{E\{\exp(2\beta^Te)\}}{E^2\{\exp(\beta^Te)\}}-1\right\}\times E\left[\int_0^L\left((R^T,Z^T)^T-\frac{G\{t,\theta_0;(R^T,Z^T)^T,X\}}{G(t,\theta_0;1,X)}\right)\frac{F_i(t,\theta;1,X)dE\{N(t)\}}{G(t,\theta_0;1,X)}\right]^{\otimes 2}.$$

It can be easily seen that V_I is the variance of $\tilde{\theta}$ when $var(e) = 0$, and V_A is a nonnegative definite matrix. Expression (5) indicates that the efficiency of $\tilde{\theta}$ improves with the increase of $\Pr(\eta = 1) = E(\eta)$ or the decrease of $E\{\exp(2\beta e)\}/E^2\{\exp(\beta^T e)\}$. When the error e is normal, $E\{\exp(2\beta^T e)\}/E^2\{\exp(\beta^T e)\} = \exp\{3\beta^T var(e)\beta/2\}$ is an increasing function of $var(e)$. In the special case that $R = g_*(X, Z) + \varepsilon$ with ε independent with (V, Δ, X, Z, e), it can be shown that V_c increases with the increase of $var(\varepsilon)$. Although the variance V_C may depend on the variance of e and other unknown quantities, estimation of V_C does not require estimating these quantities.

A drawback of the simple nonparametric approach is that it only uses the *calibration subsample* $\Omega_C = \{(V_i, \Delta_i, Z_i, W_i, R_i) : \eta_i = 1\}$ where both the error contaminated variable W_i and the instrumental variable R_i are observed. The information in the *non-calibration subsample* $\Omega_{\bar{C}} = \{(V_i, \Delta_i, Z_i, W_i) : \eta_i = 0\}$ with missing R_i is not used. When the calibration subsample is small compared to the sample size, it can be very inefficient. It is expected that the efficiency can be improved if we could use the *whole sample* $\Omega = \Omega_C \cap \Omega_{\bar{C}}$. This motivates us to develop an improved estimator.

4. GMM NONPARAMETRIC CORRECTION

Note that the nonparametric correction based on W_i only uses the whole sample (Song and Huang, 2005; Huang and Wang, 2006). The corresponding estimating equation can be written as

$$\hat{U}_F(\theta)=n^{-1}\sum_{i=1}^{n}\int_0^L\left((W_i^T, Z_i^T)^T+(c^T(\theta), 0_q^T)^T-\frac{\hat{G}(t,\theta;(W^T, Z^T)^T, W)}{\hat{G}(t,\theta;1,W)}\right)dN_i(t)=0, \quad (6)$$

where

$$c(\theta)=\frac{E\{e\exp(\beta^T e)\}}{E\{\exp(\beta^T e)\}}.$$

However, to estimate θ based on (6), the correction term $c(\theta)$ needs to be estimated by replicated measurements on W_i, which is not available in our case. Note that, if θ is known, we may estimate $c_0 = c(\theta_0)$ based on the first p-equations of (6). As we have already obtained an estimate $\tilde{\theta}$ based on (2), we may plug it in (6) and obtain an estimator of c_0 based on the calibration subsample Ω_C,

$$\hat{c}=\hat{c}(\tilde{\theta})=-\left[\int_0^L d\hat{E}\{\eta N(t)\}\right]^{-1}\int_0^L\left[d\hat{E}\{\eta W^T N(t)\}-\frac{\hat{G}(t,\tilde{\theta};\eta W, W)}{\hat{G}(t,\tilde{\theta};\eta, W)}d\hat{E}\{\eta N(t)\}\right],$$

where $\hat{E}$ is the operator for empirical mean such that $\hat{E}(a)=n^{-1}\sum_{i=1}^n a_i$.

To utilize the information on the whole sample, we propose an improved nonparametric correction estimator $\hat{\theta}(A)$ by minimizing the quadratic form

$$Q(\theta;\hat{c}, A)=\hat{U}^T(\theta;\hat{c})A\hat{U}(\theta;\hat{c}),$$

where A is a $(2p + q) \times (2p + q)$ nonzero semi-positive definite matrix and

$$\hat{U}(\theta;\hat{c})=n^{-1}\sum_{i=1}^{n}\begin{pmatrix}\int_0^L\eta_i\left\{R_i-\hat{G}(t,\theta;\eta R, W)/\hat{G}(t,\theta;\eta, W)\right\}dN_i(t)\\ \int_0^L\left\{W_i+\hat{c}-\hat{G}(t,\theta;W,W)/\hat{G}(t,\theta;1,W)\right\}dN_i(t)\\ \int_0^L\left\{Z_i-\hat{G}(t,\theta;Z,W)/\hat{G}(t,\theta;1,W)\right\}dN_i(t)\end{pmatrix}.$$

The $(2p + q)$ dimensional vector $\hat{U}(\theta; \hat{c})$ contains the estimating functions in (2) and (6), which include the information on the whole sample Ω. Since the number of estimating functions in $\hat{U}(\theta; \hat{c})$ is larger than the number of parameters $(p + q)$, there is generally no estimate for $\hat{U}(\theta; \hat{c}) = 0$. To derive an estimator, the quadratic form $Q(\theta; \hat{c}, A)$ is minimized instead. The derivation of the improved estimator has adopted similar techniques for the generalized methods of moments (GMM) (Hansen, 1982), and thus we call it the *GMM nonparametric correction* estimator. The GMM is a general methodology in econometrics literature (e.g. Cragg, 1983; Newey 1988; Newey and McFadden, 1994; Stock and Wright, 2000). It combines economic data with population moment conditions to produce estimators of parameters in statistical models. It is an extension of the method of moments to allow more moment conditions than the parameters to estimate. The GMM estimator is obtained by minimizing a quadratic form in the sample moments conditions.

The matrix A plays a role similar to weights for the estimating functions. The estimator will be different with a different choice of the matrix A. Our goal is to find an optimal matrix A_{opt} such that the estimator $\hat{\theta}(A_{opt})$ is most efficient among such estimators.

For this purpose, we first derive the asymptotic properties of $\hat{\theta}(A)$. Let I_s denote an s-dimensional identity matrix.

Theorem 2

Under conditions A–H, $\hat{\theta}(A)$ is a consistent estimator of θ_0. Further, $n^{1/2}(\hat{\theta}(A) - \theta_0)$ is asymptotically normal with mean zero and variance

$$V(A)=\{D^T(\theta_0)AD(\theta_0)\}^{-1} D^T(\theta_0)AB(\theta_0)AD(\theta_0)\{D^T(\theta_0)AD(\theta_0)\}^{-1},$$

where

$$D(\theta_0)=diag(E(\eta)I_p, I_pI_q)\int_0^L\{G(t,\theta_0;(R^T,W^T,Z^T)^T(W^T,Z^T),W) - \frac{G(t,\theta_0;(R^T,W^T,Z^T)^T,W)G(t,\theta_0;(W^T,Z^T),W)}{G(t,\theta_0;1,W)}\Bigg\} \frac{dE\{N(t)\}}{G(t,\theta_0;1,W)},$$

$$B(\theta_0)=var\{\varphi_i(\theta_0)\} \text{ with } \varphi_i(\theta_0)=\rho_i(\theta_0)+(0_p^T,\tau_i^T(\theta_0),0_q^T),\ \rho_i(\theta_0)=(\rho_{iR}^T(\theta_0),\rho_{iW}^T(\theta_0),\rho_{iZ}^T(\theta_0))^T,$$

$$\rho_{iR}(\theta_0)=\eta_i\int_0^L\left(R_i-\frac{G(t,\theta_0;\eta R,W)}{G(t,\theta_0;\eta,W)}\right)\times\left[dN_i(t)-\frac{F_i(t,\theta_0;\eta,W)}{G(t,\theta_0;\eta,W)}dE\{\eta N(t)\}\right],$$

$$\rho_{iW}(\theta_0)=\int_0^L\left\{W_i-\frac{G(t,\theta_0;W,W)}{G(t,\theta_0;1,W)}\right\}\times\left[dN_i(t)-\frac{F_i(t,\theta_0;1,W)dE\{N(t)\}}{G(t,\theta_0;1,W)}\right]+c_0\int dN_i(t),$$

$$\rho_{iZ}(\theta_0)=\int_0^L\left(Z_i-\frac{G(t,\theta_0;Z,W)}{G(t,\theta_0;1,W)}\right)\times\left[dN_i(t)-\frac{F_i(t,\theta_0;1,W)}{G(t,\theta_0;1,W)}dE\{\eta N(t)\}\right],$$

$$\tau_i^T(\theta_0)=E\left\{\int_0^L dE\{N_i(t)\}\right\}\{\xi_i(\theta_0)+\zeta_i(\theta_0)\},$$

$$\xi_i(\theta_0)=-\left[\int_0^L dE\{\eta N(t)\}\right]^{-1}\int_0^L\eta_i\left\{W_i-\frac{G(t,\theta_0;\eta W,W)}{G(t,\theta_0;\eta,W)}\right\}\times\left[dN_i(t)-\frac{F_i(t,\theta_0;\eta,W)}{G(t,\theta_0;\eta,W)}dE\{\eta N(t)\}\right]-\left[\int_0^L dE\{\eta N(t)\}\right]^{-1}c_0\left\{\int_0^L\eta_i dN_i(t)\right.$$

$$\zeta_i(\theta_0)=-\left[\int_0^L dE\{\eta N(t)\}\right]^{-1}\int_0^L\left[G\frac{(t,\theta_0;\eta W(1,W^T),W)}{G(t,\theta_0,\eta,W)}-\frac{G(t,\theta_0,\eta W,W)G(t,\theta_0,\eta W^T,W)}{G^2(t,\theta_0,\eta,W)}\right]dE\{\eta N(t)\}\times\{E(\eta)\Gamma(\theta_0)\}^{-1}\omega_i(\theta_0).$$

A consistent estimator of the variance can be obtained by substituting $\hat{\theta}(A)$ for θ_0 and the empirical means for the population means in the variance formula.

To find the optimal matrix A_{opt}, we minimize the variance $V(A)$ of the estimator $\hat{\theta}(A)$. This can be achieved by simple matrix algebra by analogy to the generalized methods of moments (Newey and McFadden, 1994) and the result is given in the following theorem.

Theorem 3

Under conditions A–H, the most efficient estimator of $\hat{\theta}(A)$ is achieved at $A_{opt} = B^{-1}(\theta_0)$ with the variance $V(A_{opt}) = \{D^T(\theta_0)B^{-1}D(\theta_0)\}^{-1}$.

The GMM estimator $\hat{\theta}(\hat{A}_{opt})$ is generally more efficient than the simple estimator $\tilde{\theta}$. This can be easily seen when there is no Z is the model. Specifically, without Z, noting that $\hat{U}(\theta;\hat{c})=(\hat{U}_c^T(\theta;(R^T,Z^T)^T),\hat{U}_F(\theta;\hat{c}))^T$, the simple nonparametric correction estimator minimizes $Q(\theta, \hat{c}, A_c)$ with $A_c = diag(I_p, 0_{q\times q})$, where $0_{q\times q}$ is a $q \times q$ zero matrix. In practice, A_{opt} can be approximated by $\hat{A}_{opt} = \tilde{B}^{-1}$ with $\tilde{B} = n^{-1}\sum\{\hat{\phi}_i(\tilde{\theta}) - \hat{E}\hat{\phi}(\tilde{\theta})\}^{\otimes 2}$, where $\hat{\phi}_i(\tilde{\theta})$ is obtained by substituting the unknown quantities in $\phi_i(\tilde{\theta})$ by their empirical estimates. The variance of $\hat{\theta}(\hat{A}_{opt})$ can be estimated by $\{\hat{D}^T\hat{A}_{opt}\hat{D}\}^{-1}\hat{D}^T\hat{A}_{opt}\hat{B}\hat{A}_{opt}\hat{D}\{\hat{D}^T\hat{A}_{opt}\hat{D}\}^{-1}$, where $\hat{D} = -\partial\hat{U}(\hat{\theta}(\hat{A}_{opt}))/\partial\theta^T$, and $\hat{B} = \sum\{\hat{\phi}_i(\hat{\theta}(\hat{A}_{opt})) - \hat{E}\hat{\phi}(\hat{\theta}(\hat{A}_{opt}))\}^{\otimes 2}$.

Remark

There could be a few variations of the above estimator by varying the data set used in estimating the correction term c_0 and the data set in the objective function $Q(\cdot)$ corresponding to the covariates W. Let Θ_c denote the former and Θ_Q the latter. Both data sets could be elements of $\{\Omega, \Omega_C, \Omega_{\bar{C}}\}$ as long as $\Theta_c \neq \Theta_Q$. Our numerical studies indicate that the performance of the GMM estimator seems similar for various choices of Θ_c and Θ_Q except in some extreme cases, such as a very small sample calibration subsample or non-calibration subsample. We use $\Theta_c = \Theta_C$ and $\Theta_Q = \Omega$ in our illustration.

5. SIMULATION STUDIES

Simulation studies were conducted to evaluate the performance of the estimators. First, we considered the case of a single covariate X, which was generated from a standard normal distribution. The instrumental variable was set as $R = 0.5X^2 + 2X + 1 + 0.5\varepsilon_1 + X\varepsilon_1 + \varepsilon_2$, where ε_1 was generated from a standard normal distribution correlated with X with correlation −0.3 which may denote a variable that was not in the proportional hazard model, and ε_2 from a normal distribution independent of X with mean 0 and variance 0.4 which may denote independent noise. The error e was generated from a normal or a skewed bimodal mixture of two normals as described in Davidian and Gallant (1993, mixing proportion $p = 0.3$ and distance between the means equal to sep = 2 times standard deviation) with mean 0 and variance $\sigma^2 = 0.1$ or 0.2. The true Cox model coefficient was taken to be $\beta_0 = -1$. The baseline hazard $\lambda_0(t) = exp\{-2\}t^{-0.5}$. The censoring time was generated from a uniform distribution on [0,40], leading to a censoring rate of about 37%. The proportion of calibration subsample $\Pr(\eta = 1)$ was set to 0.3, 0.5 or 0.7.

We carried out simulations for $n = 500$ and 2000. In each scenario, 1000 Monte Carlo data sets were simulated. For each data set, we fitted the model using (i) the “ideal” approach, in which the true values of X were used; (ii) the naive approach, in which W substituted for X in the partial likelihood estimating equation; (iii) the simple nonparametric correction estimator $\tilde{\theta}$; (iv) the GMM nonparametric correction estimator $\hat{\theta}(\hat{A}_{opt})$. For each estimator, the 95% Wald confidence interval was constructed.

The results are shown in tables 1 and 2 respectively for the normal and the mixture normal error models. The naive estimator is biased with a coverage probability well below the nominal level. The performance gets worse with the sample size growing or the error variance increasing. The nonparametric correction estimators have negligible bias close to the unachievable "ideal" estimator and the coverage probabilities are close to the nominal level. Their performance improves when the sample size increases or the error variance decreases. The GMM estimator is more efficient than the simple estimator, especially when $\Pr(\eta = 1)$ is small. For either correction approach, the standard deviations are close to the standard errors, and the efficiency improves with the increase of the proportion of calibration subsample or the decrease of the magnitude of measurement error.

Next we added a covariate $Z = \varepsilon_1$ to the proportional hazards model with $\gamma_0 = -1$. The censoring rate was 38%. The results for the normal error model with $\sigma^2 = 0.1$ and $var(\varepsilon_2) = 0.4$ are shown in Table 3. We observe similar results for estimation of β_0 as above. The estimation of γ_0 shows similar pattern. Note that the naive estimator of γ_0 also shows some bias and the coverage probability is only 83% for $n = 500$ and 52% for $n = 2000$. This indicates that estimation of the coefficient of the error free covariate Z can be affected by the measurement error on X as well.

The relationship between R and X may impact the performance of the estimators as well. We conducted simulations in the case of one covariate with normal error as described above with different instrumental variables. We considered two cases when R and X were non-linearly associated with zero linear correlation, $R = X^2 + \varepsilon$ and $R = X^4 + \varepsilon$, and compared them to the case when $R = X + \varepsilon$, where ε was normal and independent of X with mean 0 and variance 0.2. The results for $\sigma^2 = 0.1$, $\Pr(\eta = 1) = 0.5$ and $n = 2000$ are shown in Table 4. The nonparametric corrections estimators still work when $R = X^2 + \varepsilon$ or $X^4 + \varepsilon$, but the standard errors are larger than when $W = X + \varepsilon$. The performance is better when $R = X^2 + \varepsilon$ than when $R = X^4 + \varepsilon$.

We also conducted simulations to assess the sensitivity of nonparametric correction approaches to the assumption that R is independent of (T, C) given X and Z. In the single covariate model described above, the proportional hazards model can be rewritten as $log(T) = a + 2X + 2\varepsilon_*$, where a is a constant and ε_* is an extreme-value-distributed random variable with variance $\pi^2/6$ and independent of X. We replaced v the instrumental variable by $R = X + b\sqrt{6}\varepsilon_*/(10\pi)$ so that R and T are correlated given X if $b \neq 0$. We show the results for $b = 0, 0.5, 1, 2$ with normal error, $\sigma^2 = 0.1$, $\Pr(\eta = 1) = 0.5$ and $n = 500$ and 1000 in Table 5. The nonparametric correction estimators are not consistent in this case. Their performance tends to get worse with increasing b, which represents an increasing association between R and T given X. The bias may be large if violation of conditional independence is not small.

6. APPLICATION

We applied the approaches to the AIDS Clinical Trial Group (ACTG) 175 study. Our aim was to evaluate the effect of treatments for the time to AIDS or death adjusted for baseline CD4 counts. The primary analysis found ziduvudine alone to be inferior to the other three therapies; thus, further investigations focused on two treatment groups, zidovudine alone and the combination of the other three.

This dataset has been analyzed previously. By definition, baseline CD4 counts should be true CD4 counts at randomization. However, CD4 counts were only measured for less than

50% of the patients on randomization day. Huang and Wang (2000) assumed the CD4 measurements within three weeks of randomization were replicates of the underlying baseline CD4 counts. As the underlying CD4 counts may change over time during the three weeks period, these CD4 measurements may not be simple replicates of the baseline CD4 counts. We took an alternative strategy here. Assuming the CD4 counts is relatively stable within a short period, say one week, the closest measurement W within one week before randomization was taken as the baseline CD4 measurement. The closest measurement R between one to three weeks before randomization was used as an instrumental variable. Among the 2174 subjects with baseline CD4 measurement, the instrumental variable was observed among 989 patients. The median follow up time was 33 months. A total of 275 events was observed.

A proportional hazards model was adopted with two covariates, the true baseline $X = log(\text{CD4})$ and the treatment indicator $Z = I(\text{treatment} \neq \text{ziduvudine})$. The logarithm transformation was applied to the CD4 counts to achieve approximate constant variance. The same transformation was applied on the observed CD4 counts W and R. We first examine whether R is an appropriate instrumental variable. It is reasonable to assume that R is independent of the measurement error at baseline. Under this assumption, Figure 1 indicates that R is correlated with X. To be an instrumental variable, R needs to be independent of the time to AIDS or death given X and Z. This assumption seems to be appropriate based on our understanding of CD4 counts and AIDS risk, but cannot be tested from the data (Stock and Watson, 2010, chapter 12). Note that the assumption that R is a instrumental variable is weaker than R and W are replicates.

We estimated the regression coefficients using the naive, simple and GMM non-parametric correction approaches. The results are shown in Table 4. Both baseline CD4 and treatment are significant. The nonparametric correction estimates show stronger effects than the naive estimates, and the GMM estimates have smaller estimated standard errors than the simple estimates.

7. DISCUSSION

We have proposed nonparametric correction estimators for the proportional hazards model with error-contaminated covariates. The estimators are useful when no replicated observations are available on the error-prong covariates while observations available on instrumental variables.

For simplicity, we only consider the case when the dimension of the instrumental variables s equals the dimension p of the error-prone covariates. In the case of $s > p$, $\tilde{\theta}$ may be obtained by minimizing the quadratic form $\hat{U}_C^T A_C U_C$ and the optimal AC can be obtained by analogy to A_{opt}. The GMM estimator $\hat{\theta};$ can then be derived similarly as in section 4.

The function $g(\cdot)$ and the variables ε are unspecified in (1), this allows great flexibility in adopting instrumental variables. However, the format of $g(\cdot)$ may affect the efficiency of the simple and the GMM nonparametric correction estimators. The instrumental variables need not be linearly correlated with X, but cannot be independent of X. The proposed methods may break down if the instrumental variables are only weakly related with the underlying true covariates.

Our simulation studies reveal that the performance of the approaches depends on the magnitude of the measurement error, the sample size and the relationship between the error contaminated variables and the instrumental variables. When the measurement error is large, the methods might not work properly for small sample sizes with the possibility of nonconvergence and outlier estimates. This is a common issue for parametric/nonparametric correction approaches (Song and Huang, 2005). A possible improvement for the finite sample performance is to use the refined non-parametric correction technique (Song and Huang, 2005). The bootstrap confidence interval may work better when the measurement error is large (Huang and Wang, 2001).

Acknowledgments

This research was partially supported by NIH grants R01ES017030, HL121347 (Wang and Song), CA53996 (Wang) and CA152460 (Song), NSF grant DMS-1106816 (Song), and a travel award from the Mathematics Research Promotion Center of the National Science Council of Taiwan (Wang).

References

Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. The Annals of Statistics. 1982; 10:1100–1120.

Carroll RJ, Ruppert D, Crainiceanu CM, Tosteson TD, Karagas RM. Nonlinear and Nonparametric Regression and Instrumental Variables. Journal of the American Statistical Association. 2004; 99:736–750.

Carroll, RJ.; Ruppert, D.; Stefanski, LA.; Crainiceanu, CM. Meaurement Error in Nonlinear Models. New York: Chapman & Hall/CRC; 2006.

Cragg JG. More Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form. Econometrica. 1983; 51:751–763.

Dafni UG, Tsiatis AA. Evaluating Surrogate Markers of Clinical Outcome Measured With Error. Biometrics. 1998; 54:1445–1462. [PubMed: 9883544]

Davidian M, Gallant AR. The Nonlinear Mixed Effects Model With a Smooth Random Effects Density. Biometrika. 1993; 80:475–488.

Ding J, Wang JL. Modeling Longitudinal Data With Nonparametric Multiplicative Random Effects Jointly With Survival Data. Biometrics. 2008; 64:546–556. [PubMed: 17888040]

Hammer SM, Katezstein DA, Hughes MD, Gundaker H, Schooley RT, Haubrich RH, Henry WK, Lederman MM, Phair JP, Niu M, Hirsch MS, Merigan TC. for the AIDS Clinical Trials Group Study 175 Study Team . A Trial Comparing Nucleoside Monotherapy With Combination Therapy in HIV-infected Adults With CD4 Cell Counts From 200 to 500 Per Cubic Millimeter. New England Journal of Medicine. 1996; 335:1081–1089. [PubMed: 8813038]

Hansen LP. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica. 1982; 50:1029–1054.

Hu C, Lin DY. Cox Regression With Covariate Measurement Error. Scandinavian Journal of Statistics. 2002; 29:637–655.

Huang Y, Wang CY. Cox Regression With Accurate Covariates Unascertainable: A Nonparametric Correction Approach. Journal of the American Statistical Association. 2000; 95:1209–1219.

Huang Y, Wang CY. Consistent Functional Methods for Logistic Regression With Errors in Covariates. Journal of the American Statistical Association. 2001; 96:1469–1482.

Huang Y, Wang CY. Errors-In-Covariates Effect on Estimating Functions: Additivity in Limit and Nonparametric Correction. Statisitica Sinica. 2006; 16:861–881.

Li Y, Ryan L. Survival Analysis With Heterogeneous Covariate Measurement Error. Journal of the American Statistical Association. 2004; 99:724–735.

Liao X, Zucker DM, Li Y, Speigelman D. Survival Analysis with Error-Prone Time-Varying Covariates: A Risk Set Calibration Approach. Biometrics. 2011; 67:50–58. [PubMed: 20486928]

Nakamura T. Proportional Hazards Model With Covariates Subject to Measurement Error. Biometrics. 1992; 48:829–838. [PubMed: 1420844]

Newey W. Adaptive Estimation of Regression Models via Moment Restrictions. Journal of Econometrics. 1988; 38:301–339.

Newey, W.; McFadden, D. Handbook of Econometrics. Vol. 36. Elsevier Science; 1994. Large Sample Estimation and Hypothesis Testing.

Prentice R. Covariate Measurement Errors and Parameter Estimates in a Failure Time Regression Model. Biometrika. 1982; 69:331–42.

Song X, Davidian M, Tsiatia AA. A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-to-Event Data. Biometrics. 2002a; 58:742–753. [PubMed: 12495128]

Song X, Davidian M, Tsiatia AA. An Estimator for the Proportional Hazards Model With Multiple Longitudinal Covariates Measured With Error. Biostatistics. 2002b; 3:511–528. [PubMed: 12933595]

Song X, Huang Y. On Corrected Score Approach for Proportional Hazards Model With Covariate Measurement Error. Biometrics. 2005; 61:702–714. [PubMed: 16135021]

Song X, Wang CY. Semiparametric Approaches for Joint Modeling of Longitudinal and Survival Data with Time Varying Coefficients. Statistica Sinica. 2008; 27:3178–3190.

Stock, JH.; Watson, WW. Introduction to Econometrics. New-York: Addison-Wesley; 2010.

Stock JH, Wright JH. GMM With Weak Identification. Econometrica. 2000; 68:1055–1096.

Tapsoba JD, Lee SM, Wang CY. Joint Modeling of Survival Time and Longitudinal Data With Subject-Specific Change Points in The Covariates. Statistics in Medicine. 2011; 30:232–249. [PubMed: 21213341]

Tsiatis AA, Davidian M. A Semiparametric Estimator for the Proportional Hazards Model With Longitudinal Covariates Measured With Error. Biometrika. 2001; 88:447–458.

Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the Relationship of Survival to Longitudinal Data Measured With Error: Applications to Survival and CD4 Counts in Patients With AIDS. Journal of the American Statistical Association. 1995; 90:27–37.

Xu J, Zeger SL. The Evaluation of Multiple Surrogate Endpoints. Biometrics. 2001; 57:81–87. [PubMed: 11252622]

Wang CY. Corrected Score Estimator for Joint Modeling of Longitudinal and Failure Time Data. Statistica Sinica. 2006; 16:235–353.

Wang CY. Non-parametric Maximum Likelihood Estimation for Cox Regression With Subject-Specific Measurement Error. Scandinavian Journal of Statistics. 2008; 35:613–628.

Wang CY, Hsu L, Feng ZD, Prentice RL. Regression Calibration in Failure Time Regression. Biometrics. 1997; 53:131–145. [PubMed: 9147589]

Wulfsohn MS, Tsiatis AA. A Joint Model for Survival and Longitudinal Data Measured With Error. Biometrics. 1997; 53:330–339. [PubMed: 9147598]

APPENDIX A: PROOFS

Regularity Conditions

We assume the following mild regularity conditions.

A. $\lambda_0(u)$ is continuous in $[0, L]$.

B. $\Pr(V \geq L) > 0$.

C. $E(X^T X) < \infty$, $E\{R^T R < \infty$, $E(Z^T Z) < \infty$, $E(e^T e) < \infty$.

For a compact neighborhood $\mathcal{N}(\theta_0)$ of θ_0,

$E[\sup_{\theta\in\mathcal{N}(\theta_0)} X^T X \exp\{2(\beta^T X + \gamma^T Z)\}] < \infty$,

$E[\sup_{\theta\in\mathcal{N}(\theta_0)} Z^T Z \exp\{2(\beta^T X + \gamma^T Z)\}] < \infty$,

$E[\sup_{\theta\in\mathcal{N}(\theta_0)} R^T R \exp\{2(\beta^T X + \gamma^T Z)\}] < \infty$,

$E[\sup_{\theta\in\mathcal{N}(\theta_0)} \exp\{2(\beta^T e)\}] < \infty$.

D. $E(\eta) > 0$.

E. The matrix $\Gamma(\theta_0)$ defined in (4) is nonsingular.

F. $\Pr(T < C, T < L) > 0$.

G. The matrix A is positive definite.

Lemma 1

Suppose H_i is a predictable random vector with respect to the filtration $\mathcal{F}_i(t) = \{N_i(u), Y_i(u), X_i, W_i, R_i, Z_i: u \le t\}$. If $E[H_i^T H_i] < \infty$, then

$$E'[H_i N_i(t)] = \lambda_0(t) G(t, \theta_0; H, X).$$

Proof—Note that $M_i(t) = N_i(t) - \int_0^t \lambda_0(u) Y_i(u) \exp(\beta_0^T X + \gamma_0^T Z) dt$ is a martingale with respect to the filtration $\mathcal{F}_i(t)$ as $N_i(u)$ is independent of (W_i, R_i) given (X_i, Z_i). By iterated expectations and the predictability of H_i,

$$E\{H_i M_i(t)\} = E[E\{H_i M_i(t) | \mathscr{F}_i(t-)\}] = E[H_i E\{M_i(t) | \mathscr{F}_i(t-)\}] = 0.$$

Substituting M_i by $N_i(t) - \int_0^t \lambda_0(u) Y_i(u) \exp(\beta_0^T X + \gamma_0^T Z) \mathrm{d}t$ in the left side of the above equation, we have

$$E\{H_i N_i(t)\} - E\left\{H_i \int_0^t \lambda_0(u) Y_i(u) \exp(\beta_0^T X_i + \gamma_0^T Z_i) du\right\} = 0.$$

Taking derivative with respect to t, under conditions A and C together with $E[H_i^T H_i] < \infty$, we obtain

$$E'\{H_i N_i(t)\} - E\{H_i \lambda_0(t) Y_i(t) \exp(\beta_0^T X_i + \gamma_0^T Z_i)\} = 0.$$

This completes the proof.

Proof for Theorem 1: First consider the consistency. Conditions B–D ensure $G(t, \theta; \eta)$ and $G(t, \theta; 1)$ are bounded away from zero in $\mathcal{N}(\theta_0)$. Note that $\hat{U}_C(\theta)$ can be rewritten as

$$\hat{U}_C=(\theta)=\int_0^L\left\{d\hat{E}\left\{(R^T,Z^T)^T\eta_iN_i(t)\right\}-\frac{\hat{G}(t,\theta;\eta(R^TZ^T)^T,W)}{\hat{G}(t,\theta;\eta,W)}\right\}d\hat{E}\left\{\eta N(t)\right\}. \quad (7)$$

Follow the extended strong law of large numbers as given in Appendix III of Andersen and Gill (1982), under condition *C*, the four empirical processes in (7) converge almost surely (a.s.) to their limits uniformly for $t\in(0,L)$ and $\theta\in\mathcal{N}(\theta_0)$. By the chain law, $\hat{U}_C(\theta)$ converges uniformly a.s. for $\theta\in\mathcal{N}(\theta_0)$ to

$$U_C(\theta)=\int_0^L\left\{dE\left\{(R_i^T,Z_i^T)^T\eta_iN_i(t)\right\}-\frac{G(t,\theta;\eta(R^T,Z^T)^T,W)}{G(t,\theta;\eta,W)}\right\}dE\left\{\eta N(t)\right\}$$
$$=\int_0^L\left\{dE\left\{(R^T,Z^T)^T\eta N(t)\right\}-\frac{G(t,\theta;(R^T,Z^T)^T,X)}{G(t,\theta;1,X)}\right\}dE\left\{\eta N(t)\right\}.$$

By Lemma 1 and the independence of η from (V,Δ,X,Z), we have

$$dE\left\{\eta N(t)\right\}=\lambda_0(t)E(\eta)G(t,\theta_0;1,X)dt,$$
$$dE\left\{(R^T,Z^T)^T\eta N(t)\right\}=\lambda_0(t)E(\eta)G(t,\theta_0;(R^T,Z^T)^T,X)dt.$$

It follows that $U_C(\theta_0)=0$. Similarly it can be shown that $\partial\hat{U}_C(\theta)/\partial\theta$ converges uniformly a.s. to $\Gamma_\eta(\theta)=E(\eta)\Gamma(\theta)$ for $\theta\in\mathcal{N}(\theta_0)$. Under Condition E, θ_0 is the unique zero crossing for $U_C(\theta)$ in a neighborhood of θ_0. The consistency of $\tilde{\theta}$ then follows.

Next, we show the asymptotic normality. By a Taylor expansion of $\hat{U}_C(\tilde{\theta})$ at θ_0,

$$0=\hat{U}_C(\tilde{\theta})=\hat{U}_C(\theta_0)+\frac{\partial}{\partial\theta^T}\hat{U}_C(\tilde{\theta}^*)(\tilde{\theta}-\theta_0),$$

where $\tilde{\theta}^*$ lies between θ_0 and $\tilde{\theta}$ Thus

$$n^{1/2}(\tilde{\theta}-\theta_0)=\left\{\frac{\partial}{\partial\theta^T}\hat{U}_C(\tilde{\theta}^*)\right\}^{-1}n^{1/2}\hat{U}_C(\theta_0).$$

With a functional Taylor expansion and straight algebra,

$$n^{1/2}\hat{U}_C(0)=n^{-1/2}\sum_{i=1}^n\eta_i\int_0^L\left((R_i^T,Z_i)^T-\frac{G(t,\theta_0;\eta(R^T,Z^T)^T,W)}{G(t,\theta_0;\eta,W)}\right)dN_i(t)$$
$$-\int_0^L\left(\frac{F_i(t,\theta_0;\eta(R^T,Z^T)^T,W)}{G(t,\theta_0;\eta,W)}-\frac{F_i\{\theta_0;\eta,W\}G(t,\theta_0;\eta(R^T,Z^T)^T,W)}{G^2(t,\theta_0;\eta,W)}\right)\times dE\eta N(t)+o_p(1)=n^{-1/2}\sum_{i=1}^n\omega_i(\theta_0)+o_p(1). \quad (8)$$

This, together with the uniform convergence of $\partial\hat{U}_C(\theta;(R^T,Z^T)^T)/\partial\theta^T$, establishes the asymptotic normality. One can then show the consistency of the variance estimator with similar arguments.

Proof of Theorem 2: First, we consider the asymptotic properties of the estimator $\hat{c}$. Under condition F, $\int_0^L dE\{N_i(t)\}>0$. By similar arguments as for the consistency in Theorem 1, we have

$$\hat{c} \xrightarrow{a.s.} -\left[\int_0^L dE\{\eta N(t)\}\right]^{-1}\int_0^L \left[dE\{\eta W^T N(t)\}-\frac{G(t,\theta_0;\eta W,W)}{G(t,\theta_0;\eta,W)}dE\{\eta N(t)\}\right]. \quad (9)$$

With some simple algebra, it can be shown that

$$\frac{G(t,\theta_0;\eta W,W)}{G(t,\theta_0;\eta,W)}=\frac{G(t,\theta_0;X,X)}{G(t,\theta_0;1,X)}+c_0.$$

By Lemma 1, we have $dE\{\eta N(t)\} = \lambda_0(t)E(\eta)G(t, \theta_0; 1,X)$ and $dE\{\eta WN(t)\} = dE[\eta XN(t)] = \lambda_0(t)E(\eta)G(t, \theta_0;X,X)$. Thus the right side of (9) equals c_0. With a functional Taylor expansion and some algebra, it can be shown that

$$n^{1/2}\{\hat{c}(\theta_0)-c_0\})=n^{-1/2}\sum_{i=1}^{n}\xi_i(\theta_0)+o_p(1). \quad (10)$$

Applying a Taylor expansion at θ_0, together with (8), we have

$$n^{1/2}\left\{\hat{c}(\tilde{\theta})-\hat{c}(\theta_0)\right\}=n^{-1/2}\sum_{i=1}^{n}\zeta_i(\theta_0)+o_p(1). \quad (11)$$

A combination of (10) and (11) gives

$$n^{1/2}\{\hat{c}-c_0\}=n^{-1/2}\sum_{i=1}^{n}\{\xi_i(\theta_0)+\zeta_i(\theta_0)\}+o_p(1). \quad (12)$$

Now we consider the asymptotic properties of $\hat{\theta}(A)$. By the consistency of $\hat{c}$ and empirical process theory, $\hat{U}(\theta; \hat{c})$ converges uniformly a.s. to

$$U(\theta)=\int_0^L\left\{\begin{pmatrix} dE\{\eta_i R_i N_i t)\} \\ dE\{W_i N_i(t)\}+c_0 dE\{N_i(t)\} \\ Z_i\end{pmatrix}+\begin{pmatrix} G^T(t,\theta;R,W) \\ G^T(t,\theta;W,W) \\ G^T(t,\theta;W,W)\end{pmatrix}\{G(t,\theta;1,W)\}^{-1}\right\}\times dE\{\eta N(t)\}.$$

Note that $U(\theta_0) = 0$. Under condition C, θ_0 is the unique solution to $U_{(p+1;p+2q)}(\theta) = 0$, where $U_{(p+1;p+2q)}(\theta)$ denote the p+1 to p+2q elements of $U(\theta)$ (Huang and Wang, 2000). Thus θ_0 is the unique solution to $U(\theta) = 0$ and hence the unique minimum of $U^T AU$. The consistency of $\hat{\theta}(A)$ then follows.

Next we consider the asymptotic normality. Note that $\hat{U}(\theta; c)$ is linear in c, and

$$\sqrt{n}\hat{U}(\theta;\hat{c})=\sqrt{n}\hat{U}(\theta;c_0)+\frac{\partial\hat{U}(\theta;c)}{\partial c^T}\sqrt{n}(\hat{c}-c_0), \quad (13)$$

where $\partial\hat{U}^T(\theta;c)/\partial c=(0_{p\times p}, \partial\hat{U}^T_{(p+1,2p)}(\theta;c)/\partial c, 0_{p\times p})$,

$$\partial\hat{U}^T_{(p+1,2p)}(\theta;c)/\partial c=n^{-1}\sum\int_0^L dN_i(t)I_p \xrightarrow{a.s.} \left[\int_0^L dE\{N(t)\}\right] I_p,$$

and $0_{p\times p}$ is a $p \times p$ zero matrix. It can be shown by a functional Taylor expansion that

$$\sqrt{n}\hat{U}(\theta;c_0)=n^{-1/2}\sum_{i=1}^{n}\rho_i(\theta)+o_p(1).$$

Substituting this and (12) into (13), we have

$$\sqrt{n}\hat{U}(\theta;\hat{c})=n^{-1/2}\sum_{i=1}^{n}\varphi_i(\theta)+o_p(1).$$

Since $\hat{\theta}(A)$ is the minimum of $Q(\theta; \hat{c}, A)$,

$$0=\frac{\partial Q(\hat{\theta}(A);\hat{c},A)}{\partial\theta}=2\frac{\partial\hat{U}^T(\hat{\theta}(A);\hat{c})}{\partial\theta}A\hat{U}^T(\hat{\theta}(A);\hat{c}).$$

By a Taylor expansion on $\hat{U}^T(\hat{\theta}(A); \hat{c})$ at θ_0, we have

$$0=2\frac{\partial\hat{U}^T(\hat{\theta}(A);\hat{c})}{\partial\theta}A\hat{U}(\theta_0,\hat{c})+2\frac{\partial\hat{U}^T(\hat{\theta}(A);\hat{c})}{\partial\theta}A\frac{\partial\hat{U}(\hat{\theta}^*(A);\hat{c})}{\partial\theta^T}(\hat{\theta}(A)-\theta_0),$$

where $\hat{\theta}^*(A)$ lies between $\hat{\theta}(A)$ and θ_0. Thus

$$\sqrt{n}(\hat{\theta}(A)-\theta_0)=-\left(\frac{\partial\hat{U}^T(\hat{\theta}(A);\hat{c})}{\partial\theta}A\frac{\partial\hat{U}(\hat{\theta}^*(A);\hat{c})}{\partial\theta^T}\right)^{-1}\frac{\partial\hat{U}^T(\hat{\theta}(A);\hat{c})}{\partial\theta}A\hat{U}(\theta_0,\hat{c}).$$

It can be shown that $-\partial\hat{U}(\theta; \hat{c})/\partial\alpha^T$ converges uniformly a.s. to $D(\theta)$. Therefore,

$$\sqrt{n}(\hat{\theta}(A)-\theta_0)=\{D^T(\theta_0)AD(\theta_0)\}^{-1}D^T(\theta_0)An^{-1/2}\sum_{i=1}^{n}\varphi_i(\theta_0)+o_p(1).$$

where $D^T(\theta_0)AD(\theta_0)$ is positive definite under conditions *D*, *E* and *G*. The asymptotic normality follows from the central limit theorem and the Slutsky Theorem.

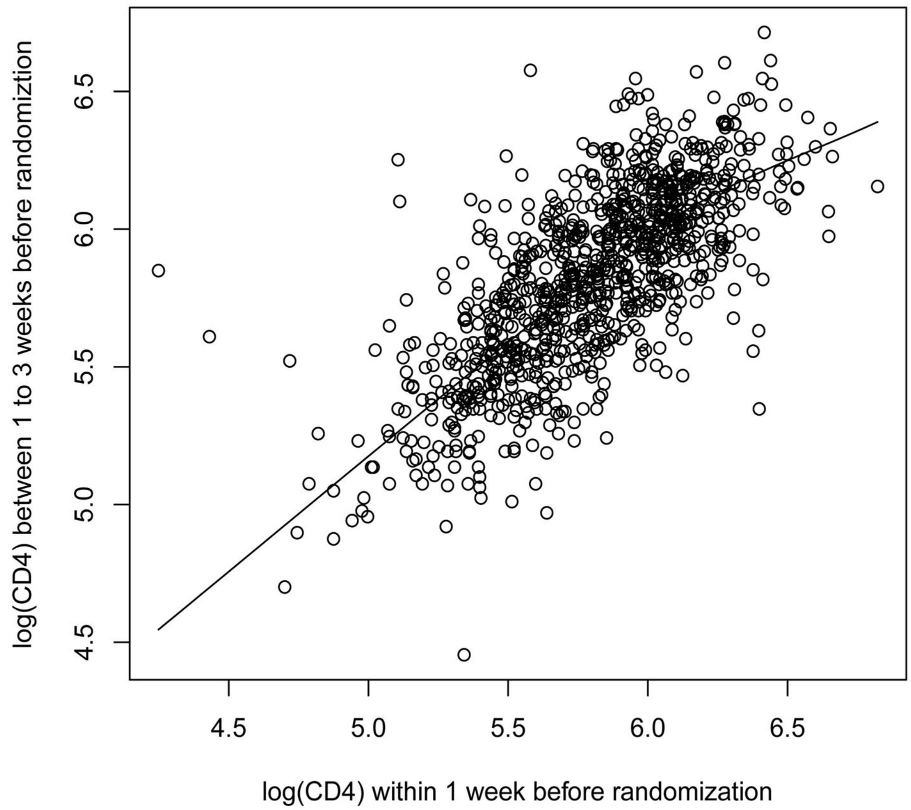

Figure 1.
Scatter plot of log(CD4) within one to three weeks versus one week before randomization. The curve was obtained by Loess smooth.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Simulation Results in the case of a single covariate contaminated with normal error.

			***n* = 500**				***n* = 2000**			
	Pr(η = 1)		**Est**	**SD**	**SE**	**CP**	**Est**	**SD**	**SE**	**CP**
		Ideal	−1.001	0.070	0.069	0.947	−0.999	0.035	0.035	0.948
$\sigma^2 = 0.1$		Naive	−0.874	0.065	0.064	0.476	−0.871	0.033	0.032	0.034
	0.3	SNC	−1.014	0.178	0.165	0.930	−1.005	0.084	0.080	0.936
		INC	−1.005	0.142	0.127	0.931	−1.003	0.068	0.063	0.947
	0.5	SNC	−1.012	0.136	0.126	0.935	−1.004	0.065	0.062	0.947
		INC	−1.007	0.118	0.107	0.930	−1.002	0.057	0.053	0.932
	0.7	SNC	−1.008	0.111	0.107	0.939	−1.001	0.056	0.053	0.935
		INC	−1.004	0.104	0.096	0.928	−0.999	0.051	0.048	0.943
$\sigma^2 = 0.2$		Naive	−0.777	0.062	0.060	0.054	−0.773	0.031	0.030	0.000
	0.3	SNC	−1.021	0.193	0.171	0.931	−1.007	0.090	0.081	0.923
		INC	−1.004	0.160	0.137	0.927	−1.002	0.074	0.066	0.931
	0.5	SNC	−1.016	0.146	0.129	0.928	−1.005	0.069	0.063	0.928
		INC	−1.008	0.130	0.111	0.921	−1.001	0.062	0.054	0.923
	0.7	SNC	−1.011	0.116	0.108	0.922	−1.002	0.059	0.053	0.919
		INC	−1.004	0.110	0.098	0.915	−0.999	0.054	0.049	0.925

SNC, simple nonparametric correction; INC, GMM nonparametric correction. SD, empirical standard deviation across simulated data sets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

Table 2

Simulation Results in the case of a single covariate contaminated with a mixture of normal error.

	Pr($\eta = 1$)		n = 500 Est	SD	SE	CP	n = 2000 Est	SD	SE	CP
		Ideal	−1.000	0.070	0.069	0.958	−1.001	0.034	0.035	0.951
$\sigma^2 = 0.1$		Naive	−0.877	0.066	0.064	0.515	0.877	0.032	0.032	0.026
	0.3	SNC	−1.021	0.184	0.167	0.932	−1.003	0.085	0.080	0.940
		INC	−1.006	0.143	0.127	0.934	−1.001	0.064	0.063	0.944
	0.5	SNC	−1.007	0.138	0.126	0.931	−1.003	0.063	0.062	0.943
		INC	−1.006	0.117	0.106	0.933	−1.002	0.054	0.053	0.942
	0.7	SNC	−1.006	0.112	0.106	0.935	−1.004	0.055	0.052	0.944
		INC	−1.006	0.103	0.096	0.936	−1.003	0.051	0.048	0.938
$\sigma^2 = 0.2$		Naive	−0.784	0.062	0.061	0.074	−0.783	0.030	0.030	0.000
	0.3	SNC	−1.027	0.195	0.170	0.929	−1.005	0.088	0.081	0.927
		INC	−1.006	0.153	0.133	0.930	−1.000	0.069	0.066	0.935
	0.5	SNC	−1.010	0.146	0.127	0.922	−1.004	0.066	0.062	0.935
		INC	−1.006	0.126	0.110	0.913	−1.002	0.058	0.054	0.938
	0.7	SNC	−1.009	0.118	0.107	0.925	−1.004	0.058	0.052	0.932
		INC	−1.005	0.110	0.098	0.923	−1.003	0.053	0.049	0.927

SNC, simple nonparametric correction; INC, GMM nonparametric correction. SD, empirical standard deviation across simulated data sets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

Table 3

Simulation Results in the case of two covariates.

	Pr(η = 1)		*n* = 500				*n* = 2000			
			Est	**SD**	**SE**	**CP**	**Est**	**SD**	**SE**	**CP**
β		Ideal	−1.002	0.071	0.072	0.951	−1.000	0.036	0.036	0.948
		Naive	−0.868	0.068	0.067	0.466	−0.864	0.034	0.033	0.026
	0.3	SNC	−1.019	0.194	0.180	0.939	−1.005	0.092	0.088	0.943
		INC	−1.007	0.155	0.139	0.944	−1.002	0.076	0.070	0.934
	0.5	SNC	−1.014	0.148	0.137	0.942	−1.003	0.071	0.068	0.935
		INC	−1.008	0.130	0.117	0.930	−1.001	0.062	0.059	0.935
	0.7	SNC	−1.012	0.118	0.115	0.955	−1.003	0.060	0.057	0.934
		INC	−1.008	0.108	0.105	0.949	−1.001	0.054	0.053	0.940
γ		Ideal	−1.002	0.073	0.072	0.941	−1.002	0.037	0.036	0.939
		Naive	−0.935	0.072	0.070	0.829	−0.933	0.037	0.035	0.518
	0.3	SNC	−1.013	0.161	0.148	0.923	−1.009	0.075	0.072	0.943
		INC	−1.007	0.107	0.095	0.926	1.003	0.050	0.048	0.943
	0.5	SNC	−1.011	0.121	0.114	0.935	−1.005	0.058	0.056	0.952
		INC	−1.007	0.097	0.087	0.920	−1.003	0.046	0.043	0.941
	0.7	SNC	−1.009	0.100	0.095	0.940	−1.002	0.051	0.047	0.935
		INC	−1.007	0.087	0.082	0.931	−1.003	0.044	0.041	0.921

SNC, simple nonparametric correction; INC, GMM nonparametric correction. SD, empirical standard deviation across simulated data sets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

Table 4

Simulation Results when $R = X + \varepsilon$, $X^2 + \varepsilon$, and $X^4 + \varepsilon$.

		Est	SD	SE	CP
	Ideal	−0.999	0.035	0.035	0.948
	Naive	−0.871	0.033	0.032	0.034
$R = X + \varepsilon$	SNC	−1.003	0.058	0.056	0.939
	INC	−1.001	0.048	0.046	0.940
$R = X^2 + \varepsilon$	SNC	−1.006	0.123	0.115	0.943
	INC	−0.979	0.116	0.103	0.913
$R = X^4 + \varepsilon$	SNC	−1.020	0.141	0.134	0.940
	INC	−0.989	0.131	0.119	0.916

SNC, simple nonparametric correction; INC, GMM nonparametric correction. SD, empirical standard deviation across simulated data sets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 5

Simulation Results when $R = X + b\sqrt{6}\varepsilon_*/(10\pi)$.

		n = 500				*n* = 1000			
		Est	**SD**	**SE**	**CP**	**Est**	**SD**	**SE**	**CP**
	Ideal	−1.009	0.073	0.070	0.940	−1.003	0.049	0.049	0.940
	Naive	−0.878	0.068	0.064	0.514	−0.873	0.046	0.045	0.210
$b = 0.0$	SNC	−1.016	0.137	0.125	0.934	−1.008	0.091	0.088	0.932
	INC	−1.007	0.118	0.105	0.925	−1.003	0.079	0.074	0.928
$b = 0.5$	SNC	−1.095	0.145	0.127	0.882	−1.085	0.096	0.088	0.843
	INC	−1.084	0.125	0.107	0.868	−1.078	0.083	0.075	0.821
$b = 1.0$	SNC	−1.179	0.155	0.129	0.722	−1.166	0.101	0.089	0.565
	INC	−1.164	0.134	0.109	0.699	−1.158	0.088	0.077	0.468
$b = 2.0$	SNC	−1.362	0.185	0.136	0.246	−1.342	0.117	0.092	0.052
	INC	−1.339	0.159	0.117	0.179	−1.329	0.102	0.081	0.024

SNC, simple nonparametric correction; INC, GMM nonparametric correction. SD, empirical standard deviation across simulated data sets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

Table 6

Results for ACTG 175 data.

	logCD4 (β)		Treatment (γ)	
	Est	**SE**	**Est**	**SE**
Naive	−1.465	0.162	−0.474	0.129
SNC	−2.359	0.398	−0.618	0.193
INC	−2.562	0.358	−0.581	0.133

SNC, simple nonparametric correction; INC, GMM nonparametric correction.