



Published in final edited form as:

J Am Stat Assoc. 2012 March 1; 107(497): 318–330. doi:10.1080/01621459.2012.656021.

Estimating Regression Parameters in an Extended Proportional Odds Model

Ying Qing Chen,

Full Member, Vaccine and Infectious Disease and, Fred Hutchinson Cancer Research Center, Seattle, WA 98109

Nan Hu,

Assistant Professor, Department of Internal Medicine, School of Medicine, University of Utah, Salt Lake City, UT 84132

Su-Chun Cheng,

Senior Research Scientist, Department of Biostatistics, Harvard University, Boston, MA 02115

Philippa Musoke, and

Associate Professor, Department of Paediatrics and Child Health, Makerere University, Kampala, Uganda

Lue Ping Zhao*

Full Member, Program in Biostatistics and Biomathematics, Fred Hutchinson Cancer Research Center, Seattle, WA 98109

Ying Qing Chen: yqchen@fhcrc.org; Nan Hu: nan.hu@hci.utah.edu; Su-Chun Cheng: scheng@jimmy.harvard.edu; Philippa Musoke: pmusoke@mujhu.org; Lue Ping Zhao: lzhaol@fhcrc.org

Abstract

The proportional odds model may serve as a useful alternative to the Cox proportional hazards model to study association between covariates and their survival functions in medical studies. In this article, we study an extended proportional odds model that incorporates the so-called “external” time-varying covariates. In the extended model, regression parameters have a direct interpretation of comparing survival functions, without specifying the baseline survival odds function. Semiparametric and maximum likelihood estimation procedures are proposed to estimate the extended model. Our methods are demonstrated by Monte-Carlo simulations, and applied to a landmark randomized clinical trial of a short course Nevirapine (NVP) for mother-to-child transmission (MTCT) of human immunodeficiency virus type-1 (HIV-1). Additional application includes analysis of the well-known Veterans Administration (VA) Lung Cancer Trial.

Keywords

Counting process; Estimating function; HIV/AIDS; Maximum likelihood estimation; Semiparametric model; Time-varying covariate

*Chen’s research was partially supported by grants from the National Institute of Allergy and Infectious Diseases (NIH/NIAID R01 AI089341, R01 AI78835) and the National Cancer Institute (NIH/NCI P01 CA053996), Hu’s research was partially supported by Huntsman Cancer Institute Support Grant (NIH/NCI P30 CA042014) and Huntsman Cancer Foundation, and Zhao’s research was partially supported by grants from the National Institute of Mental Health (NIH/NIMH R01 MH 084621) and the National Cancer Institute (NIH/NCI R01 CA 119225). The authors thank the editor, the associate editor, and two referees for their valuable comments that helped improve a previous version of the manuscript.

1. INTRODUCTION

Between November, 1997 and January, 2001, a landmark HIV/AIDS randomized prevention trial, namely the HIVNET 012, was conducted to evaluate treatment efficacy and safety of a short course NVP versus a short course zidovudine (AZT) in prevention of MTCT of HIV-1 among infants born to HIV-1 infected pregnant women in Uganda (Jackson et al., 2003). One important secondary objective of this trial is to assess infant survival between treatment arms (The HIVNET/HPTN Group, 2003). In Figure 1, we plot the Kaplan-Meier estimates of infant survival in log-log scale for both of the two treatment arms. It appears that NVP improves infant survival over time.

To estimate a treatment effect, such as comparing infant survival between the NVP and the AZT, the semiparametric Cox proportional hazards model (Cox, 1972) is often used:

$$\lambda(t|Z^*) = \lambda_0(t) \exp(\beta Z^*), \quad (1)$$

where Z^* is the treatment indicator of NVP ($Z^* = 1$) versus AZT ($Z^* = 0$), $\lambda(\cdot|Z^*)$ is the hazard function of Z^* , and $\lambda_0(\cdot)$ is the unspecified baseline hazard function of $Z^* = 0$. Here, $\beta \in \mathbb{R} \subset \mathbb{R}$ is the regression parameter to measure the treatment effect on hazard functions of infant survival. For the HIVNET 012, the partial-likelihood estimate of $\hat{\beta}$ is -0.301 (*s.e.* = 0.236). It means that NVP would reduce the hazard by 26.0% ($(1 - e^{-0.301}) \times 100\%$), implying a longer infant survival compared with AZT, although such an improvement is not statistically significant ($p > 0.20$).

A useful alternative to the Cox model is the so-called proportional odds model, which models survival functions directly (Bennett, 1983; Pettitt, 1984):

$$\log \frac{S(t|Z^*)}{1 - S(t|Z^*)} = \log \frac{S_0(t)}{1 - S_0(t)} + \beta Z^*, \quad (2)$$

where $S(\cdot|Z^*)$ is the survival function of Z^* and $S_0(\cdot)$ is the baseline survival function of $Z^* = 0$. In this model, the regression parameter β has an interpretation of survival odds ratio. In practice, such an interpretation in survival functions has a direct appeal to measuring disease risks in clinical and epidemiological studies. For example, a generalized Kolmogorov-Smirnov test procedure in Fleming, O'Fallen and O'Brien (1980) was proposed to compare survival functions via Nelson-Aalen estimates (Nelson, 1969). Additional examples of directly modeling survival functions include McCullagh (1980), Jung (1996), and a recent work of Peng and Huang (2007).

In the literature, statistical inferences for the regression parameter of interest, β , in model (2) have been extensively developed. For example, Murphy, Rossini and van der Vaart (1997) studied a maximum likelihood method based on profile likelihood; Yang and Prentice (1999) used estimating equations based on the weighted Nelson-Aalen estimates of baseline survival functions. Other works include Dabrowska and Doksum (1988), Scharfstein et al. (1998) and Zhang and Davidian (2007). Bayesian methods have also been developed (Hanson and Yang, 2007). More notably, some of the works, including Cheng, Wei and Ying (1995), Chen, Jin and Ying (2002), Kosorok, Lee and Fine (2004) and Zeng and Lin (2006, 2007), have studied model (2) under the framework of transformation models. For example, Chen, Jin and Ying (2002) developed simple martingale-based estimating equations to estimate the regression parameter, and Zeng and Lin (2006, 2007) applied a nonparametric maximum likelihood estimation (NPMLE) method to estimate the regression parameter efficiently. More references can also be found therein.

In reality, covariates are not necessarily limited to be time-independent. In the HIVNET 012, for example, follow-up information, such as maternal CD4+ counts and maternal HIV-RNA viral loads (VL) that may affect infant survival through late postnatal breastfeeding, has been collected. Other time-varying covariates, such as treatment-time interaction, can also be of practical interest, since a hypothesis testing on treatment-time interaction may serve the purpose to check the adequacy of model (2). Nevertheless, these time-varying covariates, along with the assigned treatment and other baseline covariates, may all be potentially associated with infant survival. In this paper, we hence aim to extend the proportional odds model (2) to include such time-varying covariates and develop appropriate inference procedures for the regression parameters in analysis of many datasets similar to the HIVNET 012.

We organize the rest of this article as follows. In §2, we propose an extended proportional odds model and examine a few of its properties. In §3, we present two methods, including a simple quasi partial score estimating equation approach, and a maximum likelihood estimation approach, for model inferences. In §4, we present results from our numerical studies, including Monte-Carlo simulations and methods application to the motivating HIVNET 012 Trial and the well-known Veterans Administration (VA) Lung Cancer Trial. More discussion on the extended proportional odds model is in §5. Technical proof is collected in the Appendix.

2. THE EXTENDED PROPORTIONAL ODDS MODEL

The proportional odds model (2) can be considered as a natural extension of the logistic regression model of binary outcomes to time-to-event outcomes. Suppose that T represents a non-negative time-to-event outcome. At a given time $t > 0$, denote the binary survival indicator of T by $D(t) = I(T > t)$. For Z^* , a usual logistic regression model of $D(t)$ would then assume that

$$\text{logit } \text{pr}\{D(t)=1|Z^*\} = \text{log} \frac{\text{pr}\{D(t)=1|Z^*\}}{1 - \text{pr}\{D(t)=1|Z^*\}} = \alpha(t) + \beta(t)Z^*, \tag{3}$$

where $\alpha(t)$ and $\beta(t)$ are both parameters at $t > 0$. When $\beta(\cdot)$ is assumed constant over time, i.e., $\beta(t) \equiv \beta$, then model (3) becomes the proportional odds model (2),

$$\text{logit } S(t|Z^*) = \alpha(t) + \beta Z^*, \tag{4}$$

with $\alpha(t) = \text{log}[S_0(t)/(1 - S_0(t))]$. When $\alpha(\cdot)$ is unspecified, model (4) is semiparametric.

As discussed earlier, it is of practical interest to include time-varying covariates in the proportional odds model (2), such as maternal CD4+ counts, maternal plasma HIV-RNA VL, or treatment-time interaction in the HIVNET 012 example. Let $Z^*(t)$ be the p -dimensional covariates that may include both baseline and time-varying covariates measured at t . To make $Z^*(\cdot)$'s association with the survival function $S(\cdot)$ meaningful, we consider the so-called “external” covariates, because $\text{pr}\{T > t | Z^*(t)\} \equiv 1$ for any “internal” covariates provided that $Z^*(\cdot)$ is not of the baseline level. Formally, the external time-varying covariates shall satisfy that

$$\text{pr}\{Z^*(s)|Z^*(t), T \geq t\} = \text{pr}\{Z^*(s)|Z^*(t), T = t\}, \text{ for any } s \geq t,$$

as in Kalbfleisch and Prentice (2002, pp. 196–198). Intuitively speaking, this condition essentially requires that the occurrence of a failure event at t should not affect the future path of $\mathbf{Z}^*(\cdot)$ up to anytime beyond t . This means, in the HIVNET 012 example, whether an infant survives or dies at t shall not affect the mother's future CD4+ counts or viral loads after t .

Denote $\tilde{\mathbf{Z}}(t)$ the covariate history of $\{\mathbf{Z}^*(s), 0 \leq s \leq t\}$ up to t and $\tilde{\mathbf{Z}} = \tilde{\mathbf{Z}}(\infty)$. We consider an extended proportional odds model to include the time-varying covariates $\mathbf{Z}^*(\cdot)$, as in

$$\text{logit } S(t|\tilde{\mathbf{Z}}) = -\log R(t) + \beta^T \mathbf{Z}(t), \quad (5)$$

where $\mathbf{Z}(t) = \int_0^t \mathbf{Z}^*(u) d\varphi(u, t)$ is a weighted average of $\mathbf{Z}^*(\cdot)$ on $[0, t]$, and $\varphi(u; t)$ is the pre-specified weight function such that $\int_0^t d\varphi(u; t) = 1$. Here, $R(t) = \exp\{-\alpha(t)\}$ is the baseline odds function when $\mathbf{Z}(\cdot) \equiv 0$, and the superscript T defines matrix (vector) transpose.

Apparently in this extended model, with different choices of weight function $\varphi(u; t)$, $\mathbf{Z}(t)$ may cover a wide range of covariate types in practice. Examples include:

1. $\varphi(u; t) = I(u = 0)$. This means that $\mathbf{Z}(t) = \mathbf{Z}(0)$ are for time-independent baseline covariates only, and the extended model reduces to model (2);
2. $\varphi(u; t) = I(u = t)$. This means that $\mathbf{Z}(t) = \mathbf{Z}^*(t)$ are the value of covariates measured at time t ;
3. $\varphi(u; t) = u/t$. This means that $\mathbf{Z}(t) = 1/t \int_0^t \mathbf{Z}^*(u) du$ are the average value of $\mathbf{Z}^*(\cdot)$ on $[0, t]$;
4. Suppose that $\mathbf{Z}^*(\cdot)$ are only measured at a finite number of distinct time points, $t_{(0)} < \dots < t_{(j-1)} < t_{(j)} < \dots < t_{(k)}$. Let $\varphi(u; t) = \sum_{j=1}^k I\{t_{(j-1)} \leq u = t < t_{(j)}\}$. Then $\mathbf{Z}(t) = \int_0^t \mathbf{Z}^*(u) d\varphi(u; t)$ are the so-called last-observation-carry-forward covariates.

Whatever $\varphi(u; \cdot)$ is chosen in $\mathbf{Z}(\cdot)$, β 's interpretation is straightforward: it is the odds ratio of survival probabilities associated with per unit change in $\mathbf{Z}(\cdot)$. When $\beta > 0$ (< 0), it means that larger $\mathbf{Z}(\cdot)$ are associated with increased (decreased) survival and hence decreased (increased) risk; when $\beta = 0$, there is no survival risk associated with $\mathbf{Z}(\cdot)$.

Two properties of model (5) are summarized as follows:

Property 1

Under the assumed extended proportional odds model (5),

- a. when $\mathbf{Z}(t) \equiv \mathbf{Z}(0)$ for any $t > 0$, the family of log-logistic survival distributions $\mathcal{S} = \{S(\cdot; \theta); S(t; \theta) = \theta_1 t^{-\theta_2} (1 + \theta_1 t^{-\theta_2})^{-1}, \theta_1, \theta_2 > 0\}$ is closed for model (5). That is, if a baseline survival function $S_0 \in \mathcal{S}$ and model (5) holds, then $S(\cdot|\tilde{\mathbf{Z}}) \in \mathcal{S}$;
- b. let $\lambda_0(\cdot)$ be the baseline hazard function of $\mathbf{Z}(\cdot) \equiv 0$ and $\lambda(\cdot|\tilde{\mathbf{Z}})$ be the hazard function of $\tilde{\mathbf{Z}}$, then $\lim_{t \rightarrow 0} \lambda(t|\tilde{\mathbf{Z}})/\lambda_0(t) = \exp\{\beta^T \mathbf{Z}^*(0)\}$, and $\lim_{t \rightarrow \infty} \lambda(t|\tilde{\mathbf{Z}})/\lambda_0(t) = 1$, when $\mathbf{Z}(\cdot)$ have bounded variations.

Proof of these two properties is relatively straightforward. For Property 1a, it is easy to verify that, if $S_0(t) = \theta_1 t^{-\theta_2} (1 + \theta_1 t^{-\theta_2})^{-1}$, then

$$S(t | \tilde{\mathbf{Z}}) = \frac{\theta_1 \exp\{\boldsymbol{\beta}^T \mathbf{Z}(0)\} t^{-\theta_2}}{1 + \theta_1 \exp\{\boldsymbol{\beta}^T \mathbf{Z}(0)\} t^{-\theta_2}} \in \mathcal{S},$$

under model (4). Property 1b follows an application of the usual l'Hôpital's rule.

These two properties provide some basic characterization of the extended proportional odds model. Property 1a shows that the regression parameter of the proportional odds model (2) would only affect the scale parameter in baseline log-logistic distribution functions, in the same way that the proportional hazards model would do to baseline Weibull distribution functions. Property 1b shows the so-called "converging-hazards" feature in the extended proportional odds model: under the extended model, hazard functions differ by a ratio of $\boldsymbol{\beta}$ at the initial time zero, and then converge to each other as time increases. This converging-hazards feature has been well recognized (Murphy, Rossini and van der Vaart, 1997). It can be particularly useful if the assumption of constant proportionality seems restrictive for the Cox model, as recommended by Yang and Prentice (1999).

3. INFERENCE PROCEDURES

Consider that study data are collected in a typical setting of censored time-to-event outcomes with time-varying covariates. Suppose that there are n subjects in the dataset, and let T_i and C_i be the failure and censoring times, respectively, for $i = 1, 2, \dots, n$. Given $\tilde{\mathbf{Z}}_i$, T_i and C_i are assumed to be independent. An observed dataset usually consists of n independent and identically distributed (iid) copies of $\{(X_i, \Delta_i, \tilde{\mathbf{Z}}_i(X_i))\}; i = 1, 2, \dots, n\}$, where $X_i = \min(T_i, C_i)$, $\Delta_i = I(T_i < C_i)$, and $\tilde{\mathbf{Z}}_i(X_i)$ is the covariate history of the i th subject up to X_i . Let the true values of $\boldsymbol{\beta}$ and $R(\cdot)$ be $\boldsymbol{\beta}_0$ and $R_0(\cdot)$, respectively, in the extended model (5). Subject-specific subscripts may be occasionally suppressed for their general use.

3.1 Method of moment-based estimating equations

We first consider a moment-based semiparametric estimation procedure to estimate the regression parameters in the extended model (5). Our approach is to develop a closed-form estimator for the baseline odds function $R(\cdot)$, based on a martingale representation similar to that in Chen and Jewell (2001) and Chen, Jin and Ying (2002) at the true $\boldsymbol{\beta}_0$, and then develop a class of log-rank type of estimating equations to estimate $\boldsymbol{\beta}$ in (5).

Consider $N_i(t) = I(X_i \geq t, \Delta_i = 1)$ and $Y_i(t) = I(X_i \geq t)$, and denote a filtration by $\mathcal{F}_t = \sigma\{N_i(u), Y_i(u), \tilde{\mathbf{Z}}_i; 0 \leq u \leq t, i = 1, 2, \dots, n\}$. Then we know that

$$E\{dN_i(t) | \mathcal{F}_{t-}; \boldsymbol{\beta}_0, R_0\} = \frac{Y_i(t)}{B_i(t; \boldsymbol{\beta}_0) + R_0(t)} \{dR_0(t) - R_0(t) d \log B_i(t; \boldsymbol{\beta}_0)\} \tag{6}$$

where $B_i(t; \boldsymbol{\beta}) = \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(t)\}$. Furthermore, let

$M_i(t; \boldsymbol{\beta}, R) = N_i(t) - \int_0^t Y_i(u) \{B_i(u; \boldsymbol{\beta}) + R(u)\}^{-1} \{dR(u) - R(u) d \log B_i(u; \boldsymbol{\beta})\}$ for $i = 1, 2, \dots, n$. Then $\{M_i(t; \boldsymbol{\beta}_0, R_0); t \geq 0\}$ are martingales with respect to the filtration \mathcal{F}_t . In addition,

$$\sum_{i=1}^n \{B_i(t; \boldsymbol{\beta}_0) + R_0(t)\} dN_i(t) - \sum_{i=1}^n Y_i(t) \{dR_0(t) - R_0(t) d \log B_i(t; \boldsymbol{\beta}_0)\} = \sum_{i=1}^n \{B_i(t; \boldsymbol{\beta}_0) + R_0(t)\} dM_i(t; \boldsymbol{\beta}_0, R_0).$$

Since $E[\{B_i(t; \beta) + R_0(t)\} dM_i(t; \beta, R_0)] = 0$ by equation (6), we thus consider the following estimating equation of $R(t)$ for any t such that $\sum_j Y_j(t) > 0$, as if β were known:

$$\sum_{i=1}^n [\{B_i(t; \beta) + R(t)\} dN_i(t) - Y_i(t) \{dR(t) - R(t) d \log B_i(t; \beta)\}] = 0. \tag{8}$$

This estimating equation is indeed a first-order stochastic differential equation of $R(\cdot)$, i.e.,

$$dR(t) + R(t) d \log P_n(t; \beta) = dQ_n(t; \beta), \tag{9}$$

where

$$P_n(t; \beta) = \exp \left[- \int_0^t \sum_{i=1}^n \{dN_i(u) + Y_i(u) d \log B_i(u; \beta)\} / \sum_{i=1}^n Y_i(u) \right], \text{ and}$$

$$Q_n(t; \beta) = \int_0^t \sum_{i=1}^n B_i(u; \beta) dN_i(u) / \sum_{i=1}^n Y_i(u).$$

Thus, we solve (9) to obtain a closed-form solution for an estimator of $R(\cdot)$,

$$\widehat{R}_n(t; \beta) = P_n(t; \beta)^{-1} \int_0^t P_n(u - ; \beta) dQ_n(u; \beta). \tag{10}$$

Here, $P_n(t-)$ denotes the left-continuous version of $P_n(t)$. Comparing equations (7) and (8), it is true that

$$\widehat{R}_n(t; \beta_0) - R_0(t) = - P_n(t; \beta_0)^{-1} \sum_{i=1}^n \int_0^t \frac{P_n(u - ; \beta_0) \{B_i(u; \beta_0) + R_0(u)\}}{\sum_j Y_j(u)} dM_i(u; \beta_0, R_0). \tag{11}$$

Consider that $0 < \tau = \inf\{t: \text{pr}(X > t) = 0\} < \infty$. We have the following asymptotic properties for this estimator of $R(\cdot)$:

Lemma 2—For a fixed constant $t \in [0, \tau)$, under the conditions specified in the Appendix, as $n \rightarrow \infty$,

- a. $\widehat{R}_n(\cdot; \beta_0)$ is consistent on $[0, t]$ almost surely. That is, $\|\widehat{R}_n(\beta_0) - R_0\| = \sup_{u \in [0, t]} |\widehat{R}_n(u; \beta_0) - R_0(u)|$ converges to 0 almost surely.
- b. $n^{1/2} \{\widehat{R}_n(t; \beta_0) - R_0(t)\}$ is asymptotically mean-zero normal.

Proof of this Lemma follows the standard martingale theory of counting processes. When the covariates are all time-independent, $\widehat{R}_n(\cdot)$ in (10) is equivalent to the ones in Yang and Prentice (1999) for estimating the baseline odds function, except that $P_n(\cdot)$ in (10) is a Nelson-Aalen type of estimator, instead of the Kaplan-Meier estimator. Nevertheless, $\widehat{R}_n(\cdot)$ in (10) maintains the same meaning of weighted empirical odds functions even in the presence of time-varying covariates.

In order to estimate the regression parameter β in model (5), we consider the following estimating functions:

$$S_n(\beta, R_0) = \sum_{i=1}^n \int_0^\tau Z_i(t) \{ [B_i(t) + R_0(t; \beta)] dN_i(t) - Y_i(t) \{ dR_0(t) - R_0(t) d \log B_i(t; \beta) \} \},$$

They are unbiased at the true parameter β_0 , given the fact of (6). By replacing $R_0(\cdot)$ with its estimator $\widehat{R}_n(\cdot; \beta)$ and some algebra shown in the Appendix, we obtain these estimating equations for β_0 :

$$S_n(\beta) = \sum_{i=1}^n \int_0^\tau \{ Z_i(t) - \overline{Z}(t) \} \left[\{ B_i(t; \beta) + \widehat{R}_n(t; \beta) \} dN_i(t) + Y_i(t) \widehat{R}_n(t; \beta) d \log B_i(t; \beta) \right],$$

where $\overline{Z}(t) = \sum_j Y_j(t) Z_j(t) / \sum_j Y_j(t)$. Apparently, $S_n(\beta)$ are not necessarily continuous or monotone, particularly when $Z(\cdot)$ are not differentiable. Hence, we consider the parameter estimator $\hat{\beta}_n$ to satisfy that $S_n(\hat{\beta}_n) = 0$, as defined for the estimating equations of rank-type in Tsiatis (1990).

The following theorem summarizes the asymptotic properties of $\hat{\beta}_n$ that can be used in statistical inferences of β_0 :

Theorem 3—Under the conditions specified in the Appendix, there exists a neighborhood \mathcal{N} containing β_0 as an interior point such that, as $n \rightarrow \infty$,

- a. $\hat{\beta}_n \in \mathcal{N}$ is strongly consistent;
- b. $n^{1/2}(\hat{\beta}_n - \beta_0)$ converges weakly to a mean-zero normal variate with variance-covariance $U^{-1} V (U^{-1})^T$, where

$$U = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \int_0^\tau \{ Z_i(t) - \overline{Z}(t) \} \left[\left\{ \frac{\partial \widehat{R}_n(t; \beta_0)}{\partial \beta} + e^{\beta_0^T Z_i(t)} Z_i(t) \right\} dN_i(t) + Y_i(t) \left\{ \frac{\partial \widehat{R}_n(t; \beta_0)}{\partial \beta} \beta_0^T + R_0(t) \right\} dZ_i(t) \right]^T,$$

$$V = \lim_{n \rightarrow \infty} \int_0^\tau \frac{1}{n} \sum_{i=1}^n \frac{Y_i(t) \xi_i(t)^{\otimes 2} \{ dR_0(t) - R_0(t) d \log B_i(t; \beta_0) \}}{B_i(t; \beta_0) + R_0(t)},$$

and

$$\xi_i(t; \beta) = \{ Z_i(t) - \overline{Z}(t) \} \{ B_i(t; \beta) + R_0(t) \} - \frac{P_n(t; \beta) \{ B_i(t; \beta) + R_0(t) \}}{\sum_j Y_j(t)} \left[\sum_{k=1}^n \int_0^\tau \frac{Z_k(u) - \overline{Z}(u)}{P_n(u; \beta)} Y_k(u) d \log \{ R_0(u) + B_k(u; \beta) \} \right];$$

- c. U and V can be consistently estimated by

$$\widehat{U}_n = n^{-1} \sum_{i=1}^n \int_0^\tau \{ Z_i(t) - \overline{Z}(t) \} \left[\left\{ \frac{\partial \widehat{R}_n(t; \widehat{\beta}_n)}{\partial \beta} + e^{\widehat{\beta}_n^T Z_i(t)} Z_i(t) \right\} dN_i(t) + Y_i(t) \left\{ \frac{\partial \widehat{R}_n(t; \widehat{\beta}_n)}{\partial \beta} \widehat{\beta}_n^T + \widehat{R}_n(t; \widehat{\beta}_n) \right\} dZ_i(t) \right]^T,$$

and

$$\widehat{V}_n = n^{-1} \sum_{i=1}^n \int_0^\tau \frac{Y_i(t) \widehat{\xi}_i(t)^{\otimes 2} \{ d\widehat{R}_n(t; \widehat{\beta}_n) - \widehat{R}_n(t; \widehat{\beta}_n) d \log B_i(t; \widehat{\beta}_n) \}}{B_i(t; \widehat{\beta}_n) + \widehat{R}_n(t; \widehat{\beta}_n)},$$

respectively, where $\hat{\xi}_A(t) = \xi_A(t; \hat{\beta}_n, \hat{R}_n)$.

Proof of this theorem is outlined in the Appendix. According to this theorem, the proposed moment-based estimating equations yield a usual sandwich-type of variance estimate for $\hat{\beta}_n$.

In addition, user-defined weight functions can be included in $S_n(\cdot)$ to obtain a class of weighted estimators of β :

$$\sum_{i=1}^n \int_0^\tau W_n(t) \{Z_i(t) - \bar{Z}(t)\} \left[\{B_i(t; \beta) + \widehat{R}_n(t; \beta)\} dN_i(t) + Y_i(t) \widehat{R}_n(t; \beta) d\log B_i(t; \beta) \right] = 0, \quad (12)$$

where $W_n(\cdot)$ are \mathcal{F}_t -predictable weight functions converging to a deterministic function of $w(\cdot)$. For example, a popular choice of weight functions can be a Prentice-Wilcoxon-type of $W(t) = \widehat{S}_{KM}(t-)$, which is a left-continuous version of the Kaplan-Meier estimate of the baseline survival function $S_0(\cdot)$, given its semiparametric optimality when $\mathbf{Z}(\cdot)$ is time-independent. Denote the solution to the weight estimating equations by $\hat{\beta}_{w,n}$. Then all the asymptotic properties of $\hat{\beta}_{w,n}$ can be derived similarly as in Theorem 3:

Corollary 4—Under the same conditions of Theorem 3, as $n \rightarrow \infty$, the following hold for some pre-specified weight functions $W_n(\cdot) \rightarrow w(\cdot)$ almost surely:

- a. $\hat{\beta}_{w,n} \in \mathcal{N}$ converges to β_0 almost surely;
- b. $n^{1/2}(\hat{\beta}_{w,n} - \beta_0)$ converges weakly to a mean-zero normal variate with variance-covariance $U_w^{-1} V_w (U_w^{-1})^T$, where

$$U_w = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \int_0^\tau W_n(t) \{Z_i(t) - \bar{Z}(t)\} \left[\left\{ \frac{\partial \widehat{R}_n(t; \beta_0)}{\partial \beta} + e^{\beta_0^T Z_i(t)} Z_i(t) \right\} dN_i(t) + Y_i(t) \left\{ \frac{\partial \widehat{R}_n(t; \beta_0)}{\partial \beta} \beta_0^T + R_0(t) \right\} dZ_i(t) \right]^T,$$

$$V_w = \lim_{n \rightarrow \infty} \int_0^\tau \frac{1}{n} \sum_{i=1}^n \frac{Y_i(t) \xi_{w,i}(t)^{\otimes 2} \{dR_0(t) - R_0(t) d\log B_i(t; \beta_0)\}}{B_i(t; \beta_0) + R_0(t)},$$

and

$$\xi_{w,i}(t; \beta) = W_n(t) \{Z_i(t) - \bar{Z}(t)\} \{B_i(t; \beta) + R_0(t)\} - \frac{P_n(t; \beta) \{B_i(t; \beta) + R_0(t)\}}{\sum_j Y_j(t)} \left[\sum_{k=1}^n \int_0^\tau \frac{W_n(u) \{Z_k(u) - \bar{Z}(u)\} Y_k(u)}{P_n(u; \beta)} d\log \{R_0(u) + B_k(u; \beta)\} \right];$$

- c. U_w and V_w can be consistently estimated by

$$\widehat{U}_{w,n} = n^{-1} \sum_{i=1}^n \int_0^\tau W_n(t) \{Z_i(t) - \bar{Z}(t)\} \left[\left\{ \frac{\partial \widehat{R}_n(t; \widehat{\beta}_n)}{\partial \beta} + e^{\widehat{\beta}_n^T Z_i(t)} Z_i(t) \right\} dN_i(t) + Y_i(t) \left\{ \frac{\partial \widehat{R}_n(t; \widehat{\beta}_n)}{\partial \beta} \widehat{\beta}_n^T + \widehat{R}_n(t; \widehat{\beta}_n) \right\} dZ_i(t) \right]^T,$$

and

$$\widehat{V}_{w,n} = n^{-1} \sum_{i=1}^n \int_0^\tau \frac{Y_i(t) \widehat{\xi}_{w,i}(t)^{\otimes 2} \{d\widehat{R}_n(t; \widehat{\beta}_n) - \widehat{R}_n(t; \widehat{\beta}_n) d\log B_i(t; \widehat{\beta}_n)\}}{B_i(t; \widehat{\beta}_n) + \widehat{R}_n(t; \widehat{\beta}_n)},$$

respectively, where $\hat{\xi}_{w,A}(t) = \xi_A(t; \hat{\beta}_{w,n}, \hat{R}_n)$.

Apparently, one use of the weighted estimating equations in (12) is that data analysts may be able to choose optimal weight functions to minimize the sandwich estimate of variance to reach semiparametric efficiency, for example, using a “sample-splitting” technique (Lin and Ying, 1994). In general, moment-based semiparametric estimation approaches do not guarantee that optimal weight functions can be easily obtained, particularly in the presence of time-varying covariates. To see this difficulty, consider the hazard function of a univariate covariate $\tilde{Z}(\cdot)$:

$$\lambda(t|\tilde{Z}) = \frac{\lambda_0(t)}{1 - S_0(t) + S_0(t)\exp\{\beta Z(t)\}} - \frac{\{1 - S_0(t)\}\beta Z'(t)}{1 - S_0(t) + S_0(t)\exp\{\beta Z(t)\}}, \tag{13}$$

Thus an optimal weight function $w_{\text{opt}}(\cdot)$ for the weighted estimating equations (12) would nominally satisfy that:

$$w_{\text{opt}}(t) \propto \lim_{\beta \rightarrow 0} \frac{1}{\beta} \log \left\{ \frac{\lambda(t|\tilde{Z})}{\lambda_0(t)} \right\} = -S_0(t) - \{1 - S_0(t)\}\lambda_0(t)\{\log Z(t)\}',$$

(Bickel and Kwon, 2001). When $Z(\cdot)$ is time-independent, it is easy to see that an optimal weight function would be proportional to $S_0(\cdot)$, the baseline survival function. However, when $Z(\cdot)$ is time-varying, it is less straightforward to obtain a simple form of $w_{\text{opt}}(\cdot)$ without estimating $\lambda_0(\cdot)$, although in practice some kernel estimates of $\lambda_0(\cdot)$ can be used (Tsiatis, 1990).

3.2 Maximum likelihood method

Nevertheless, we can apply a nonparametric maximum likelihood estimation (NPMLE) procedure to estimate the extended model, following the approach recently advocated by Zeng and Lin (2006, 2007). Denote $\Lambda_0(t)$ the cumulative hazard function of $\lambda_0(t)$, i.e., $\Lambda_0(t) = \int_0^t \lambda_0(u) du$. Then under the extended model, it is straightforward to write out a likelihood function for β and Λ_0 , which is proportional to

$$\mathcal{L} = \prod_{i=1}^n \prod_{t \leq \tau} \left\{ Y_i(t) \Lambda'(t|\tilde{Z}_i) \right\}^{dN_i(t)} \exp \left\{ - \int_0^\tau Y_i(u) d\Lambda(u|\tilde{Z}_i) \right\}, \tag{14}$$

where

$$\Lambda'(t|\tilde{Z}_i) = \{-\log S(t|\tilde{Z}_i)\}' = \frac{\Lambda_0'(t) - [1 - \exp\{-\Lambda_0(t)\}]\beta^T dZ_i(t)}{1 - \exp\{-\Lambda_0(t)\} + \exp\{-\Lambda_0(t) + \beta^T Z_i(t)\}}.$$

Apparently, the maximum of \mathcal{L} does not exist if Λ_0 are restricted to be absolutely continuous. Instead, we allow Λ_0 to be discrete and replace $d\Lambda_0(t)$ by the jump size of Λ_0 at t , $\Lambda_0\{t\}$. Then calculation of the NPMLE of β and Λ_0 is equivalent to maximizing \mathcal{L} with respect to β and $\Lambda_0(\cdot)$ being the step functions with jumps at observed event times X_i , $i = 1, 2, \dots, n$. Denote the NPMLE of β and $\Lambda_0(\cdot)$ by $\hat{\beta}_n$ and $\hat{\Lambda}_{0,n}(\cdot)$, respectively. Then the NPMLEs of $\hat{\beta}_n$ and $\hat{\Lambda}_{0,n}(\cdot)$ have the following asymptotic properties:

Theorem 5—Under the conditions specified in the Appendix,

- a. $|\check{\beta}_n - \beta_0| + \sup_{t \in [0, \tau]} |\check{\Lambda}_{0,n}(t) - \Lambda_0(t)|$ converges to 0 almost surely;
- b. $n^{1/2}(\check{\beta}_n - \beta_0, \check{\Lambda}_{0,n} - \Lambda_0)$ converges weakly to a zero-mean Gaussian process in $\mathcal{R}^p \times \mathcal{H}(\mathcal{A})$, where $\mathcal{A} = \{h(u) \text{ of bounded variation on } [0, \tau]: \|h(u)\| \leq 1\}$. And the limiting covariance matrix of $n^{1/2}(\check{\beta}_n - \beta_0)$ reaches the semiparametric efficiency bound.

The NPMLE for the extended proportional odds model is a special example under the general NPMLE theory for semiparametric regression models in Zeng and Lin (2007). So it is straightforward to follow the general theory to establish the desired asymptotic properties in Theorem 5, which is omitted here due to space limit. Nevertheless, details of these proofs can be found in Appendix B of Zeng and Lin (2007).

As shown in Theorems 3 and 5, both the moment-based estimation procedure and the NPMLE procedure provide asymptotically tractable estimators for the regression parameters. In particular, when the extended model is correctly specified, the NPMLE procedure yields asymptotically semiparametric efficient estimators, which may have a great advantage in theory and in practical inferences. Nevertheless, the moment-based procedure would yield a simpler closed-form estimator of the baseline odds function and generally robust sandwich estimates of variance. In fact, through appropriate weight functions, moment-based estimators may gain substantial efficiency.

4. NUMERICAL STUDIES

4.1 Computing Algorithms

For either estimation procedure proposed earlier, computing can be challenging. For the moment-based estimating equations, they are not necessarily always smooth or monotone, standard numerical approaches, such as Newton-Raphson, do not work well, particularly when sample size is small. For lower dimensions, a direct grid search or bisection methods can be used to estimate the regression parameters. For higher dimensions, random search technique such as the simulated annealing has been suggested more effective (Lin and Ying, 1995). In our calculation, we use a recursive bisection method proposed by Huang (2002) to search for roots. That is, having solved for the first k components of β , $(\beta_1, \beta_2, \dots, \beta_k)$, we will use a one-dimensional bisection algorithm to solve for β_{k+1} . Details on the recursive bisection algorithm is implemented can be found in Chen and Jewell (2001).

For the NPMLE procedure, computing is more demanding because maximization would occur at all of the observed event times. In our calculation, we use an optimization algorithm called fmi-nunc in MATLAB for unconstrained nonlinear optimization, suggested by Zeng and Lin (2007). For this optimization algorithm, we also provide Hessian matrix so that the algorithm converges faster and more reliably, although the Hessian matrix is not required.

In both calculations, we use initial values from MLE of parametric extended proportional odds models, with baseline survival functions chosen from various parametric families, such as exponential and log-logistic. With these initial values, we find that calculation of moment-based estimates is quite reliable and much faster than the NPMLE's, similar to the experience in Yang and Prentice (1999), which also reported that moment-based estimation showed great reduction in computing time relative to the NPMLE.

4.2 Simulation Studies

Simulation studies are conducted to assess the validity of our proposed inference procedures. In these simulation studies, we choose our simulation parameters similar to those in Zucker & Yang (2006). In each simulated dataset, n observations are generated. For each observation, we generate a time-independent covariate Z_1 that follows a continuous uniform

distribution $\text{Unif}[0, 4]$. For the purpose of demonstration, time-varying covariates are calculated by the interaction between Z_1 and time t : $Z_2(t) = Z_1 \cdot t$. For the proportional odds model that would generate failure times, we consider a baseline odds function $F(t|Z=0)/\{1 - F(t|Z=0)\} = 0.01 \cdot t$. Censoring times are generated such that the censoring proportion of each simulated data is around 30%. We present our simulation results in Tables 1 and 2. Each cell in the tables is based on 10,000 simulated datasets. Sample size n is selected to be 100, 300 or 500, representing relatively small, medium or large sample sizes, respectively.

In Table 1, we present simulation results from fitting the extended proportional odds model with Z_1 only, i.e., $\text{logit } S(t|Z_1) = -\log R(t) + \beta Z_1$. This is indeed a proportional odds model with time-independent covariates. Three estimation procedures are used to evaluate their performance in estimating the regression parameter for Z_1 :

1. ME-0: solving unweighted moment-based estimating equations $S_n(\beta) = 0$;
2. ME-W: solving weighted moment-based estimating equations $S_{n,w}(\beta) = 0$, where the weight function is chosen to be the Kaplan-Meier estimate of the baseline survival function;
3. NPMLE: maximizing nonparametric likelihood function \mathcal{L} .

As shown in the table, all these three estimation procedures for this model with time-independent covariates yield virtually unbiased estimates, and reasonable coverage probabilities. Among them, NPMLE estimates apparently outperform the unweighted moment-based estimation approach uniformly. However, when the weight function is incorporated in the moment-based estimating equations, efficiency of the moment-based estimates is greatly improved, almost as good as their NPMLE counterparts.

In Table 2, we present simulation results from the extended proportional odds model that include both Z_1 and $Z_2(\cdot)$ as covariates, i.e., $\text{logit } S(t|Z) = -\log R(t) + \beta Z_1 + \gamma Z_2(t)$. As shown in Table 2, both estimates for the time-independent and the time-varying covariates are also virtually unbiased for all of the three estimating procedures, regardless of different sample sizes and true parameter values. However, when sample size is relatively small, the discrepancy between empirical and large-sample approximation variance estimates tends to be big. For relatively large sample sizes, our large-sample approximation appears to behave well: variance estimates are consistent between the empirical and their large-sample approximation ones, and coverage probabilities are around their nominal value of 95%. Again, NPMLE estimates outperform unweighted moment-based estimation procedure by a noticeable margin in their variances. Such an advantage is less noticeable when weighted moment-based estimation procedure is used, even though the chosen weight function is not necessarily the one to reach the semiparametric efficiency, as shown in our earlier calculation for the extended model with time-dependent covariates.

4.3 Data Analyses

In this section, we demonstrate our proposed methods by analyzing the motivating example of the HIVNET 012 trial, and the well-known VA Lung Cancer Trial (Prentice, 1973).

Example 1—In the HIVNET 012 Trial, a total of 626 HIV-1 infected pregnant women in Uganda were recruited and randomized to receive either NVP or AZT at more than 36 weeks of gestation. Complete medical histories and physical examination of all participants were collected before their entry to the study, on enrollment, at delivery, at discharge from hospital and at 7 days and 6 weeks after delivery. Among those 619 women analyzed in the primary analysis (Jackson et al., 2003), 308 women were randomized to receive AZT, and 311 were randomized to receive NVP. Their median ages were 25 and 24 ($p > 0.1000$); their

CD4+ counts measured at the baseline were 426 and 459 ($p > 0.2500$); and their HIV-1 RNA viral copies were 27,800 and 25,247 ($p > 0.7000$), all respectively. The total follow-up time is 18 months. At the end of the trial, the primary efficacy analysis showed that the HIV-1 transmission risks in the AZT and the NVP groups were 10.4% and 8.2% ($p > 0.3000$) at birth, 21.3% and 11.9% ($p < 0.0300$) by age 6–8 weeks, and 25.1% and 13.1% ($p < 0.0001$) by age 14–16 weeks, which supported that NVP significantly lowered the vertical transmission risk in less-developed countries.

According to the HIVNET 012 study design, the NVP was administered only once in order to maintain certain plasma drug concentration for up to 7 days. Preliminary studies showed that its potent antiviral effect usually persisted for 1 to 2 weeks, followed by a rapid development of detectable viral resistance to the NVP by 6 to 8 weeks after ingestion (Eshleman et al., 2001). Its long-term effect to suppress the maternal viral load could be diminished, which might lead to a decreasing effect on the ultimate infant survival, as time progresses. Therefore, the proportional odds model may serve as a reasonable tool given its feature of modeling converging hazard functions.

In our analysis, we first fit the Cox model with these covariates: the treatment indicator of NVP versus AZT, the treatment-time interaction, and maternal CD4+ counts and viral loads (VL). Here, maternal CD4+ counts and VLs are grouped into three categories: low maternal risk category (CD4+ > 350 and VL ≤ 50,000), intermediate maternal risk category (CD4+ > 350 and VL > 50,000, or CD4+ ≤ 350 and VL > 50,000), and high maternal risk category (CD4+ ≤ 350 and VL > 50,000). An estimate of the regression parameter for treatment-time interaction is -0.003 ($s.e. = 0.0003$) with a 95% CI $(-0.004, -0.003)$, which is statistically significantly different from no interaction ($p < 0.001$). As a result, the NVP administration is not considered to have a constant effect on the hazard functions of infant survival. The comparison of their respective survival functions would then depend on both the regression coefficients and the baseline hazard function, which shall lead to a complicated interpretation of the NVP's effect in terms of the infants' survival functions. Nevertheless, the negative interaction means that NVP and time are potentially antagonistic in hazard functions. This may mean that the NVP's effect on the infants' hazard functions may be reduced during the observation period.

We also fit an extended proportional odds model (2) with the same set of covariates. The estimates of regression parameters are tabulated in Table 3. As shown in Table 3, the interaction between time and the administration of NVP to HIV-infected mothers during labor and delivery is not statistically significant ($p = 0.08$). The main effect of the NVP administration is nevertheless statistically significant. That is, under the assumed proportional odds model, the NVP administration would improve the odds of infants' overall survival by $e^{0.265} - 1 = 30.3\%$ ($p = 0.01$) uniformly over time. For the maternal CD4+ counts and VLs, when compared with the low maternal risk category, the intermediate and high maternal risk categories appear to be associated with $1 - e^{-0.026} = 2.6\%$ and $1 - e^{-0.052} = 5.1\%$ of loss of infant survival, respectively. Additional NPML of fitting the same model yields a similar result: the administration of NVP would significantly improve the odds of infant survival by $e^{0.276} - 1 = 31.7\%$. This significant improvement of infant survival due to NVP has a profound impact: its broader administration shall improve the overall human life expectancy, especially in resource-limited settings.

Example 2—The well-known VA Lung Cancer Trial dataset in Prentice (1973) has been analyzed extensively in the statistical literature. The originally published dataset contains only two baseline covariates, performance score Z_{PS} on a scale of 0 to 100, and tumor type Z_{TU} indicating large, adeno, small and squamous types, and the censored time-to-event outcome of lung cancer survival. Here, Z_{TU} are 3-dimensional dummy vectors with the large

type as the baseline category. We apply our proposed extended model to this dataset to allow comparison between our proposed method and the existing ones in the literature. Moreover, we demonstrate that the extended model can also serve as a goodness-of-fit tool for the usual proportional odds model.

Specifically, we analyze a subgroup of 97 patients without prior therapy to study the association between the lung cancer survival and Z_{PS} and Z_{TU} . In Bennett (1983), logit $\{S(\cdot)\}$ are plotted by $Z_{PS} \leq 50$ and $Z_{PS} > 50$, which leads to a visual justification of assuming constant effect of Z_{PS} in a parametric proportional odds model

$$\text{logit } S(t|Z_{PS}) = -\log R(t) + \beta_1 Z_{PS} + \gamma^T Z_{TU},$$

where $R(\cdot)$ is of log-logistic distributions. Most subsequent works, including Pettitt (1984), Cheng, Wei and Ying (1995) and Yang and Prentice (1998), all similarly assume that the performance score would have constant effect in their proportional odds models, despite the fact that the cutoff of 50 is chosen for convenience.

In our extended model, we instead use Z_{PS} of its original scale. Results from a series of model-fitting by the extended proportional odds model are tabulated in Table 4. When Z_{PS} is the only covariate in the extended model, we obtain a significant parameter estimate, as shown by Model I of Table 4. However, when we include an additional time-varying interaction between Z_{PS} and t in Model II, the parameter estimate of this interaction is then 0.016 (*s.e.* = 0.003, $p < 0.001$), which is highly significant. This means that the difference in log survival odds is not necessarily constant throughout time for Z_{PS} , if Z_{TU} are not included. When Z_{TU} are the only covariates in the extended model, small type, compared with large type, appears significantly associated with the lung cancer survival in Model III, while all of the tumor-time interactions are not significant in Model IV. This may indicate that it is reasonable to assume a constant effect for Z_{TU} .

Nevertheless, when no interaction is included in the proposed model, all of the estimates we obtain are similar to the ones in the literature, as shown by Model V in Table 4. This validates our proposed inference procedures that are able to replicate what have been in the literature in absence of time-varying covariates. Moreover, when the time-varying interactions of $(Z_{PS}, Z_{TU}^T)^T$ and t are included in the proposed extended model, we find that only the interaction of Z_{PS} and t is marginally significant at α -level 10%. This may suggest it is reasonable to assume constant covariate effect in the proportional odds model.

As a summary, even when all the covariates are only collected at baseline, the extended proportional odds model can serve as a useful tool to assess covariate-time interaction, and leads to an approach of goodness-of-fit assessment for the usual proportional odds model.

5. DISCUSSION

In this article, we study an extended proportional odds model in presence of time-varying covariates. This extended model directly models survival functions, with an appealing interpretation of regression parameters in absolute risk. In the literature, inclusion of time-varying covariates in the usual proportional odds model has been discussed. For example, a specific version of our proposed model (5) was discussed in Yang and Prentice (1999). Although the self-consistency integral equations proposed in Yang and Prentice (1999) work with the usual proportional odds model elegantly, they are yet to be developed for time-varying covariates. Compared with the integral equations, the ordinary differential equations

used in this article yield a simple closed-form estimator for the baseline odds function, which eventually makes the estimation of regression parameters more straightforward.

Nevertheless, the popular Cox proportional hazards model can also incorporate time-varying covariates. For example, the Cox proportional hazards model would usually assume that

$$\lambda\{t|\mathbf{Z}^*(t)\}=\lambda_0(t)\exp\{\boldsymbol{\beta}^T\mathbf{Z}^*(t)\}, \tag{15}$$

as in Kalbfleisch and Prentice (2002, pp. 96–98). It is then apparent that the association of $\mathbf{Z}^*(\cdot)$ in survival functions is not always straightforward in model (15), because it implies that

$$\Lambda\{t|\mathbf{Z}(t)\}=\Lambda_0(t)\int_0^t\exp\{\boldsymbol{\beta}^T\mathbf{Z}(u)\}\psi(u;t)du,$$

where $\Lambda(t)=\int_0^t\lambda(u)du$, and $\psi(u;t)=\lambda_0(u)/\Lambda_0(t)$ such that $\int_0^t\psi(u;t)du=1$. Mathematically, unless $\mathbf{Z}(\cdot)$ are constant almost everywhere on $[0, t]$, the regression parameter $\boldsymbol{\beta}$ does not yield a direct comparison between survival functions. This has been the same complication in studying association of survival functions for other hazard-based regression models, for example, the joint hazards model of Chen and Jewell (2000), in presence of time-varying covariates.

Without time-varying covariates, both the proportional odds model and the Cox proportional hazards model are special examples of the transformation model (Cheng, Wei and Ying, 1995). Recent effort has been focused on developing transformation models with time-varying covariates (Kosorok, Lee and Fine, 2004; Zeng and Lin, 2007). For example, in the latest Zeng and Lin (2007), an extended transformation model with time-varying covariates assumes that

$$\Lambda(t|\tilde{\mathbf{Z}})=G\left[\Lambda_0(t)\int_0^t\exp\{\boldsymbol{\beta}^T\mathbf{Z}(u)\}\psi(u;t)du\right], \tag{16}$$

where $G(\cdot)$ is a continuously differentiable and strictly increasing transformation function. Similar to the Cox proportional hazards model, however, this model does not provide direct comparison of survival functions either, unless $\mathbf{Z}(\cdot)$ are constant almost everywhere.

One direct extension of the transformation model to include time-varying covariates, as suggested by a reviewer, is to consider the following model,

$$S(t|\tilde{\mathbf{Z}})=\exp(-G[\Lambda_0(t)\exp\{\boldsymbol{\beta}^T\mathbf{Z}(t)\}]), \tag{17}$$

where $\Lambda_0(\cdot)$ is an unspecified monotonically increasing function with $\Lambda_0(0)=0$, and $G(\cdot)$ is a specified monotonically increasing function with $G(0)=1$. As shown in this model, the effect of time-varying covariates are also directly related the survival, which includes the extended proportional odds model and the extended Cox model. This in fact appealing in offering a formal way to justify between these two models. Both martingale based estimating equations and self-consistent integral equations can be used in estimating regression parameters, however, there may not be closed-form solutions for the baseline function. This extension shall be explored in future.

In this article, we mainly focus on estimating the regression parameters in the extended model, given that the major interest of our motivating examples in randomized clinical trials is to compare treatment effect. One particular feature of the extended model, however, is that it may be used to estimate subject-specific survival probabilities given a covariate profile. To serve this purpose, it is also critical to develop more asymptotic properties for estimating the baseline odds function. Conceptually, this shall be straightforward, because of the establishment of consistency and asymptotic normality of the regression parameters, as shown in Cheng, Wei and Ying (1997). Extensive theory calculation and reliable computing algorithms for the asymptotic variance-covariance of regression parameter estimators and baseline odds function estimators are yet to be developed in both moment-based estimating equations and the NPMLE approaches in our future research.

References

- Bennett S. Analysis of Survival Data by the Proportional Odds Model. *Statistics in Medicine*. 1983; 2:273–277. [PubMed: 6648142]
- Bickel PJ, Kwon J. Inference for Semiparametric Models: Some Questions and an Answer. *Statistica Sinica*. 2001; 11:863–886.
- Chen K, Jin Z, Ying Z. Semiparametric Analysis of Transformation Models with Censored Data. *Biometrika*. 2002; 89:659–668.
- Chen YQ, Jewell NP. On a General Class of Hazards Regression Models. *Biometrika*. 2001; 88:687–702.
- Cheng SC, Wei LJ, Ying Z. Analysis of Transformation Models with Censored Data. *Biometrika*. 1995; 82:835–45.
- Cheng SC, Wei LJ, Ying Z. Predicting Survival Probabilities with Semi-parametric Transformation Models. *Journal of the American Statistical Association*. 1997; 92:227–235.
- Cox DR. Regression Models and Life Tables” (with discussion). *Journal of Royal Statistical Society, Series B*. 1972; 34:187–220.
- Dabrowska DM, Doksum KA. Estimation and Testing in a Two-Sample Generalized Odds-Rate Model. *Journal of the American Statistical Association*. 1988; 83:744–749.
- Eshleman SH, Mracna M, Guay LA, Deseyve M, Cunningham S, Mirochnick M, Musoke P, Fleming T, Fowler MG, Mofenson LM, Mmiro F, Jackson JB. Selection and fading of resistance mutations in women and infants receiving nevirapin to prevent HIV-1 vertical transmission (HIVNET 012). *Lancet*. 2001; 15:1951–1957.
- Fleming TR, O’Fallon JR, O’Brien PC. Modified Komogorov-Smirnov Test Procedures with Application to Arbitrarily Right-Censored Data. *Biometrics*. 1980; 36:607–625.
- Hanson T, Yang MG. Bayesian Semiparametric Proportional Odds Models. *Biometrics*. 2007; 63:88–95. [PubMed: 17447933]
- Huang YJ. Calibration Regression of Censored Lifetime Medical Cost. *Journal of the American Statistical Association*. 2002; 97:318–327.
- Jackson JB, Musoke P, Fleming T, Guay LA, Bagenda D, Allen M, Nakabiito C, Sherman J, Bakaki P, Owor M, Ducar C, Deseyve M, Mwatha A, Emel L, Duefield C, Mirochnick M, Fowler MG, Mofenson L, Miotti P, Gigliotti M, Bray D, Mmiro F. Intrapartum and Neonatal Single-Dose Nevirapine Compared with Zidovudine for Prevention of Mother-to-Child Transmission of HIV-1 in Kampala, Uganda: 18-month Follow-Up of the HIVNET 012 Randomised Trial. *Lancet*. 2003; 362:859–868. [PubMed: 13678973]
- Jung S-H. Regression Analysis for Long-Term Survival Rate. *Biometrika*. 1996; 83:227–232.
- Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. 2. Wiley; Hoboken: 2002.
- Kosorok MR, Lee BL, Fine JP. Robust Inference for Univariate Proportional Hazards Frailty Regression Models. *The Annals of Statistics*. 2004; 32:1448–1491.
- Lin DY, Ying Z. Semiparametric Inference for the Accelerated Life Model with Time-Dependent Covariates. *Journal of Statistical Planning and Inferences*. 1995; 44:47–63.

- McCullagh P. Regression Models for Ordinal Data. *Journal of the Royal Statistical Society, Series B.* 1980; 42:109–142.
- Murphy SA, Rossini AJ, van der Vaart AW. Maximum Likelihood Estimation in the Proportional Odds Model. *Journal of the American Statistical Association.* 1996; 92:968–976.
- Nelson W. Hazard Plotting for Incomplete Failure Data. *Journal of Quality Technology.* 1969; 1:27–52.
- Peng L, Huang YJ. Survival analysis with temporal covariate effects. *Biometrika.* 2007; 94:719–733.
- Pettitt A. Proportional Odds Models for Survival Data and Estimates Using Ranks. *Applied Statistics.* 1984; 33:169–175.
- Prentice RL. Exponential survivals with censoring and explanatory variables. *Biometrika.* 1973; 60:279–88.
- Scharfstein DO, Tsiatis AA, Gilbert PB. Semiparametric Efficient Estimation in the Generalized Odds-Rate Class of Regression Models for Right-Censored Time-to-Event Data. *Lifetime Data Analysis.* 1998; 4:355–391. [PubMed: 9880995]
- The HIVNET/HPTN Group. The HIVNET 012 Protocol. The HIV Prevention Trials Network; 2003. HIVNET 012: Phase IIB Trial to Evaluate the Efficacy of Oral Nevirapine and The Efficacy of Oral AZT in Infants Born to HIV-Infected Mothers in Uganda for Prevention of Vertical HIV Transmission - a Study of the HIVNET/HPTN Group. <http://www.hptn.org>
- Tsiatis AA. Estimating Regression Parameters Using Linear Rank Tests for Censored Data. *Annals of Statistics.* 1990; 18:354–372.
- Yang S, Prentice RL. Semiparametric Inference in the Proportional Odds Regression Model. *Journal of the American Statistical Association.* 1999; 94:125–136.
- Zeng D, Lin DY. Efficient Estimation of Semiparametric Transformation Models for Counting Processes. *Biometrika.* 2006; 93:627–40.
- Zeng D, Lin DY. Maximum Likelihood Estimation in Semiparametric Regression Models with Censored Data. *Journal of the Royal Statistical Society, Series B.* 2007; 69:1–30.
- Zhang M, Davidian M. ‘Smooth’ Semiparametric Regression Analysis for Arbitrarily Censored Time-to-Event Data. *Biometrics.* 2007; 64:567–576. [PubMed: 17970813]
- Zucker DM, Yang S. Inference for a Family of Survival Models Encompassing the Proportional Hazards and Proportional Odds Model. *Statistics in Medicine.* 2006; 25:995–1014. [PubMed: 16220492]

APPENDIX: ASYMPTOTIC PROPERTIES

For ease of presentation, in this Appendix we outline a proof of the stated asymptotic properties in Theorems 3, assuming that $\mathbf{Z}(\cdot)$ and regression parameter $\boldsymbol{\beta}$ are scalar. Following the proof, it is straightforward to establish asymptotic properties in higher dimensions.

Regularity Conditions

We first assume some general regularity conditions for the asymptotic results summarized in Theorem 3, Corollary 4 and Theorem 5. Additional assumptions, if needed for a specific result, are stated in the proof.

- C1** $\beta_0 \in \mathcal{B} \subset \mathcal{R}$ lies in the interior of a compact set \mathcal{C} .
- C2** The baseline density function f_0 of T and its derivative f_0' are bounded, satisfying that

$$\int_0^\infty \left\{ \frac{f_0'(t)}{f_0(t)} \right\}^2 f_0(t) dt < \infty, \text{ and } \int_0^\infty t^\varepsilon f_0(t) dt < \infty \text{ for some } \varepsilon > 0.$$

- C2'** Baseline cumulative hazard function $\Lambda_0(\cdot) > 0$ is strictly increasing and continuously differentiable.
- C3** The density function g of C are uniformly bounded, i.e., $\sup_t g(t) < \infty$.
- C4** τ is finite such that $\text{pr}(T > \tau) > 0$ and $\text{pr}(C = \tau) > 0$.
- C5** There exists $\kappa_1 > 0$ and $\kappa_2 > 0$ such that $\sup_{|s-t| \leq n^{-\kappa_2}} n^{-1} \sum_i |Z_i(s) - Z_i(t)| = O(n^{-1/2-\kappa_1})$. And, for any positive sequence $d_n \rightarrow 0$, there exists $\kappa_3 > 0$ such that $\sup_{|s-t| \leq d_n} n^{-1} \sum_i |Z_i(s) - Z_i(t)| = o(\max(d_n^{\kappa_3}, n^{-\kappa_3}))$.

These conditions are mainly used to facilitate our proofs, which are not necessarily the weakest. Nevertheless, most of these conditions have been commonly used in the literature. Specifically, C1 and C2' are used in Zeng and Lin (2006) to justify the asymptotics for NPMLE, and C2-C4 are used in Chen, Jin and Ying (2001). C5 is the smoothness condition for time-dependent covariates, which was used in Lin and Ying (1995). C5 is satisfied when $Z_i(\cdot)$ are smooth in the sense that they have uniformly bounded derivatives or when $Z_i(\cdot)$ are step functions with random between-jump times having uniformly bounded densities.

Proof of Theorem 3

To show the asymptotic results in this theorem, we first establish a martingale representation of $S_n(\beta_0)$.

Consider that

$$\begin{aligned} S_n(\beta_0) &= \sum_{i=1}^n \int_0^\tau Z_i(t) \left[\{B_i(t) + \widehat{R}_n(t; \beta_0)\} dN_i(t) - Y_i(t) \{d\widehat{R}_n(t; \beta_0) - \widehat{R}_n(t; \beta_0) d\log B_i(t)\} \right] \\ &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left[\{B_i(t) + \widehat{R}_n(t; \beta_0)\} dN_i(t) + Y_i(t) \widehat{R}_n(t; \beta_0) d\log B_i(t) \right], \end{aligned}$$

where $\bar{Z}(t) = \sum_j Y_j(t) Z_j(t) / \sum_j Y_j(t)$ as defined in §3. Replace $N_i(t)$ by

$$N_i(t) = M_i(t) + \int_0^t Y_i(u) \{B_i(u; \beta_0) + R_0(u)\}^{-1} \{dR_0(u) - R_0(u) d\log B_i(u; \beta_0)\},$$

and then we have

$$\begin{aligned}
 S_n(\beta_0) &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left\{ B_i(t; \beta_0) + \widehat{R}_n(t; \beta_0) \right\} \\
 &\times \left[M_i(t) + \frac{Y_i(t) dR_0(t) - R_0(t) d\log B_i(t; \beta_0)}{B_i(t; \beta_0) + R_0(t)} \right] + Y_i(t) \widehat{R}_n(t; \beta_0) d\log B_i(t; \beta_0) \\
 &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left\{ B_i(t; \beta_0) + \widehat{R}_n(t; \beta_0) \right\} dM_i(t) \\
 &+ \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left\{ B_i(t; \beta_0) + \widehat{R}_n(t; \beta_0) \right\} \left[\frac{Y_i(t) dR_0(t) - R_0(t) d\log B_i(t; \beta_0)}{B_i(t; \beta_0) + R_0(t)} \right] \\
 &+ \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} Y_i(t) \widehat{R}_n(t; \beta_0) d\log B_i(t; \beta_0) \\
 &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left\{ B_i(t; \beta_0) + \widehat{R}_n(t; \beta_0) \right\} dM_i(t) \\
 &+ \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \frac{\widehat{R}_n(t; \beta_0) - R_0(t)}{B_i(t; \beta_0) + R_0(t)} Y_i(t) dR_0(t) \\
 &+ \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left\{ \widehat{R}_n(t; \beta_0) - R_0(t) \right\} \frac{Y_i(t) B_i(t; \beta_0) d\log B_i(t; \beta_0)}{B_i(t; \beta_0) + R_0(t)}.
 \end{aligned}$$

Further replace $\widehat{R}_n(t; \beta_0) - R_0(t)$ by its martingale representation in (11), and we have

$$\begin{aligned}
 S_n(\beta_0) &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left\{ B_i(t; \beta_0) + \widehat{R}_n(t; \beta_0) \right\} dM_i(t) \\
 &+ \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \frac{\widehat{R}_n(t; \beta_0) - R_0(t)}{B_i(t; \beta_0) + R_0(t)} Y_i(t) d\{R_0(t) + B_i(t; \beta_0)\} \\
 &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left\{ B_i(t; \beta_0) + \widehat{R}_n(t; \beta_0) \right\} dM_i(t) \\
 &+ \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left[- \int_0^t \frac{P_n(u; \beta_0) \sum_k \{B_k(u; \beta_0) + R_0(u)\}}{P_n(t; \beta_0) \sum_j Y_j(u)} dM_k(u; \beta_0, R_0) \right] \\
 &\quad \times Y_i(t) d\log\{R_0(t) + B_i(t; \beta_0)\}.
 \end{aligned}$$

Therefore, as a result of integration by parts, we have

$$\begin{aligned}
 S_n(\beta_0) &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left\{ B_i(t; \beta_0) + \widehat{R}_n(t; \beta_0) \right\} dM_i(t), \\
 &- \sum_{i=1}^n \int_0^\tau \frac{P_n(t; \beta_0) \{B_i(t; \beta_0) + R_0(t)\}}{\sum_j Y_j(t)} \left[\sum_{k=1}^n \int_t^\tau \frac{\{Z_k(u) - \bar{Z}(u)\} Y_k(u)}{P_n(u)} d\log\{R_0(u) + B_k(u)\} \right] dM_i(t) \\
 &= \sum_{i=1}^n \int_0^\tau \xi_i(t; \beta_0) dM_i(t).
 \end{aligned}$$

where

$$\xi_i(t; \beta) = \{Z_i(t) - \bar{Z}(t)\} \{B_i(t; \beta) + R_0(t)\} - \frac{P_n(t; \beta) \{B_i(t; \beta) + R_0(t)\}}{\sum_j Y_j(t)} \left[\sum_{k=1}^n \int_t^\tau \frac{\{Z_k(u) - \bar{Z}(u)\} Y_k(u)}{P_n(u; \beta)} d\log\{R_0(u) + B_k(u; \beta)\} \right],$$

as defined in Theorem 3. By this martingale representation of $S_n(\beta_0)$, it is apparent that $n^{-1} S_n(\beta_0) \rightarrow_P 0$ by the weak law of large numbers (WLLN).

Let $\mu(t) = \lim_n \sum_i Y_i(t) Z_i(t) / \sum_j Y_j(t) = \mathfrak{E} \{ Y(t) Z(t) / \mathfrak{E} Y(t) \}$ and

$$R(t;\beta) = \lim_n \widehat{R}_n(t;\beta) = \frac{\lim_n \int_0^t P_n(t;\beta) dQ_n(t;\beta)}{\lim_n P_n(t;\beta)} \\ = \frac{\int_0^t \exp\left[-\int_0^u \mathcal{E}\{dN(s)+\beta Y(s)dZ(s)\}/\mathcal{E}Y(s)\right] \mathcal{E}\{\exp\{\beta Z(u)\}dN(u)\}/\mathcal{E}Y(u)}{\exp\left[-\int_0^t \mathcal{E}\{dN(u)+\beta Y(u)dZ(u)\}/\mathcal{E}Y(u)\right]}$$

It is apparent $R_0(t) = R(t; \beta_0)$ under the assumed model (5). Furthermore, let

$$H_1(t) = \int_0^t \frac{\mathcal{E}\{Y(u)dZ(u)\}}{\mathcal{E}Y(u)} \text{ and } dH_2(t;\beta_0) = \frac{H_1(t)\mathcal{E}\{e^{\beta_0 Z(t)}dN(t)\}}{\mathcal{E}Y(t)} - \frac{\mathcal{E}\{e^{\beta Z(t)}Z(t)dN(t)\}}{\mathcal{E}Y(t)}$$

Then

$$R'_\beta(t;\beta) = \frac{d}{d\beta} \frac{\int_0^t \exp\left[-\int_0^u \mathcal{E}\{dN(s)+\beta Y(s)dZ(s)\}/\mathcal{E}Y(s)\right] \mathcal{E}\{\exp\{\beta Z(u)\}dN(u)\}/\mathcal{E}Y(u)}{\exp\left[-\int_0^t \mathcal{E}\{dN(u)+\beta Y(u)dZ(u)\}/\mathcal{E}Y(u)\right]} \\ = \left\{ \frac{1}{\exp\left[-\int_0^t \mathcal{E}\{dN(u)+\beta Y(u)dZ(u)\}/\mathcal{E}Y(u)\right]} \right\}^2 \\ \times \left(\int_0^t \exp\left[-\int_0^u \mathcal{E}\{dN(s)+\beta Y(s)dZ(s)\}/\mathcal{E}Y(s)\right] \left[-\int_0^u \mathcal{E}\{Y(s)dZ(s)\}/\mathcal{E}Y(s) \right] \right. \\ \times \mathcal{E}\{\exp\{\beta Z(u)\}dN(u)\}/\mathcal{E}Y(u) \cdot \exp\left[-\int_0^t \mathcal{E}\{dN(u)+\beta Y(u)dZ(u)\}/\mathcal{E}Y(u)\right] \\ \left. + \int_0^t \exp\left[-\int_0^u \mathcal{E}\{dN(s)+\beta Y(s)dZ(s)\}/\mathcal{E}Y(s)\right] \mathcal{E}\{\exp\{\beta Z(u)\}Z(u)dN(u)\}/\mathcal{E}Y(u) \right. \\ \times \exp\left[-\int_0^t \mathcal{E}\{dN(u)+\beta Y(u)dZ(u)\}/\mathcal{E}Y(u)\right] \\ \left. - \int_0^t \exp\left[-\int_0^u \mathcal{E}\{dN(s)+\beta Y(s)dZ(s)\}/\mathcal{E}Y(s)\right] \mathcal{E}\{\exp\{\beta Z(u)\}dN(u)\}/\mathcal{E}Y(u) \right. \\ \left. \exp\left[-\int_0^t \mathcal{E}\{Y(u)dZ(u)\}/\mathcal{E}Y(u)\right] \right).$$

As a result,

$$R'_\beta(t;\beta_0) = -\frac{1}{S_0(t)} \left[\int_0^t S_0(u)dH_2(u) + R_0(u)\exp\{-H_1(t)\} \right].$$

We assume an additional regularity condition on $R'_\beta(t;\beta_0)$ (C6):

$$|\mathcal{E}\{e^{\beta Z(t)}Z(t)\} + R'_\beta(t;\beta_0)d\mathcal{E}N(t) + \{\beta R(t;\beta)\}'_{\beta=\beta_0} \mathcal{E}Y(t)dZ(t)| > 0.$$

This condition can be generally satisfactory, in particular when $Z(\cdot)$ is time-independent.

Now consider an arbitrary $\beta \neq \beta_0$. Denote

$$\sum_{i=1}^n \int_0^{\tau_i} \{Z_i(t) - \mu(t)\} [\{B_i(t;\beta) - B_i(t;\beta_0) + R(t;\beta) - R_0(t)\}dN_i(t) + Y_i(t)\{\beta R(t;\beta) - \beta_0 R_0(t)\}dZ_i(t)]$$

by $s_n(\beta)$. Apparently, $s_n(\beta_0) = 0$. Assume that there exists $\epsilon > 0$ such that (C7):

$$\text{pr}\{|Z_i(t) - \mu(t)| > \epsilon, i=1, 2, \dots, n\} > 0,$$

which essentially means covariate processes cannot be identical for all the study subjects.

Then under conditions (C6) and (C7), we know that $\|\lim_n n^{-1} s'_n(\beta_0)\| > 0$. Without loss of generality, we simply assume that $s'_n(\beta_0) > 0$. Therefore, there must exist a neighborhood of β_0 and an $N_0 > 0$, $U(\beta_0)$ say, such that $n > N_0$, $s_n(\beta)$ is monotonically increasing with a unique zero-crossing at β_0 in $U(\beta_0)$.

We know that

$$\begin{aligned} & n^{-1} S_n(\beta) - n^{-1} S_n(\beta_0) \\ = & n^{-1} \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left[\{B_i(t;\beta) - B_i(t;\beta_0) + \widehat{R}_n(t;\beta) - \widehat{R}_n(t;\beta_0)\} dN_i(t) \right. \\ & \left. + Y_i(t) \{ \widehat{R}_n(t;\beta) d\log B_i(t;\beta) - \widehat{R}_n(t;\beta_0) d\log B_i(t;\beta_0) \} \right] = n^{-1} s_n(\beta), \end{aligned}$$

by the definition of $s_n(\beta)$. Hence, $n^{-1} \|S_n(\beta) - s_n(\beta)\| \rightarrow 0$. Moreover, since

$$S'_n(\beta) = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left[\{B'_i(t;\beta) + \widehat{R}'_n(t;\beta)\} dN_i(t) + Y_i(t) \{ \beta \widehat{R}'_n(t;\beta) \}'_\beta dZ_i(t) \right],$$

and

$$S'_n(\beta) = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \mu(t)\} \left[\{e^{\beta Z_i(t)} Z_i(t) + R'_\beta(t;\beta)\} dN_i(t) + Y_i(t) \{ \beta R(t;\beta) \}'_\beta dZ_i(t) \right],$$

with an application of the Glivenko-Cantelli Lemma, we know that $n^{-1} \|S'_n(\beta) - s'_n(\beta)\| \rightarrow_r 0$ uniformly in $\beta \in U(\beta_0)$ under condition (C5). Therefore, for any $\epsilon > 0$, there exists sufficiently large n and $\nu > 0$, such that $S_n(\beta_0 - \epsilon) \cdot S_n(\beta_0 + \epsilon) < 0$, and $\text{pr}\{\hat{\beta}_n \in (\beta_0 - \epsilon, \beta_0 + \epsilon)\} > 1 - \nu$, which then leads to the consistency of $\hat{\beta}_n$.

Moreover, by the martingale central limit theorem (MCLT), $n^{-1/2} S_0(\beta_0)$ is asymptotically mean-zero normal with variance-covariance matrix

$$\begin{aligned} V = \lim_{n \rightarrow \infty} V_n &= \lim_{n \rightarrow \infty} \langle n^{-1/2} S_n(\beta_0), n^{-1/2} S_n(\beta_0) \rangle(\tau) = \lim_{n \rightarrow \infty} \int_0^\tau \frac{1}{n} \sum_{i=1}^n \xi_i(u) \otimes^2 Y_i(u) d\Lambda_i(u) \\ &= \lim_{n \rightarrow \infty} \int_0^\tau \frac{1}{n} \sum_{i=1}^n \frac{Y_i(u) \xi_i(u) \otimes^2 [\{dR_0(u) - R_0(u) d\log B_i(u; \beta_0)\}]}{B_i(u; \beta_0) + R_0(u)}, \end{aligned}$$

as in Yang & Prentice (1999).

For the asymptotic normality of $\hat{\beta}_n$, we consider that

$$\begin{aligned} \frac{\partial S_n(\beta_0)}{\partial \beta} &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left[\left\{ \frac{\partial B_i(t; \beta_0)}{\partial \beta} + \frac{\partial \widehat{R}_n(t; \beta_0)}{\partial \beta} \right\} dN_i(t) \right. \\ &\quad \left. + Y_i(t) \frac{\partial \widehat{R}_n(t; \beta_0)}{\partial \beta} d\log B_i(t; \beta) + Y_i(t) \widehat{R}_n(t; \beta_0) d\frac{\log B_i(t; \beta_0)}{\partial \beta} \right]^T \\ &= \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left[\left\{ e^{\beta_0 Z_i(t)} Z_i(t) + \frac{\partial \widehat{R}_n(t; \beta_0)}{\partial \beta} \right\} dN_i(t) \right. \\ &\quad \left. + Y_i(t) \frac{\partial \widehat{R}_n(t; \beta_0)}{\partial \beta} d\log B_i(t; \beta) + Y_i(t) \widehat{R}_n(t; \beta_0) dZ_i(t) \right]^T. \end{aligned}$$

Under the assumed regularity conditions, it is true that $n^{-1} S_n(\beta_0)/\beta$ converges to

$$U = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \left[\left\{ \frac{\partial \widehat{R}_n(t; \beta_0)}{\partial \beta} + e^{\beta_0 Z_i(t)} Z_i(t) \right\} dN_i(t) + Y_i(t) \left\{ \frac{\partial \widehat{R}_n(t; \beta_0)}{\partial \beta} \beta_0 + R_0(t) \right\} dZ_i(t) \right].$$

Via Taylor expansion, we obtain that $n^{1/2}(\hat{\beta}_n - \beta_0) \doteq \{n^{-1} S_n(\beta_0)/\beta\}^{-1} \cdot n^{-1/2} S_n(\beta_0)$. Therefore, an application of Delta-method coupled with the consistency of $\hat{\beta}_n$ leads to the conclusions that

$$n^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, U^{-1} V (U^{-1})^T),$$

as stated in the theorem.

In addition, since Corollary 4 establishes same asymptotic properties for the weighted estimators, proof of Corollary 4 follow essentially same steps, except that additional weight functions would be included in all calculations. To save space, we here only present the major steps to prove Theorem 3.

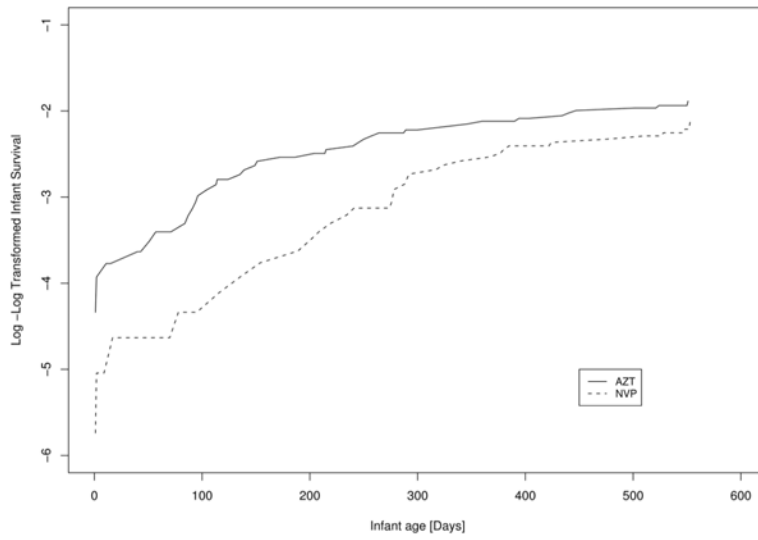


Figure 1.
Kaplan-Meier estimates of HIVNET 012 infant survival

Table 1

Summary of simulation studies with time-independent covariates only

Sample Size		β						
n	β_0	Method	Bias	SSE	MSE	CP		
100	0	ME-0	0.0006	0.294	0.299	0.951		
100	0	ME-W	0.0007	0.246	0.248	0.947		
100	0	NPML	-0.0005	0.245	0.255	0.955		
100	1	ME-0	0.0003	0.308	0.296	0.953		
100	1	ME-W	-0.0003	0.255	0.247	0.954		
100	1	NPML	0.0009	0.253	0.244	0.952		
300	0	ME-0	0.0004	0.132	0.139	0.957		
300	0	ME-W	-0.0001	0.110	0.116	0.949		
300	0	NPML	0.0008	0.109	0.114	0.951		
300	1	ME-0	0.0009	0.180	0.176	0.949		
300	1	ME-W	0.0002	0.151	0.149	0.950		
300	1	NPML	-0.0004	0.149	0.148	0.951		
500	0	ME-0	-0.0002	0.107	0.108	0.947		
500	0	ME-W	0.0001	0.090	0.090	0.951		
500	0	NPML	0.0006	0.090	0.092	0.948		
500	1	ME-0	0.0005	0.128	0.125	0.951		
500	1	ME-W	-0.0003	0.108	0.103	0.950		
500	1	NPML	0.0002	0.107	0.104	0.952		

SSE, sample standard error of estimates; ME-0, unweighted moment-based estimates; MEW, weighted moment-based estimates; NPML, nonparametric maximum likelihood estimates; MSE, mean of estimated standard errors; CP, 95% nominal coverage probabilities.

Table 2

Summary of simulation studies with time-varying covariates

Sample Size	β (Time-invariant)						$\hat{\gamma}$ (Time-variant)					
	n	β_0	Method	Bias	SSE	MSE	CP	γ_0	Method	Bias	SSE	MSE
100	0	ME-0	-0.0002	0.175	0.172	0.942	0	ME-0	0.0004	0.173	0.179	0.966
100	0	ME-W	0.0008	0.170	0.169	0.957	0	ME-W	0.0002	0.159	0.167	0.960
100	0	NPMLE	0.0006	0.168	0.163	0.954	0	NPMLE	0.0003	0.158	0.164	0.948
100	1	ME-0	-0.0009	0.548	0.518	0.959	0	ME-0	0.0008	0.270	0.266	0.951
100	1	ME-W	-0.0007	0.529	0.497	0.949	0	ME-W	0.0008	0.233	0.251	0.957
100	1	NPMLE	0.0001	0.509	0.487	0.943	0	NPMLE	-0.0010	0.240	0.246	0.947
100	0	ME-0	-0.0008	0.403	0.395	0.941	1	ME-0	0.0002	0.346	0.345	0.954
100	0	ME-W	0.0007	0.388	0.386	0.936	1	ME-W	-0.0006	0.320	0.320	0.947
100	0	NPMLE	0.0006	0.383	0.379	0.955	1	NPMLE	0.0005	0.315	0.315	0.929
100	1	ME-0	0.0009	0.609	0.596	0.940	1	ME-0	-0.0008	0.355	0.357	0.945
100	1	ME-W	0.0004	0.578	0.576	0.947	1	ME-W	0.0009	0.331	0.331	0.955
100	1	NPMLE	0.0005	0.583	0.562	0.948	1	NPMLE	0.0005	0.326	0.320	0.949
300	0	ME-0	0.0001	0.020	0.018	0.955	0	ME-0	0.0004	0.050	0.058	0.957
300	0	ME-W	0.0004	0.019	0.017	0.957	0	ME-W	0.0003	0.047	0.055	0.936
300	0	NPMLE	0.0003	0.019	0.017	0.951	0	NPMLE	0.0002	0.045	0.053	0.939
300	1	ME-0	-0.0005	0.428	0.429	0.951	0	ME-0	0.0010	0.243	0.250	0.945
300	1	ME-W	0.0007	0.420	0.422	0.956	0	ME-W	0.0006	0.225	0.236	0.956
300	1	NPMLE	0.0002	0.412	0.473	0.946	0	NPMLE	0.0001	0.220	0.227	0.950
300	0	ME-0	0.0005	0.303	0.308	0.952	1	ME-0	-0.0001	0.162	0.169	0.958
300	0	ME-W	0.0001	0.291	0.303	0.939	1	ME-W	0.0004	0.152	0.158	0.923
300	0	NPMLE	0.0004	0.291	0.297	0.948	1	NPMLE	0.0008	0.147	0.153	0.956
300	1	ME-0	0.0001	0.351	0.356	0.953	1	ME-0	-0.0000	0.228	0.228	0.940
300	1	ME-W	0.0009	0.345	0.347	0.945	1	ME-W	0.0007	0.212	0.213	0.957
300	1	NPMLE	-0.0008	0.332	0.337	0.952	1	NPMLE	-0.0003	0.208	0.208	0.937
500	0	ME-0	0.0000	0.089	0.091	0.945	0	ME-0	0.0001	0.041	0.037	0.950
500	0	ME-W	-0.0003	0.085	0.090	0.961	0	ME-W	0.0009	0.038	0.034	0.946
500	0	NPMLE	0.0006	0.084	0.087	0.954	0	NPMLE	0.0008	0.037	0.034	0.969

Sample Size		$\hat{\beta}$ (Time-invariant)					$\hat{\gamma}$ (Time-variant)					
n	β_0	Method	Bias	SSE	MSE	CP	γ_0	Method	Bias	SSE	MSE	CP
500	1	ME-0	0.0002	0.352	0.360	0.950	0	ME-0	0.0006	0.221	0.223	0.955
500	1	ME-W	0.0003	0.348	0.353	0.942	0	ME-W	0.0002	0.210	0.209	0.936
500	1	NPML	-0.0007	0.337	0.341	0.951	0	NPML	0.0003	0.200	0.202	0.948
500	0	ME-0	0.0006	0.251	0.257	0.949	1	ME-0	-0.0005	0.253	0.252	0.948
500	0	ME-W	0.0003	0.242	0.247	0.948	1	ME-W	0.0006	0.236	0.234	0.950
500	0	NPML	0.0005	0.241	0.246	0.956	1	NPML	0.0007	0.232	0.227	0.962
500	1	ME-0	0.0000	0.302	0.300	0.951	1	ME-0	-0.0002	0.187	0.183	0.946
500	1	ME-W	0.0008	0.292	0.291	0.947	1	ME-W	0.0003	0.177	0.170	0.949
500	1	NPML	-0.0001	0.289	0.283	0.943	1	NPML	0.0004	0.168	0.166	0.952

SSE, sample standard error of estimates; ME-0, unweighted moment-based estimates; MEW, weighted moment-based estimates; NPML, nonparametric maximum likelihood estimates; MSE, mean of estimated standard errors; CP, 95% nominal coverage probabilities.

Table 3

Summary of fitting proportional odds models for HIVNET012 data

Parm	Est.	s.d.	CI
NVP	0.265	0.121	(0.027, 0.502)
NVP×Time	-8.52×10^{-5}	6.20×10^{-5}	$(-2.10 \times 10^{-4}, 4.00 \times 10^{-5})$
CD4+ and VL versus Low risk			
Intermediate risk	-0.026	0.018	(-0.061, 0.009)
High risk	-0.052	0.033	(-0.116, 0.012)

NVP, indicator of Nevirapine (versus AZT); Est., parameter estimate; *s.d.*, standard error; CI, 95% confidence interval; Low risk, CD4+>350, VL < 50,000; Intermediate risk, CD4+>350, VL>50,000 or CD4+ < 350, VL < 50,000; High risk, CD4+ < 350, VL>50,000.

Table 4

Summary of fitting proportional odds models for Veteran Administration data

Parm	Model I			Model II			Model III		
	Est.	s.d.	CI	Est.	s.d.	CI	Est.	s.d.	CI
PS	-0.134	0.014	(-0.161, -0.107)	-0.089	0.019	(-0.126, -0.052)	-	-	-
PS×Time	-	-	-	0.016	0.003	(0.010, 0.022)	-	-	-
Tumor type versus Large									
AD	-	-	-	-	-	-	1.292	0.367	(0.573, 2.011)
AD×Time	-	-	-	-	-	-	-	-	-
SM	-	-	-	-	-	-	1.972	0.486	(1.019, 2.925)
SM×Time	-	-	-	-	-	-	-	-	-
SQ	-	-	-	-	-	-	-0.105	0.376	(-0.842, 0.632)
SQ×Time	-	-	-	-	-	-	-	-	-
Model IV									
Parm	Est.	s.d.	CI	Est.	s.d.	CI	Est.	s.d.	CI
PS	-	-	-	-0.059	0.012	(-0.083, -0.035)	-0.049	0.021	(-0.090, -0.008)
PS×Time	-	-	-	-	-	-	-0.008	0.005	(-0.018, 0.002)
Tumor type versus Large									
AD	0.958	0.520	(-0.061, 1.977)	1.521	0.492	(0.557, 2.485)	1.222	0.625	(-0.003, 2.447)
AD×Time	-0.228	0.127	Z(-0.477, 0.021)	-	-	-	-0.169	0.225	(-0.610, 0.272)
SM	1.210	0.671	(-0.105, 2.525)	1.451	0.614	(0.249, 2.655)	1.141	0.781	(-0.390, 2.672)
SM×Time	-0.054	0.232	(-0.509, 0.401)	-	-	-	-0.011	0.223	(-0.449, 0.425)
SQ	0.072	0.591	(-1.086, 1.230)	-0.009	0.563	(-1.112, 1.094)	0.021	0.603	(-1.161, 1.203)
SQ×Time	0.082	0.092	(-0.098, 0.262)	-	-	-	0.071	0.104	(-0.133, 0.275)

PS, performance score; AD, adeno; SM, small; SQ, squamous; Est., parameter estimate; s.d., standard error; CI, 95% confidence interval; -, corresponding covariates excluded in the model.