On Corrected Score Approach for Proportional Hazards Model with Covariate Measurement Error

Xiao Song
*University of Washington, songx@u.washington.edu*

Yijian Huang
*Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, yhuang5@sph.emory.edu*

Suggested Citation
http://biostats.bepress.com/uwbiostat/paper226

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.
Copyright © 2011 by the authors
1. Introduction

In clinical trials, the proportional hazards model is often used to characterize the relationship between time to failure and covariates. True values of the covariates are required to implement the partial likelihood inference procedure. However, covariates might be measured with error. One example is the AIDS Clinical Trial Group (ACTG) 175 study, a randomized trial to compare zidovudine alone, zidovudine plus didanosine, zidovudine plus zalcitabine, and didanosine alone, in HIV-infected subjects (Hammer et al., 1996). We are interested in the effect of baseline CD4 count on time to AIDS or death. It is well known that the observation of CD4 count is subject to substantial instrumental measurement error and biological variation. Naive approaches that ignore measurement error might lead to biased estimation (Prentice, 1982). Various approaches have been proposed to deal with measurement error and they can be classified in terms of approximate versus consistent estimation and structural versus functional modeling (Carrol, Ruppert, and Stefanski, 1995, Chapter 1.2). The regression calibration approach (e.g. Prentice, 1982; Tsiatis, DeGruttola and Wulfsohn, 1995; Wang et al., 1997) is a frequently used approximate method, which reduces bias relative to naive approaches but may still give erroneous results (Tsiatis and Davidian, 2001). The likelihood based approaches (e.g. DeGruttola and Tu, 1994; Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997) are consistent. However, they rely on normality or other parametric assumptions for both the error and the underlying true covariates. The semiparametric likelihood approaches (Hu, Tsiatis and Davidian, 1998; Song, Davidian and Tsiatis, 2002b) relax the distributional assumptions on the underlying covariates. Unfortunately, they might be computationally intensive like other likelihood based approaches. Two attractive alternatives are the conditional score approach (Tsiatis and Davidian, 2001) and the corrected score approach (Nakamura, 1992; Huang and Wang, 2000, 2003). The former requires the error to be normal, while the latter does not. Both methods are easy to implement and the estimators are consistent without distributional assumptions on the underlying true covariates. The corrected score approach is sometime called correction approach and can be further classi-
fied as parametric correction (Nakamura, 1992) and nonparametric correction (Huang and Wang, 2000, 2003) based on whether there are parametric assumptions on the error.

The essence of the correction approach is to remove bias by correcting the biased “score function”, while the key idea of the conditional score approach is to “condition away” the nuisance random parameters based on sufficient statistics. When the error distribution is assumed to be normal, both parametric correction and conditional score methods have been developed for several commonly used generalized linear models. They are generally different, as in the case of Poisson regression. Moreover, they might not be applicable to the same score equations. For example, for the logistic regression, the conditional score approach (Stefanski and Carroll, 1985) can be used, while Stefanski (1989) showed that there is no standard corrected score estimator in general and Huang and Wang (2001) proposed an alternative correction method based on related estimating functions. An exception is the proportional hazards model, for which both techniques can be used and the estimators are asymptotically equivalent in the case of normal error (Tsiatis and Davidian, 2001). However, their relative finite sample performance is not clear. In addition, for both approaches, the estimating equations might have multiple solutions or no finite solution. This essentially is not a computational issue and hence is unsolvable by simply improving the computing algorithm and computer speed.

In this paper, we investigate the relative performance of the conditional score estimator and the parametric correction estimator in the case of normal error with small sample and large measurement error and discover that the former is superior, which may be explained by the unbiasedness of its estimating equation. Noting that the former can be viewed as a refinement of the latter, we are motivated to extend this refinement to both the parametric correction method and the nonparametric correction method in general. The refined estimators are asymptotically equivalent to their standard counterparts but have better finite
sample properties. We give the model definition in Section 2. The conditional score estimator and the parametric correction estimator are compared in Section 3. In Section 4, we refine correction estimators in general. Section 5 presents an analysis of the ACTG 175 data. Possible extensions are discussed in Section 6.

2. Model Definition

For simplicity, we only consider time-independent covariates. Extension to time-dependent covariates is discussed in Section 6. For subject $i = 1, \ldots, n$, let $T_i$ denote failure time and $C_i$ denote censoring time. The observed survival data are $V_i = \min\{T_i, C_i\}$ and $\Delta_i = I(T_i \leq C_i)$. Let $X_i$ denote $q$ time-independent covariates with observed value $W_i = X_i + e_i$, where $e_i$ is the error with the variance $\Sigma$. If a covariate is not error-contaminated, the corresponding error variance is 0. We assume that $(T_i, C_i, X_i, e_i)$ are independent and identically distributed across $i$, and $e_i$ is independent of $(T_i, C_i, X_i)$. The hazard of failure $\lambda_i(t)$ depends on $X_i$ through the proportional hazards model

$$
\lambda_i(t) = \lim_{dt \to 0} dt^{-1} \Pr(t \leq T_i < t + dt | T_i \geq t, X_i, C_i) = \lambda_0(t) \exp(\beta_0^T X_i),
$$

where $\lambda_0(t)$ is an unspecified baseline function, $\beta_0$ is a $q$-dimensional vector of parameters. Equation (1) makes the explicit assumption that censoring is noninformative. Our interest focuses on estimation of $\beta_0$.

3. Parametric Correction and Conditional Score Estimators

3.1 Estimating Equations

Let $L$ be a positive constant such that $\Pr(V \geq L) > 0$. Under the normal error assumption, when $\Sigma$ is known, the standard parametric correction estimating equation (Nakamura, 1992) is

$$
\sum_{i=1}^{n} \int_{0}^{L} \left\{ W_i + \Sigma \beta - \frac{\sum_{j=1}^{n} Y_j(u)W_j \exp(\beta^T W_j)}{\sum_{j=1}^{n} Y_j(u) \exp(\beta^T W_j)} \right\} dN_i(u) = 0,
$$

(2)
where \( N_i(u) = I(V_i \leq u, \Delta_i = 1) \) is the counting process and \( Y_i(u) = I(V_i \geq u) \) is the at risk process. In contrast, with simple algebra, the conditional score estimating equation (Tsiatis and Davidian, 2001; Song, Tsiatis and Davidian, 2002a) can be rewritten as

\[
\sum_{i=1}^{n} \int_0^L \left\{ W_i + \Sigma \beta - \frac{\sum_{j=1}^{n} Y_j(u) \{ W_j + \Sigma \beta dN_j(u) \} \exp [\beta^T \{ W_j + \Sigma \beta dN_j(u) \}] \} \sum_{j=1}^{n} Y_j(u) \exp [\beta^T \{ W_j + \Sigma \beta dN_j(u) \}] \} \right\} dN_i(u) = 0.
\]

(3)

Both (2) and (3) are summations across all death times and reduce to the usual partial likelihood score equations when there is no measurement error. The only difference between (2) and (3) lies in the “average” terms, where the error contaminated observations \( W_j \) in (2) are replaced by the adjusted values \( W_j + \Sigma \beta dN_j(u) \) in (3). Thus, at each failure time, if we treat \( W_i + \Sigma \beta \) as pseudo “observations”, the “average” term in (3) is a weighted average of these pseudo “observations” in the risk set, while the “average” term in (2) is a weighted average of the unadjusted covariates \( W_j \). In fact, we can show that the conditional score estimating equation is unbiased and the proof is sketched in Appendix A. This suggests that the conditional score estimator might perform better in the case of finite samples.

The estimating equation (2) is the first-order correction termed by Nakamura (1992). The consistency of this estimator was later proved by Kong and Gu (1999). Nakamura (1992) also proposed an approximate second-order parametric correction, which performs better than the first order correction when \( \beta^T \Sigma \beta \) is small. Here we are interested in comparing these estimators in the case of small sample and large measurement error.

### 3.2 Simulation Studies

We conducted simulation studies to compare the parametric correction and conditional score estimators in the case of normal error under the following scenarios. We considered the simple case of a single covariate \( X \) with one observation per subject. The error variance \( \Sigma = \sigma^2 \) was assumed to be known. We generated \( X \) from a normal distribution with mean
2 and variance 1. The error was generated from a normal distribution with mean 0 and variance $\sigma^2 = 0.3$ and 0.6. The true Cox model coefficient was taken to be $\beta_0 = -1$. The baseline hazard $\lambda_0(u)$ was a constant 1. Censoring was generated from a uniform distribution on $[0,7]$, leading to a censoring rate of about 61%.

We carried out simulations for $n = 300$ and 600. For each scenario, 1000 Monte Carlo data sets were simulated. For each data set, we fitted the model using the “ideal” approach, in which the true values of $X$ are used, the first and second-order parametric correction approaches and the conditional score approach, for which the variances are estimated via the sandwich technique. The Newton-Raphson method was used to solve the estimating equations with the naive estimate as the starting value. If the Newton-Raphson method failed to find a solution or the distance between the found solution and the naive estimate was larger than 4, the bisection method was then used. If more than one solution was found, the one closest to the naive estimate was selected.

The results are given in Table 1. The conditional score estimator is closer to the unachievable “ideal” estimator in all the cases, while the parametric correction estimators show some bias in the case of large error. For all these three estimators, the empirical standard deviation could be much larger than the estimated standard error in the case of small sample size and large error. This is caused by some extreme outliers as indicated by the scatter plots in Figure 1. There are much more outliers for the parametric correction methods than for the conditional score method. We presented the Q-Q plots of the estimates in Figure 2. It seems that the distribution of the conditional score estimator is closer to the normal than those of the parametric correction estimators. The second-order parametric correction does not perform better than the first order parametric correction in the case of small sample and large measurement error. In the case of $\sigma^2 = 0.6$, we failed to get the second-order parametric estimates for 14 and 4 data sets for $n = 300$ and 600, respectively. Typical score
plots for two simulation data sets are shown in Figure 3. The estimates are close to the truth in Figure 3(a) but are outliers in Figure 3(b). The unbiasedness of the conditional score estimating equation might account for the superiority of this estimator.

4. Refinement for Correction Method

The conditional score estimator can be viewed as a refinement of the parametric correction estimator. However, the conditional score approach requires the normality assumption of the error. In this section, we extend this refinement to correction methods in general.

4.1 One Observation Per Subject

4.1.1 Estimating Equations

Assume $D(\beta) = E\{e_i \exp(\beta^T e_i)\} / E\{\exp(\beta^T e_i)\}$ is known. The standard correction estimator under any error distribution (Huang and Wang 2003) is the solution to the estimating equation

$$\tilde{U}(\beta, D) = n^{-1} \sum_{i=1}^{n} \int_{0}^{L} \left\{ W_i + D(\beta) - \frac{\tilde{S}_1(u, \beta, W)}{\tilde{S}_0(u, \beta, W)} \right\} dN_i(u) = 0, \quad (4)$$

where, for $r = 0, 1$, $\tilde{S}_r(u, \beta, W) = n^{-1} \sum_{i=1}^{n} \tilde{S}_{ri}(u, \beta, W)$ and

$$\tilde{S}_{ri}(u, \beta, W) = Y_i(u) W_i^r \exp\{\beta^T W_i\}.$$

Note that when the error is normal, $D(\beta) = \Sigma \beta$ and (4) reduces to (2). By mimicking the weighted average form of the conditional score approach in the normal case, we propose the refined correction estimating equation

$$\hat{U}(\beta, D) = n^{-1} \sum_{i=1}^{n} \int_{0}^{L} \left\{ W_i + D(\beta) - \frac{S_1(u, \beta, W, D)}{S_0(u, \beta, W, D)} \right\} dN_i(u) = 0, \quad (5)$$

where, for $r = 0, 1$, $S_r(u, \beta, W, D) = n^{-1} \sum_{i=1}^{n} S_{ri}(u, \beta, W, D)$ and

$$S_{ri}(u, \beta, W, D) = Y_i(u) \{W_i + D(\beta) dN_i(u)\}^r \exp\{\beta^T \{W_i + D(\beta) dN_i(u)\}\}.$$

Like the standard correction estimating equation, (5) reduces to the usual partial likelihood score equations when there is no error. In the case of normal error, (5) reduces to the conditional score estimating equation (3).
4.1.2 Properties Now we consider the large sample properties of \( \hat{\beta} \), the refined correction estimator of \( \beta_0 \). Under similar regularity conditions as those given in Huang and Wang (2000), we can show that \( \hat{\beta} \) exists asymptotically in a compact neighborhood of \( \beta_0 \) and is consistent and asymptotically normal. The asymptotic variance can be consistently estimated by

\[
\hat{V} = n^{-1} \hat{A}^{-1}(\hat{\beta}) \hat{B}(\hat{\beta}) \{ \hat{A}^{-1}(\hat{\beta}) \}^T,
\]

where \( \hat{A}(\beta) = \partial \hat{U}(\beta)/\partial \beta^T \), \( \hat{B}(\beta) = n^{-1} \sum_{i=1}^{n} \hat{\omega}_i(\beta; W, D) \hat{\omega}_i^T(\beta; W, D) \), and

\[
\hat{\omega}_i(\beta; W, D) = \int_0^L \left[ \left\{ W_i + D(\beta) - \frac{S_1(u, \beta, W, D)}{S_0(u, \beta, W, D)} \right\} dN_i(u) \right. \\
- \left. \left\{ S_{1i}(u, \beta, W, D) - \frac{S_{0i}(u, \beta, W, D) S_1(u, \beta, W, D)}{S_0(u, \beta, W, D)} \right\} \right] n^{-1} \sum_{j=1}^{n} dN_j(u) \right).
\]

The arguments are similar to those in Huang and Wang (2000) and hence omitted for brevity.

In addition, we can show that the refined and standard correction estimators are asymptotically equivalent. The proof is sketched in Appendix B. We discuss the existence and tail behaviors of these approaches in Section 4.2 for replication data.

To investigate the performance of the correction estimators in the case of non-normal distributions, we conducted simulation studies under the same scenarios as those in Section 3.2 except that the error was generated from a bimodal mixture of normals as described in Davidian and Gallant (1993) with several settings of \( p \), the mixing proportion, and \( \text{sep} \), the distance between the means of the two normals in unit of the common standard deviation. We observe similar phenomena as those in the normal case. The results are omitted for brevity. Thus the refined correction approach performs better in general than the standard correction approach in situations of small sample size and large measurement error.

4.2 Refined Correction Estimators With Replication Data

In practice, \( D(\cdot) \) is usually unknown. As a general technique when measurement error exists, \( D(\cdot) \) can be estimated from internal or external assessment data, such as replication,
validation or instrumental data. If the error is assumed to have a parametric distribution with parameter $\theta$, $D(\cdot)$ can be estimated based on a consistent estimator of $\theta$, such as the methods of moments estimator. However, it might be difficult to make appropriate distributional assumptions for the error in practice. Alternatively, we can estimate $D(\cdot)$ nonparametrically with no distributional assumption required for the error, as proposed by Huang and Wang (2003) for the standard nonparametric correction estimator. We call the former refined parametric correction and the latter refined nonparametric correction in accordance to the standard ones.

As illustration, we give the refined parametric and nonparametric correction estimators when replicated observations are available. These estimators can be easily extended to the cases when other error assessment data are available. The consistency and asymptotic normality remain if $D(\cdot)$ is substituted by a regular asymptotically linear estimator by arguments similar to those when $D(\cdot)$ is known.

For simplicity, we assume that only one covariate, $X_1$, is measured with error; this can be easily extended to multiple error-contaminated covariates. Let $X_2$ denote $q_2 = q - 1$ covariates that are exactly measured. Thus $X = (X_1, X_2^T)^T$. Write $\beta = (\beta_1, \beta_2^T)^T$ and $e = (e_1, 0_{q_2}^T)^T$ correspondingly, where $0_{q_2}$ is a $q_2 \times 1$ zero vector. For the $i$th subject, let $W_{i11}, \ldots, W_{imi}$ be $m_i$ observations of $X_{1i}$, that is, $W_{1ij} = X_{1i} + e_{1ij}$ and $e_{1ij}$ are independent and identically distributed. Thus $D(\beta) = \{D_1(\beta), 0_{q_2}^T\}^T$, where $D_1(\beta) = \frac{E\{e_{1ij} \exp(\beta_1 e_{1ij})\}}{E\{\exp(\beta_1 e_{1ij})\}} - E(e_1)$. Write $W_{ij} = (W_{1ij}, X_{2i}^T)^T$, and $A_1$ as the operator averaging over $j = 1$ through $m_i$, that is, $A_1 h(W_{ij}) = m_i^{-1} \sum_{j=1}^{m_i} h(W_{ij})$.

4.2.1 Parametric Correction As a simple extension of (4), for replication data, the standard parametric correction estimating equation is obtained by replacing $W_i$ by the average $A_1 W_{ij}$ and $D(\beta)$ by $D_i(\beta) = \{D_{1i}(\beta), 0_{q_2}^T\}^T$ in (4), where

$$D_{1i}(\beta) = \frac{E\{A_1 e_{ij} \exp(\beta_1 A_1 e_{ij})\}}{E\{\exp(\beta_1 A_1 e_{ij})\}} - E(A_1 e_{ij}).$$
Hence the corresponding refined estimating equation is
\[ \hat{U}_1(\beta, D) = n^{-1} \sum_{i=1}^{n} \int_{0}^{L} \left[ A_1 W_{ij} + D_i(\beta) - \frac{S^#_1(u, \beta, W, D)}{S^*_0(u, \beta, W, D)} \right] dN_i(u) = 0, \]  
(6)

where, for \( r = 0, 1, S^#_r(u, \beta, W, D) = n^{-1} \sum_{i=1}^{n} S^#_r(u, \beta, W, D) \) and
\[ S^*_r(u, \beta, W, D) = Y_i(u) \{ A_1 W_{ij} + D_i(\beta) dN_i(u) \}^r \exp \left[ \beta^T \{ A_1 W_{ij} + D_i(\beta) dN_i(u) \} \right]. \]

A consistent estimator of \( D_{1i}(\cdot), \hat{D}_{P,1i}(\cdot) \), is substituted in (6). When \( m_i = m \) is a constant, it is simple to show that \( D_{1i}(\beta) = \hat{D}_{P,1i}(\beta/m) \) and (6) reduces to the conditional score estimating equations for normal error with \( \hat{D}_{P,1i}(\beta) = \hat{\Sigma}_1 \beta/m \), where \( \hat{\Sigma}_1 \) is the methods of moments estimator for \( \Sigma_1 \), the variance of \( e_1 \). We call this correction type A parametric correction.

Following Huang and Wang (2003), an alternative set of standard parametric correction estimating equation can be written as
\[ \hat{U}_2(\beta, D) = n^{-1} \sum_{i=1}^{n} \int_{0}^{L} \left[ A_1 \{ W_{ij} + D_i(\beta) \} - \frac{\hat{S}^*_1(u, \beta, W)}{S^*_0(u, \beta, W)} \right] dN_i(u) = 0, \]  
(7)

where, for \( r = 0, 1, \hat{S}^*_r(u, \beta, W) = n^{-1} \sum_{i=1}^{n} \hat{S}^*_r(u, \beta, W) \) and
\[ \hat{S}^*_r(u, \beta, W) = A_1 Y_i(u) W_{ij}^r \exp \left[ \beta^T \{ W_{ij} + D_i(\beta) dN_i(u) \} \right]. \]

The corresponding refined correction estimator can be obtained from
\[ \hat{U}_2(\beta, D) = n^{-1} \sum_{i=1}^{n} \int_{0}^{L} \left[ A_1 \{ W_{ij} + D_i(\beta) \} - \frac{S^*_1(u, \beta, W, D)}{S^*_0(u, \beta, W, D)} \right] dN_i(u) = 0, \]  
(8)

where, for \( r = 0, 1, S^*_r(u, \beta, W, D) = n^{-1} \sum_{i=1}^{n} S^*_r(u, \beta, W, D) \) and
\[ S^*_r(u, \beta, W, D) = A_1 Y_i(u) \{ W_{ij} + D_i(\beta) dN_i(u) \}^r \exp \left[ \beta^T \{ W_{ij} + D_1(\beta) dN_i(u) \} \right], \]

with \( D_1(\cdot) \) in (8) replaced by some parametric estimator \( \hat{D}_{P,1}(\cdot) \). For example, if the error is normal, \( \hat{D}_{P,1}(\beta) = \hat{\Sigma}_1 \beta \). This approach is called type B parametric correction. The type A correction might be more intuitive, but the type B correction surprisingly could have better finite sample performance, as we illustrate in Section 4.3.
4.2.2 Nonparametric Correction

The standard nonparametric correction estimator proposed by Huang and Wang (2003) can be obtained from (7) with \( D_1(\cdot) \) replaced by a nonparametric estimator \( \hat{D}_{NP,1}(\cdot) \). For simplicity, here we use

\[
\hat{D}_{NP,1}(\beta_1) = \frac{n^{-1} \sum_{i=1}^{n} I(m_i \geq 2) A_2(W_{1ij} - W_{1ij'}) \exp(\beta_1 W_{1ij})}{n^{-1} \sum_{i=1}^{n} I(m_i \geq 2) A_1 \exp(\beta_1 W_{1ij})},
\]

where \( A_2 \) is an operator averaging over \( G_i = \{ (W_{1ij}, W_{1ij'}) : j \neq j' \} \), the set of ordered pairs of observations for subject \( i \) with \( m_i \geq 2 \), that is,

\[
A_2 h(W_{1ij}, W_{1ij'}) = \{ m_i(m_i - 1) \}^{-1} \sum_{(W_{1ij}, W_{1ij'}) \in G_i} h(W_{1ij}, W_{1ij'}).
\]

The corresponding refined estimator can be obtained from (8) with \( D_1(\cdot) \) replaced by \( \hat{D}_{NP,1}(\cdot) \). The correction using \( \hat{D}_{NP,1}(\cdot) \) is called type 1 nonparametric correction. This can be easily extended to other choices of \( \hat{D}_{NP,1}(\cdot) \).

Huang and Wang (2000) proposed an alternative standard nonparametric correction approach with the estimating equation given by

\[
n^{-1} \sum_{i=1}^{n} \int_{0}^{L} \left\{ A_1 W_{ij} - \frac{\sum_{j=1}^{n} A_2 Y_j(u) W_{jk} \exp\{\beta^T W_{jk'}\}}{\sum_{j=1}^{n} A_1 Y_j(u) \exp\{\beta^T W_{jk}\}} \right\} dN_i(u) = 0. 
\] (9)

An important difference between (8) and (9) is that the latter can only be applied to replication data, while the former can be utilized for various error assessment data. However, with simple algebra, we can rewrite (9) as

\[
n^{-1} \sum_{i=1}^{n} \int_{0}^{L} \left\{ A_1 W_{ij} + \hat{D}^*_{NP}(\beta, u) - \frac{\tilde{S}^*_1(u, \beta, W)}{\tilde{S}^*_0(u, \beta, W)} \right\} dN_i(u) = 0, 
\] (10)

where \( \hat{D}^*_{NP}(\beta, u) = (\hat{D}^*_{NP,1}(\beta_1, u), 0_{q_2})^T \), and

\[
\hat{D}^*_{NP,1}(\beta_1, u) = \frac{n^{-1} \sum_{i=1}^{n} I(m_i \geq 2) Y_i(u) A_2 \{ W_{1ij} - W_{1ij'} \} \exp\{\beta_1 W_{1ij}\}}{n^{-1} \sum_{i=1}^{n} I(m_i \geq 2) Y_i(u) A_1 \exp\{\beta_1 W_{1ij}\}}.
\]

(10) has a form similar to (7) with \( D_1(\beta_1) \) replaced by \( \hat{D}^*_{NP,1}(\beta_1, u) \), which is analogous to \( \hat{D}_{NP,1}(\beta_1) \) with only at-risk subjects used in computation. This form is an improvement.
over (9) since it allows including subjects with no replicated observations and can be easily extended to other error assessment data. More importantly, it draws a connection between the two standard nonparametric correction estimators. Applying the refining technique on (10), an alternative refined nonparametric correction estimator can be obtained from (8) with $D_1(\beta_1)$ replaced by $\hat{D}_{NP,1}^*(\beta_1, u)$. The correction using $\hat{D}_{NP,1}^*(\beta_1, u)$ is called type 2 nonparametric correction.

4.2.3 Properties The consistency and asymptotical normality of the refined estimators for replication data can be shown by arguments similar to those for a single observation per subject with $D(\cdot)$ known. The variances for the parametric correction estimators can be obtained by the sandwich technique. The estimators for the variances for the refined nonparametric estimators are given in Appendix C.

For finite samples, the standard nonparametric correction estimators might not exist (Huang and Wang, 2000, 2003). Now we consider the existence of the other estimators in the case of a single covariate $X$. Let $W_{i1}, \ldots, W_{im_i}$ be the observations of $X$ for the $i$th subject, $\hat{D}(\beta)$ denote any parametric or nonparametric estimator for $D(\beta)$ and $\hat{U}(\beta, \hat{D})$ denote any standard or refined correction estimating function. Let $u_1 < \ldots < u_R$ be the death times. Define $W_{\text{max}, a} = \max\{(i,j): dN_i(u_a) = 1\} W_{ij}$ and $W_{\text{min}, a} = \min\{(i,j): dN_i(u_a) = 1\} W_{ij}$ for $a = 1, \ldots, R$. If there exists a failure time $u_{a_0}$ such that $W_{\text{min}, a_0} < W_{\text{max}, a_0}$, then we can show that

A. For the refined type 1 and type 2 nonparametric approaches, $0 < \lim_{\beta \to -\infty} \hat{U}(\beta, \hat{D}) < +\infty$ and $-\infty < \lim_{\beta \to +\infty} \hat{U}(\beta, \hat{D}) < 0$;

B. For the parametric approaches, if

$$\hat{D}(\beta) \text{ is continuous, } \lim_{\beta \to -\infty} \hat{D}(\beta) = -\infty, \text{ and } \lim_{\beta \to +\infty} \hat{D}(\beta) = +\infty, \quad (11)$$

then
i. For the refined type A parametric approach, \( \lim_{\beta \to -\infty} \hat{U}(\beta, \hat{D}) \geq 0 \) and \( \lim_{\beta \to +\infty} \hat{U}(\beta, \hat{D}) \leq 0 \), with the inequalities hold if and only if there exist two subjects failing at the same time with different \( A_1 W_{ij} \). If there are no tied failure times, there exist \( b_1 \) and \( b_2, b_1 < b_2 \), such that \( U(\beta, \hat{D}) < 0 \) for \( \beta < b_1 \) and \( U(\beta, \hat{D}) > 0 \) for \( \beta > b_2 \).

ii. For the refined type B parametric approach, \( 0 < \lim_{\beta \to -\infty} \hat{U}(\beta, \hat{D}) < +\infty \) and \(-\infty < \lim_{\beta \to +\infty} \hat{U}(\beta, \hat{D}) < 0; \)

iii. For the standard type A and type B parametric approaches, \( \lim_{\beta \to -\infty} \hat{U}(\beta, \hat{D}) = -\infty \) and \( \lim_{\beta \to +\infty} \hat{U}(\beta, \hat{D}) = +\infty \).

The proof is given in Appendix D(2004). Thus the refined nonparametric estimators always exist regardless the error distribution, so does the refined type B parametric estimator under condition (11), for example, when the error is normal or a mixture of normals. This is also the case for the refined type A parametric correction when there exist two subjects failing at the same time with different \( A_1 W_{ij} \) under condition (11). In contrast, the standard parametric correction approaches always have a zero-crossing around which it is not monotonically decreasing. However, this solution is not the right one since the limit of the standard parametric correction estimating function converges uniformly over a compact neighborhood of \( \beta_0 \) to the limit of the ideal partial likelihood score function, which is monotonically decreasing. Hence, the standard parametric estimating equations always have an odd number of solutions and it will lead to an outlier when there is only one solution. Similarly, the refined type A parametric estimating equation also has inconsistent solutions in the case of no tied failure times.

4.3 Simulation Studies

We used the same scenarios as those in Section 3.2 except that the error was generated from a skewed bimodal mixture of normals as described in Davidian and Gallant (1993)
(p = 0.3 and sep=2) with mean 0 and variance $\sigma^2 = 0.6$ and 1.2 and each subject had two replicated observations on $X$. For simplicity, for the parametric approaches, we assume only the error variance is unknown, which is estimated using the methods of moments. For each data set, we fitted the model using the “ideal” approach and the various correction methods described above. Our major interest is in the relative performance of the refined estimators compared to their standard counterparts. The results are given in Table 2. Figure 4 presents the scatter plots of refined estimates and the corresponding standard estimates versus the ideal estimates. Various standard estimators perform differently. However, for each one of them, the corresponding refined estimator has smaller bias, better coverage and is closer to the normal (Q-Q plots not shown). Among the refined estimators, the type 2 nonparametric approach has the smallest empirical deviation and works best in all these cases. The standard type 1 or type 2 nonparametric estimates are not found for some simulated data sets. For all the standard estimators, the empirical standard deviation can be much larger than the estimated standard error in the case of small sample size and large error because of outliers. These problems become worse as the error variance increases and the sample size decreases. Figure 5 presents the typical score plots from two simulated data sets. In Figure 5(a), all approaches have a solution close to the truth. In Figure 5(b), the standard parametric estimates are outliers and no solutions are found for the standard nonparametric approaches.

We also have carried out simulations when the error is normal or a mixture of normals with different parameters $p$ and sep. Besides to similar results as those shown above, we further observed that the refined type A parametric correction approach might generate outliers but the number of outliers are much less than the corresponding standard approach. When the error variance is smaller, the refined parametric estimators may be more efficient than the refined nonparametric estimators for moderate sample size. Overall, the simulation evidence suggests that the refined correction estimators perform better than their standard
counterparts, the nonparametric estimators are robust to the error distribution and the parametric estimators can be biased if the error distribution is misspecified. Hence we recommend using the refined nonparametric estimators in practice.

5. Application to ACTG 175 Data

We apply the correction approaches to the ACTG 175 data. We focus on assessing the effect of baseline CD4 count on time to AIDS or death in antiretroviral-naive patients. In the study, 1,067 patients had no antiretroviral therapy at baseline, among which 1,036 patients had two CD4 measurements within 3 weeks of randomization and 31 patients had only one CD4 measurement, which we treated as baseline CD4 measurements. Figure 6 shows the Q-Q plot of the deviations from the average of log(CD4) for each subject. It seems that the error distribution deviates slightly from the normal. The estimated variances are 0.033 for error and 0.076 for the true underlying log(CD4), with a ratio of 0.43. The primary analysis found zidovudine alone to be inferior to the other three therapies; thus, our further investigations focused on two treatment groups, zidovudine alone and the combination of the other three. We consider a Cox regression model with two covariates, log(CD4) and a treatment indicator, 0 for zidovudine alone and 1 for the combination of the other three.

We fit the model using the standard and refined correction approaches with mean zero normal error assumption for the parametric methods. The results are shown in Table 3. In this case, all the approaches have similar estimates for both coefficients. This is not surprising since the sample size is large and as we observed in simulations, this is the case for most data sets. To illustrate the possible deviance between each pair of refined and standard approaches, we took random subsamples of size 210 from the 1,087 patients and fit the model for 1000 such subsamples. Figure 7 shows the scatter plots of the standard estimates versus the corresponding refined estimates across all these subsamples. There exists obvious difference between some of the refined and standard correction estimates.
from the type B parametric correction and type 1 and type 2 nonparametric correction approaches. In addition, no standard type 1 nonparametric correction estimates were found for 46 subsamples.

6. Discussion

We have refined the correction approach for the proportional hazards model with covariate measurement error. The refined estimators are asymptotically equivalent to the standard counterparts, but have better finite sample performance in the case of small sample and large measurement error. We have shown that the refined estimators exist in the case of a single covariate using replication data, with some restriction on the error distribution for the parametric estimators. The investigation of the existence of these estimators in the case of multiple covariates confronts similar difficulties as those for systems of nonlinear equations and remains an open problem. Based on the general existence of the refined estimators when $D(\beta_0)$ is substituted in the estimating function, we conjecture that the refined estimators might exist more generally than the standard ones in the case of multiple covariates.

The model as presented here includes only error contaminated time-independent covariates. It can be further extended to the case of time-dependent covariates measured with error based on the normal case conditional score approach (Tsiatis and Davidian, 2001; Song, Davidian and Tsiatis, 2002a).

ACKNOWLEDGEMENTS

This research was supported by NIH grants U01-AI46702 (Song), CA90747 and AI29168 (Huang), and career development funds from the University of Washington Department of Biostatistics (Song). Part of this work was done when the second author was associate member at the Fred Hutchinson Cancer Research Center.
Résumé

REFERENCES


**Appendix A**

*Unbiasedness of the Conditional Score Estimating Equation*

Since $T$ is continuous, $\Pr(\sum_{i=1}^n dN_i(t) \leq 1) = 1$. Let $X = (X_1, \ldots, X_n)$, $W = (W_1, \ldots, W_n)$, $Y(t) = \{Y_1(t), \ldots, Y_n(t)\}$, $dN(t) = \{dN_1(t), \ldots, dN_n(t)\}$. Define $\mathcal{H}(t) = \{Y(t), \sum_{i=1}^n dN_i(t) = 1\}$. Given $X_i$ and $Y_i(t)$, $dN_i(t)$ is Bernoulli distributed with probability $\lambda_i(t) \exp(\beta_0^T X_i)Y_i(t)dt$.

Therefore, the conditional distribution of $dN(t)$ given $\mathcal{H}(t)$ and $X$ is

$$p\{dN(t)|\mathcal{H}(t), X\} = \prod_{i=1}^n \left\{ \frac{Y_i(t) \exp(\beta_0^T X_i)}{\sum_{j=1}^n Y_j(t) \exp(\beta_0^T X_j)} \right\}^{dN_i(t)} ,$$

where $\theta_0 = 1$. Note that the conditional distribution of $W_i$ given $X_i$ is normal with mean $X_i$ and variance $\Sigma$. Thus, with some simple algebra, we can show

$$p\{dN(t), W|\mathcal{H}(t), X\} = h_1 \{dN(t), W, Y\} h_2 \{\xi(t), X, Y\} ,$$

where

$$h_1 \{dN(t), W, Y\} = \prod_{i=1}^n (2\pi)^{-q/2} |\Sigma|^{-1/2} \{Y_i(t)\}^{dN_i(t)} \exp \left\{ -W_i^T \Sigma^{-1} W_i/2 \right\} ,$$

$$h_2 \{\xi(t), X, Y\} = \left\{ \sum_{i=1}^n Y_i(t) \exp(\beta_0^T X_i) \right\}^{-1} \prod_{i=1}^n \exp \left\{ -X_i^T \Sigma^{-1} X_i/2 \right\} \exp \left\{ -\xi_i^T (t) \Sigma^{-1} X_i \right\} ,$$

and $\xi(t) = \{\xi_1(t), \ldots, \xi_n(t)\}$, $\xi_i = W_i + \Sigma \beta_0 dN_i(t)$. Thus $\xi(t)$ is a sufficient statistic for $X$, which implies that the conditional distribution of $dN(t)$ given $\xi(t)$ and $\mathcal{H}(t)$ does not depend
on X. With some algebra, we can show that

\[ \Pr \{dN_i(t) = 1|H(t), \xi(t)\} = \frac{Y_i(t) \exp\{\beta^T \xi_i\}}{\sum_{j=1}^n Y_j(t) \exp\{\beta^T \xi_j\}}. \]  

(A.1)

Let \( V(1) \leq \ldots \leq V(n) \) be the ordered statistics for \( V_1, \ldots, V_n \). Then from (A.1),

\[ E\{dN_i(V(k))|H(V(k)), \xi(V(k)), V(k)\} = \frac{Y_i(V(k)) \exp\{\beta^T \xi_i(V(k))\}}{\sum_{j=1}^n Y_j(V(k)) \exp\{\beta^T \xi_j(V(k))\}}. \]  

(A.2)

Note that the conditional score estimating function (3) can be written as

\[ \hat{U}_{CS}(\beta) = \sum_{k=1}^n \eta(\beta, V(k)), \]

where

\[ \eta(\beta, V(k)) = \left[ \sum_{i=1}^n \xi_i(V(k))dN_i(V(k)) - \frac{\sum_{i=1}^n Y_i(V(k))\xi_i(V(k)) \exp\{\beta^T \xi_i(V(k))\}}{\sum_{j=1}^n Y_j(V(k)) \exp\{\beta^T \xi_j(V(k))\}} \sum_{i=1}^n dN_i(V(k)) \right]. \]

Using (A.2), it is simple to show that \( E\{\eta(\beta_0, V(k))|H(V(k)), \xi(V(k)), V(k)\} = 0 \). It follows that \( E\{\hat{U}_{CS}(\beta_0)\} = 0 \). That is, the conditional score estimating equation is unbiased.

**APPENDIX B**

Asymptotic Equivalence of the Refined and Standard Corrected Score Estimators

For simplicity, we only give the proof when \( D(\cdot) \) is known. The proof is analogous when \( D(\cdot) \) is estimated parametrically or nonparametrically.

Using arguments similar to those in Huang and Wang (2000), we can show that

\[ n^{1/2} \hat{U}(\beta) = n^{-1/2} \sum_{i=1}^n \omega_i(\beta) + o_p(1), \]

where

\[ \omega_i(\beta) = \int_0^L \left\{ W_i + D(\beta)dN_i(u) - \frac{s_1(u, \beta)}{s_0(u, \beta)} \right\} \left[ dN_i(u) - \frac{S_0(u, \beta, W, D)}{s_0(u, \beta)} dE\{N(u)\} \right] \]

where \( s_r(u, \beta) = \lim_{n \to \infty} S_r(u, \beta, W, D) = \lim_{n \to \infty} \tilde{S}_r(u, \beta, W) \) for \( r = 0, 1 \). Similarly, the standard corrected score function can be represented as

\[ n^{1/2} \tilde{U}(\beta) = n^{-1/2} \sum_{i=1}^n \tilde{\omega}_i(\beta) + o_p(1), \]
where
\[ \tilde{\omega}_i(\beta) = \int_0^L \left\{ W_i - \frac{s_1(u, \beta)}{s_0(u, \beta)} \right\} \left[ dN_i(u) - \frac{\tilde{S}_0(u, \beta, W)}{s_0(u, \beta)} \right] + D(\beta) dN_i(u). \]

Hence
\[ n^{1/2} \left\{ \hat{U}(\beta) - \tilde{U}(\beta) \right\} = n^{-1/2} \sum_{i=1}^n \left( \int_0^L \left\{ W_i - \frac{s_1(u, \beta)}{s_0(u, \beta)} \right\} \left[ \frac{S_0(u, \beta, W, D) - \tilde{S}_0(u, \beta, W)}{s_0(u, \beta)} \right] dE\{N(u)\} \right) \]
\[ - \int_0^L D(\beta) dN_i(u) \frac{S_0(u, \beta, W, D)}{s_0(u, \beta)} dE\{N(u)\} \]
\[ + o_p(1). \]

Since \( S_0(u, \beta, W, D) - \tilde{S}_0(u, \beta, W) = 0 \) almost surely with respective to the Lebesgue measure, the integrations in (B.3) are equal to \( 0_p(1) \). Similarly, the integrations in (B.4) are equal to \( 0_p(1) \). Therefore, \( n^{1/2} \left\{ \hat{U}(\beta) - \tilde{U}(\beta) \right\} = o_p(1) \).

Next, using a Taylor series expansion, we have
\[ 0 = n^{1/2} \hat{U}(\hat{\beta}) = n^{1/2} \hat{U}(\beta_0) + \hat{A}(\beta^*) n^{1/2} (\hat{\beta} - \beta_0), \]
where \( \beta^* \) is on the line between \( \hat{\beta} \) and \( \beta_0 \). Similarly, letting \( \tilde{\beta} \) be the correction estimator, we have
\[ 0 = n^{1/2} \tilde{U}(\tilde{\beta}) = n^{1/2} \tilde{U}(\beta_0) + \tilde{A}(\beta^{**}) n^{1/2} (\tilde{\beta} - \beta_0), \]
where \( \tilde{A}(\beta) = \partial \tilde{U}(\beta)/\partial \beta^T \), and \( \beta^{**} \) is on the line between \( \tilde{\beta} \) and \( \beta_0 \). We can show \( \hat{A}(\beta^*) \) and \( \tilde{A}(\beta^{**}) \) converges to \( A(\beta_0) = \partial U(\beta)/\partial \beta^T \) with arguments similar to those in Huang and Wang (2000). Therefore,
\[ n^{1/2} (\tilde{\beta} - \beta_0) = n^{1/2} (\hat{\beta} - \beta_0) - n^{1/2} (\tilde{\beta} - \beta_0) \]
\[ = - \left\{ \hat{A}(\beta^*) \right\}^{-1} n^{1/2} \hat{U}(\beta_0) + \left\{ \tilde{A}(\beta^{**}) \right\}^{-1} n^{1/2} \tilde{U}(\beta_0) \]
\[ = - \left\{ A(\beta_0) \right\}^{-1} n^{1/2} \left\{ \hat{U}(\beta_0) - \tilde{U}(\beta_0) \right\} + o_p(1) \]
\[ \Rightarrow o_p(1). \]
APPENDIX C

Estimation of Variances for the Refined Nonparametric Estimators

Let \( \hat{\beta}_{NP,1} \) be the refined type 1 nonparametric estimator and \( \hat{\beta}_{NP,2} \) be the refined type 2 nonparametric estimator. The variance of \( \hat{\beta}_{NP,1} \) can be estimated by

\[
n^{-1} \hat{A}_{NP}^{-1}(\hat{\beta}_{NP,1}, \hat{D}_{NP,1}) \hat{B}_{NP}(\hat{\beta}_{NP,1}, \hat{D}_{NP,1}) \{ \hat{A}_{NP}^{-1}(\hat{\beta}_{NP,1}, \hat{D}_{NP,1}) \}^T,
\]

where \( \hat{A}_{NP}(\beta, D) = \frac{\partial \hat{V}_2(\beta, D)}{\partial \beta} \), \( \hat{B}_{NP}(\beta, D) = n^{-1} \sum_{i=1}^n \hat{\omega}_{NP,i}(\beta, D) \hat{\omega}_{NP,i}^T(\beta, D) \), and

\[
\hat{\omega}_{NP,i}(\beta, D) = \int_0^L \left( \left[ A_i \{ W_{ij} + D(\beta) \} - \frac{S_1^*(u, \beta, W, D)}{S_0^*(u, \beta, W, D)} \right] dN_i(u) - \left\{ \frac{S_1^*(u, \beta, W, D)}{S_0^*(u, \beta, W, D)} - \frac{S_0^*(u, \beta, W, D) S_1^*(u, \beta, W, D)}{\{ S_0^*(u, \beta, W, D) \}^2} \right\} \right. \\
\left. + \left[ \frac{\partial}{\partial D_1} \left( \frac{S_1^*(u, \beta, W, D)}{S_0^*(u, \beta, W, D)} \right) - 1 \right] I(m_i \geq 2) Q(\beta_1, D_1) \right\} n^{-1} \sum_{j=1}^n dN_j(u),
\]

\[
Q(\beta_1, D_1) = A_2 \left\{ (W_{1ij} - W_{1ij' - D_1}) \frac{I(m_i \geq 2) \exp(\beta_1 W_{1ij})}{n^{-1} \sum_{i=1}^n I(m_i \geq 2) A_1 \exp(\beta_1 W_{1ij})} \right\}.
\]

The variance of \( \hat{\beta}_{NP,2} \) can be estimated by

\[
n^{-1} \hat{A}_{NP}(\hat{\beta}_{NP,2}, \hat{D}_{NP,1})^{-1} \hat{B}_{NP}(\hat{\beta}_{NP,2}, \hat{D}_{NP,1}) \{ \hat{A}_{NP}(\hat{\beta}_{NP,2}, \hat{D}_{NP,1}) \}^T,
\]

with \( Q(\beta_1, D_1) \) replaced by

\[
Q^*(\beta_1, D_1, u) = A_2 \left\{ (W_{1ij} - W_{1ij' - D_1}) \frac{I(m_i \geq 2) Y_i(u) \exp(\beta_1 W_{1ij})}{n^{-1} \sum_{i=1}^n I(m_i \geq 2) Y_i(u) A_1 \exp(\beta_1 W_{1ij})} \right\}.
\]

APPENDIX D

Existence of the Correction Estimators for Replication Data

Suppose there is only a single covariate \( X \). For each death time \( u_k \) \((k = 1, \ldots, R)\), let

\[
W_{max,k} = \max_{\{i,j:dN_i(u_k)=1\}} W_{ij}, \quad \bar{W}_{max,k} = \max_{\{i,j:dN_i(u_k)=1\}} A_i W_{ij}, \quad W_{max}^* = \max_{\{i,j:m_i \geq 2\}} W_{ij},
\]

\[
W_{max,i} = \max_j W_{ij}, \quad j_i = \arg \max_j W_{ij}, \quad \bar{W}_{max,i} = \max_j A_i W_{ij}.
\]

21
For the refined type 1 nonparametric correction approach,

\[
\hat{D}_+ \equiv \lim_{\beta \to +\infty} \hat{D}(\beta) = \lim_{\beta \to +\infty} \frac{\sum_{i=1}^n I(m_i \geq 2)A_2(W_{ij} - W_{ij}) \exp(\beta W_{ij})}{\sum_{i=1}^n I(m_i \geq 2)A_1 \exp(\beta W_{ij})}
\]

\[
= \lim_{\beta \to +\infty} \frac{\sum_{i=1}^n I(m_i \geq 2)A_2(W_{ij} - W_{ij}) \exp \{\beta (W_{ij} - W_{ij}^*)\}}{\sum_{i=1}^n I(m_i \geq 2)A_1 \exp \{\beta (W_{ij} - W_{ij}^*)\}}
\]

\[
= \frac{\sum_{i=1}^n I(m_i \geq 2)I(W_{ij}^* = W_{ij}^*) \sum_{j' \neq j} (W_{ij}^* - W_{ij'})}{\sum_{i=1}^n I(m_i \geq 2)I(W_{ij}^* = W_{ij}^*)} > 0.
\]

If \(dN_i(u_k) = 1\),

\[
\lim_{\beta \to +\infty} A_1 \left\{ W_{ij} + \hat{D}(\beta) \right\} = \frac{S_1'(u_k, \beta, W, \hat{D})}{S_0'(u_k, \beta, W, \hat{D})}
\]

\[
= A_1 W_{ij} + \hat{D}_+ + \lim_{\beta \to +\infty} \frac{\sum_{i'=1}^n A_1 Y_{ij}(u_k) \left\{ W_{ij'} + \hat{D}(\beta)dN_i(u_k) \right\} \exp \left[ \beta \left\{ W_{ij'} + \hat{D}(\beta)dN_i(u_k) \right\} \right]}{\sum_{i'=1}^n A_1 Y_{ij}(u_k) \exp \left[ \beta \left\{ W_{ij'} + \hat{D}(\beta)dN_i(u_k) \right\} \right]}
\]

\[
= A_1 W_{ij} + \hat{D}_+ - \left( W_{\max,k} + \hat{D}_+ \right) \leq 0,
\]

where "<" holds for those subjects with \(A_1 W_{ij} < W_{\max,k}\). Hence \(\lim_{\beta \to +\infty} \hat{U}(\beta, \hat{D}) < 0\).

With similar arguments, we can show that \(\lim_{\beta \to -\infty} \hat{U}(\beta, \hat{D}) > 0\).

The proof for the refined type 2 nonparametric correction approach is similar.

For the parametric correction methods, assume \(\lim_{\beta \to +\infty} \hat{D}(\beta) = +\infty\), \(\lim_{\beta \to -\infty} \hat{D}(\beta) = -\infty\).

For the standard type A parametric correction, let

\[
\tilde{S}_r(u, \beta, W) = \sum_{i=1}^n Y_i(u) \left\{ A_1 W_{ij} \right\}^r \exp \{\beta A_1 W_{ij}\},
\]

http://biostats.bepress.com/uwbiostat/paper226
Thus if \( dN_i(u_k) = 0 \), we have

\[
\lim_{\beta \to +\infty} \frac{\tilde{S}_1^\#(u, \beta, W)}{S_0^\#(u, \beta, W)} = \lim_{\beta \to +\infty} \frac{\sum_{i=1}^n Y_i(u) A_i W_{ij} \exp \{ \beta A_i W_{ij} \}}{\sum_{i=1}^n Y_i(u) \exp \{ \beta A_i W_{1ij} \}}
\]

\[
= \lim_{\beta \to +\infty} \frac{\sum_{i=1}^n Y_i(u) A_i W_{ij} \exp \{ \beta ( A_i W_{1ij} - \max_{i:Y_i(u)>0} W_{\max,i}) \}}{\sum_{i=1}^n Y_i(u) \exp \{ \beta ( A_i W_{1ij} - \max_{i:Y_i(u)>0} W_{\max,i}) \}}
\]

\[
= \max_{\{i:Y_i(u)>0\}} \frac{W_{\max,i}}{W_{\max,i}}.
\]

Thus if \( dN_i(u_k) = 1 \),

\[
\lim_{\beta \to +\infty} A_i W_{ij} + \hat{D}(\beta) - \frac{\tilde{S}_1^\#(u_k, \beta, W)}{S_0^\#(u_k, \beta, W)} = +\infty.
\]

Hence \( \lim_{\beta \to +\infty} \hat{U}(\beta, D) = +\infty \). Similarly, we can show that \( \lim_{\beta \to -\infty} \hat{U}(\beta, D) = -\infty \).

For the standard type B parametric correction,

\[
\lim_{\beta \to +\infty} \frac{\tilde{S}_1^\#(u, \beta, W)}{S_0^\#(u, \beta, W)} = \lim_{\beta \to +\infty} \frac{\sum_{i=1}^n A_i Y_i(u) W_{ij} \exp \{ \beta W_{ij} \}}{\sum_{i=1}^n A_i Y_i(u) \exp \{ \beta W_{ij} \}}
\]

\[
= \lim_{\beta \to +\infty} \frac{\sum_{i=1}^n A_i Y_i(u) W_{ij} \exp \{ \beta ( W_{ij} - \max_{i:Y_i(u)>0} W_{\max,i}^*) \}}{\sum_{i=1}^n A_i Y_i(u) \exp \{ \beta ( W_{ij} - \max_{i:Y_i(u)>0} W_{\max,i}^*) \}}
\]

\[
= \max_{\{i:Y_i(u)>0\}} \frac{W_{\max,i}^*}{W_{\max,i}^*}.
\]

Hence, with similar arguments as those for the standard type A parametric correction, we have \( \lim_{\beta \to +\infty} \hat{U}(\beta, D) = +\infty \). Similarly, we can show that \( \lim_{\beta \to -\infty} \hat{U}(\beta, D) = -\infty \).

For refined type A parametric correction, if \( dN_i(u_k) = 1 \), then

\[
\lim_{\beta \to +\infty} A_i W_{ij} + \hat{D}(\beta) - \frac{S_1^\#(u_k, \beta, W, D)}{S_0^\#(u_k, \beta, W, D)}
\]

\[
= \lim_{\beta \to +\infty} \left( A_i W_{ij} + \hat{D}(\beta) \right)
\]

\[
\frac{\sum_{i'Y_i'(u_k)A_1 \{W_{i'j} + \hat{D}(\beta)dN_{i'}(u_k)\} \exp \left[ \beta A_1 \left\{ W_{i'j} + \hat{D}(\beta)dN_{i'}(u_k) \right\} \right]}{\sum_{i'Y_i'(u_k)A_1 \{W_{i'j} + \hat{D}(\beta)dN_{i'}(u_k)\} \exp \left[ \beta A_1 \left\{ W_{i'j} + \hat{D}(\beta)dN_{i'}(u_k) \right\} \right]}.
\]

23
\[ \lim_{\beta \to +\infty} \left( A_1 W_{ij} + \hat{D}(\beta) \right) - \frac{\sum_{i'=1}^{n} Y_{i'}(u_k) \left\{ A_1 W_{i'j} + \hat{D}(\beta) dN_{i'}(u_k) \right\} \exp \left[ \beta \left\{ A_1 W_{i'j} + \hat{D}(\beta) dN_{i'}(u_k) - W_{\max,k} - \hat{D}(\beta) \right\} \right]}{\sum_{i'=1}^{n} A_1 Y_{i'}(u_k) \exp \left[ \beta \left\{ A_1 W_{i'j} + \hat{D}(\beta) dN_{i'}(u_k) - W_{\max,k} - \hat{D}(\beta) \right\} \right]} \]

\[ = \lim_{\beta \to +\infty} A_1 W_{ij} + \hat{D}(\beta) - W_{\max,k} - \hat{D}(\beta) \]

\[ = A_1 W_{ij} - W_{\max,k} \leq 0, \]

where "<" holds for those subjects with \( A_1 W_{ij} < W_{\max,k} \). Hence \( \lim_{\beta \to +\infty} \hat{U}(\beta, \hat{D}) \leq 0 \), with "<" holds if and only there exists two subjects failing at the same time with different \( A_1 W_{ij} \). With similar arguments we can show \( \lim_{\beta \to -\infty} \hat{U}(\beta, \hat{D}) \geq 0 \) with ">" holds under the same condition. If there are no tied failure times, \( dN_{i'}(u_k) = 0 \) for \( i \neq i' \). Together with \( \lim_{\beta \to +\infty} \hat{D}(\beta) = +\infty \), there exists \( b_2 \), such that \( A_1 \left\{ W_{ij} + \hat{D}(\beta) dN_{i}(u_k) \right\} \geq A_1 \left\{ W_{i'j} + \hat{D}(\beta) dN_{i'}(u_k) \right\} \) for \( \beta > b_2 \) and \( i \neq i' \). Hence

\[ S_1^# \left( u_k, \beta, W, \hat{D} \right) < A_1 \left\{ W_{ij} + \hat{D}(\beta) dN_{i}(u_k) \right\} S_0^# \left( u_k, \beta, W, \hat{D} \right) \]

for \( \beta > b_2 \), which implies \( \hat{U}(\beta, \hat{D}) > 0 \) for \( \beta > b_2 \). Similarly, we can show that there exists \( b_1 < b_2 \) such that \( \hat{U}(\beta, \hat{D}) < 0 \) for \( \beta < b_1 \).

For the refined type B parametric correction, the estimating equation is

\[ \hat{U}(\beta, \hat{D}) = n^{-1} \sum_{i=1}^{n} \int_{0}^{L} \left\{ A_1 W_{ij} + \hat{D}(\beta) - \frac{S_1^* \left( u, \beta, W, \hat{D} \right)}{S_0^* \left( u, \beta, W, \hat{D} \right)} \right\} dN_i(u), \]

If \( dN_i(u_k) = 1 \), then

\[ \lim_{\beta \to +\infty} A_1 W_{ij} + \hat{D}(\beta) - \frac{S_1^* \left( u_k, \beta, W, \hat{D} \right)}{S_0^* \left( u_k, \beta, W, \hat{D} \right)} = \lim_{\beta \to +\infty} \left( A_1 W_{ij} + \hat{D}(\beta) \right) \]
\[
\sum_{\ell'=1}^{n} A_{1\ell'}(u_k) \{ W_{\ell'j} + D(\beta)dN_{\ell'}(u_k) \} \exp \left[ \beta \{ W_{\ell'j} + D(\beta)dN_{\ell'}(u_k) \} \right] \]
\[
\sum_{\ell'=1}^{n} A_{1\ell'}(u_k) \exp \left[ \beta \{ W_{\ell'j} + D(\beta)dN_{\ell'}(u_k) \} \right]
\]
\[
= \lim_{\beta \to +\infty} \left( A_{1ij} + D(\beta) \right)
\[
- \sum_{\ell'=1}^{n} A_{1\ell'}(u_k) \{ W_{\ell'j} + D(\beta)dN_{\ell'}(u_k) \} \exp \left[ \beta \{ W_{\ell'j} + D(\beta)dN_{\ell'}(u_k) - W_{\text{max},k} - D(\beta) \} \right] \]
\[
\sum_{\ell'=1}^{n} A_{1\ell'}(u_k) \exp \left[ \beta \{ W_{\ell'j} + D(\beta)dN_{\ell'}(u_k) - W_{\text{max},k} - D(\beta) \} \right]
\]
\[
= \lim_{\beta \to +\infty} A_{1ij} + D(\beta) - W_{\text{max},k} - D(\beta)
\[
= A_{1ij} - W_{\text{max},k}
\]
\[
\leq 0,
\]
where “<” holds for those subjects with \( A_{1ij} < W_{\text{max},k} \). Hence \( \lim_{\beta \to +\infty} \hat{U}(\beta, \hat{D}) < 0 \).

With similar arguments we can show \( \lim_{\beta \to -\infty} \hat{U}(\beta, \hat{D}) > 0 \).
Table 1
Simulation results in the case of normal error with $\Sigma$ known. I, “ideal” method; CS, conditional score; PC, first-order parametric correction; PC2, second-order parametric correction

<table>
<thead>
<tr>
<th>$n$</th>
<th>I</th>
<th>CS</th>
<th>PC</th>
<th>PC2</th>
<th>CS</th>
<th>PC</th>
<th>PC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>Est</td>
<td>-1.005</td>
<td>-1.021</td>
<td>-1.023</td>
<td>-1.028</td>
<td>-1.077</td>
<td>-0.304</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.109</td>
<td>0.177</td>
<td>0.465</td>
<td>0.187</td>
<td>0.354</td>
<td>2.338</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.110</td>
<td>0.160</td>
<td>0.169</td>
<td>0.163</td>
<td>0.243</td>
<td>0.290</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.955</td>
<td>0.938</td>
<td>0.951</td>
<td>0.941</td>
<td>0.935</td>
<td>0.868</td>
</tr>
<tr>
<td>600</td>
<td>Est</td>
<td>-1.006</td>
<td>-1.027</td>
<td>-1.035</td>
<td>-1.030</td>
<td>-1.065</td>
<td>-0.866</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.077</td>
<td>0.128</td>
<td>0.132</td>
<td>0.130</td>
<td>0.221</td>
<td>1.294</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.077</td>
<td>0.115</td>
<td>0.118</td>
<td>0.116</td>
<td>0.174</td>
<td>0.197</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.962</td>
<td>0.955</td>
<td>0.950</td>
<td>0.951</td>
<td>0.955</td>
<td>0.951</td>
</tr>
</tbody>
</table>
Table 2
Simulation results for replication data with skewed bimodal mixture of normals error. I, “ideal” method, RPA, refined type A parametric correction; SPA, standard type A parametric correction; RPB, refined type B parametric correction; SPB, standard type B parametric correction; RNP1, refined type 1 nonparametric correction; SNP1, standard type 1 nonparametric correction; RNP2, refined type 2 nonparametric correction; SNP2, standard type 2 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; NC, proportion of simulated data sets with no finite solutions found.

<p>| n    | Est | RPA | SPA | RPB | SPB | RNP1 | SNP1 | RNP2 | SNP2 | RPA | SPA | RPB | SPB | RNP1 | SNP1 | RNP2 | SNP2 |
|------|-----|-----|-----|-----|-----|------|------|------|------|------|-----|-----|-----|-----|------|------|------|------|
| 300  | 1.013 | -1.033 | -1.049 | -1.020 | -1.008 | -1.030 | -1.093 | -1.013 | -1.050 | -1.060 | -0.743 | -1.014 | -0.195 | -1.011 | -1.219 | -0.990 | -1.130 |
| SD   | 0.113 | 0.176 | 0.188 | 0.174 | 0.797 | 0.205 | 0.512 | 0.163 | 0.192 | 0.256 | 1.807 | 0.219 | 2.311 | 0.226 | 1.971 | 0.189 | 0.953 |
| SE   | 0.011 | 0.159 | 0.168 | 0.152 | 0.185 | 0.170 | 0.295 | 0.153 | 0.176 | 0.217 | 0.305 | 0.183 | 0.351 | 0.189 | 1.766 | 0.181 | 0.372 |
| CP   | 0.946 | 0.955 | 0.965 | 0.929 | 0.963 | 0.938 | 0.944 | 0.945 | 0.966 | 0.952 | 0.940 | 0.887 | 0.832 | 0.896 | 0.872 | 0.926 | 0.930 |
| NC   | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.018 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.079 | 0.000 | 0.079 | 0.037 |
| 600  | -1.001 | -1.005 | -1.011 | -0.999 | -1.014 | -1.007 | -1.030 | -0.996 | -1.011 | -1.013 | -1.033 | -0.996 | -0.959 | -1.006 | -1.034 | -0.983 | -1.072 |
| SD   | 0.078 | 0.112 | 0.114 | 0.110 | 0.118 | 0.145 | 0.459 | 0.110 | 0.118 | 0.153 | 0.169 | 0.150 | 0.807 | 0.187 | 0.608 | 0.137 | 1.424 |
| SE   | 0.077 | 0.109 | 0.112 | 0.109 | 0.115 | 0.125 | 0.155 | 0.109 | 0.115 | 0.147 | 0.160 | 0.142 | 0.198 | 0.162 | 0.255 | 0.139 | 0.203 |
| CP   | 0.958 | 0.956 | 0.957 | 0.954 | 0.958 | 0.941 | 0.945 | 0.951 | 0.956 | 0.952 | 0.962 | 0.938 | 0.959 | 0.895 | 0.890 | 0.944 | 0.957 |
| NC   | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.011 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.047 | 0.000 | 0.011 |</p>
<table>
<thead>
<tr>
<th></th>
<th>log(CD4)</th>
<th></th>
<th>treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>SE</td>
<td>Est</td>
</tr>
<tr>
<td>RPA</td>
<td>−2.267</td>
<td>0.418</td>
<td>−0.574</td>
</tr>
<tr>
<td>SPA</td>
<td>−2.268</td>
<td>0.419</td>
<td>−0.574</td>
</tr>
<tr>
<td>RPB</td>
<td>−2.135</td>
<td>0.373</td>
<td>−0.564</td>
</tr>
<tr>
<td>SPB</td>
<td>−2.137</td>
<td>0.374</td>
<td>−0.565</td>
</tr>
<tr>
<td>RNP1</td>
<td>−2.307</td>
<td>0.444</td>
<td>−0.564</td>
</tr>
<tr>
<td>SNP1</td>
<td>−2.311</td>
<td>0.447</td>
<td>−0.564</td>
</tr>
<tr>
<td>RNP2</td>
<td>−2.245</td>
<td>0.423</td>
<td>−0.564</td>
</tr>
<tr>
<td>SNP2</td>
<td>−2.248</td>
<td>0.425</td>
<td>−0.565</td>
</tr>
</tbody>
</table>

Table 3
Results for ACTG 175 data. RPA, refined type A parametric correction; SPA, standard type A parametric correction; RPB, refined type B parametric correction; SPB, standard type B parametric correction; RNP1, refined type 1 nonparametric correction; SNP1, standard type 1 nonparametric correction; RNP2, refined type 2 nonparametric correction; SNP2, standard type 2 nonparametric correction.
Figure 1. Scatter plots of the conditional score (CS) estimates, the first-order (PC) and second-order (PC2) parametric correction estimates versus the “ideal” estimates in the case of normal error with $\Sigma$ known.

(a) $n = 300, \sigma^2 = 0.3$

(b) $n = 300, \sigma^2 = 0.6$

(c) $n = 600, \sigma^2 = 0.3$

(d) $n = 600, \sigma^2 = 0.6$
Figure 2. Q-Q plots of the “ideal” (I) estimates, the conditional score (CS) estimates and the first-order (PC) and second-order (PC2) parametric correction estimates relative to the normal distribution in the case of normal error with $\Sigma$ known.
Figure 3. Typical score plots for two simulation data sets in the case of normal error with $D(\cdot)$ known. I, “ideal” method; CS, conditional score; PC, first-order parametric correction; PC2, second-order parametric correction.
Figure 4. Scatter plots of refined estimates and the corresponding standard estimates versus the “ideal” estimates in the case of normal error with replication data. RPA, refined type A parametric correction; SPA, standard type A parametric correction; RPB, refined type B parametric correction; SPB, standard type B parametric correction; RNP1, refined type 1 nonparametric correction; SNP1, standard type 1 nonparametric correction; RNP2, refined type 2 nonparametric correction; SNP2, standard type 2 nonparametric correction.
Figure 5. Typical score plots for two simulation data sets with replication data. Left panels are for the parametric approaches and right panels are for the nonparametric approaches. I, “ideal” method; RPA, refined type A parametric correction; SPA, standard type A parametric correction; RPB, refined type B parametric correction; SPB, standard type B parametric correction; RNP1, refined type 1 nonparametric correction; SNP1, standard type 1 nonparametric correction; RNP2, refined type 2 nonparametric correction; SNP2, standard type 2 nonparametric correction.
Figure 6. Q-Q plot of the deviations from the average of log(CD4).
Figure 7. Scatter plots of refined estimates versus the corresponding standard estimates for subsamples of the ACTG data. RPA, refined type A parametric correction; SPA, standard type A parametric correction; RPB, refined type B parametric correction; SPB, standard type B parametric correction; RNP1, refined type 1 nonparametric correction; SNP1, standard type 1 nonparametric correction; RNP2, refined type 2 nonparametric correction; SNP2, standard type 2 nonparametric correction.