Published in final edited form as:

Biom J. 2017 May ; 59(3): 405–419. doi:10.1002/bimj.201600003.

Cox proportional hazards models with left truncation and time-varying coefficient: Application of age at event as outcome in cohort studies

Minjin Kim¹, Myunghee Cho Paik^{*1}, Jiyeong Jang², Ying K. Cheung³, Joshua Willey⁴, Mitchell S. V. Elkind⁴, Ralph L. Sacco⁵

¹Department of Statistics, Seoul National University, Seoul, 151-747, Korea

²Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago, IL 60612, USA

³Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY 10032, USA

⁴Department of Neurology, Neurological Institute, Columbia University, New York, NY 10032, USA

⁵Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL 33136, USA

Abstract

When analyzing time-to-event cohort data, two different ways of choosing a time scale have been discussed in the literature: time-on-study or age at onset of disease. One advantage of choosing the latter is interpretability of the hazard ratio as a function of age. To handle the analysis of age at onset in a principled manner, we present an analysis of the Cox Proportional Hazards model with time-varying coefficient for left-truncated and right-censored data. In the analysis of Northern Manhattan Study (NOMAS) with age at onset of stroke as outcome, we demonstrate that well-established risk factors may be important only around a certain age span and less established risk factors can have a strong effect in a certain age span.

Keywords

Estimating equation; Local linear fitting; Profile likelihood; Time-to-event cohort data; Time-varying coefficient

1 Introduction

Time-to-event is defined as the duration from a certain time of origin and the time when the event of interest occurs. While the choice of the endpoint of the duration is straightforward,

*Corresponding author: myungheechopaik@snu.ac.kr, Phone: +82-2-880-6764, Fax: +82-2-883-6144.

Additional supporting information including source code to reproduce the results may be found in the online version of this article at the publisher's web-site

Conflict of interest

The author has declared no conflict of interest.

different choices of the time of origin are possible in longitudinal follow-up studies. The common practice sets the starting point as the recruitment time, and an alternative choice is time at birth. By the first choice, the time scale of analysis is follow-up time or time-on-study, and by the second choice, the age at event. Pencina et al. (2007) addressed the important issue of choosing a time-scale analyzing time-to-event data. Korn et al. (1997) first suggested using time at birth as the point of origin, thus, age at onset of disease being the time scale. One reason why age at onset is a less popular choice is that subjects are not followed from birth and those subjects who died before reaching the recruitment age have no chance to be sampled. This yields left-truncated and right-censored (LTRC) data. Many authors warned that left truncation should be taken care of in the analysis to avoid bias (e.g., Wang et al., 1986; Tsai et al., 1987). To draw a valid inference for the analysis with age at onset as outcome by properly taking the left truncation into account, the left truncation time, and age at onset should be quasi-independent (Tsai, 1990), that is, the two are independent in the observed region. Unlike independence between failure time and censoring, quasi-independence is testable in one sample case (Tsai, 1990) or in regression setting (Cheng and Wang, 2015).

There are advantages choosing age at onset as the time scale in cohort studies. First, the estimated survival function of age at onset of disease can be useful for clinicians and health care professionals. Another advantage is clinical interpretability of time-varying coefficient. In studies with clinically meaningful origins involving actions, such as start of treatment, transplant, or infection, the hazard ratio as a function of follow-up time provides important clinical information on how risk changes since the action point. However in longitudinal follow-up studies where recruitment time is the time of origin, the hazard ratio as a function of follow-up time may not be meaningful. In this case, analysis using age at onset as a time scale provides the effects of risk factor as a function of age, which can be useful in clinical practice. To incorporate this feature in the analysis using age at onset as the time scale in a principled manner, we need survival regression models with time-varying coefficients for LTRC data.

The Cox proportional hazards (PH) model (Cox, 1972) has been frequently used to analyze time-to-event data. While the Cox PH model is a popular survival regression model, the proportionality assumption may be violated for some variables. This implies that the hazard ratio is not constant over time and one remedy to this problem is to consider time-varying coefficient. The Cox model with time-varying coefficient has been studied by many researchers. Zucker and Karr (1990) provided the asymptotic behavior of the maximum penalized partial likelihood estimator with an appropriate penalty on a roughness of the time-varying coefficients. Gray (1992, 1994) proposed a spline-based method with prespecified knots to estimate time-varying coefficients and test proportionality. Martinussen et al. (2002) proposed to circumvent the problem by estimating cumulative coefficient functions. Other techniques for inference include exploiting the histogram sieve by Murphy and Sen (1991), the kernel-weighted partial likelihood approach by Cai and Sun (2003) and Tian et al. (2005), and a multivariable fractional polynomial time approach by Sauerbrei et al. (2007). See also Hennerfeind et al. (2006), Kauermann and Khomski (2006), Kneib and Fahrmeir (2007). Yu and Lin (2010) considered the case where the Cox model with time-

varying coefficients for right censored data when time-constant coefficients are present at the same time and when failure times are correlated.

In this paper, we consider the Cox models with time-varying and time-constant coefficients for the LTRC data. The work is motivated by the Northern Manhattan Study (NOMAS) whose primary goal is to identify risk factors for stroke. Previously, the NOMAS has been analyzed using follow-up time as outcome. Using age at onset as outcome with time-varying coefficients, risk factors can be identified depending on age. When the all coefficients are time-constant, Andersen et al. (1993) showed that the partial likelihood approach for right-censored data can be modified for the LTRC data using a proper risk set adjusting for left truncation times. We propose an estimator adopting the profile approach of Yu and Lin (2010) and Andersen et al. (1993) and derive inferential procedures for the Cox models with time-varying and time-constant coefficient for the LTRC data. While asymptotic properties of the proposed estimator are fairly straightforward, computation is extensive to use in applications. To alleviate this problem, we also present a backfitting estimator that can be computed mainly using built-in R functions. We revisit the NOMAS and conduct the proposed analysis using age at onset as time scale by accommodating left truncation and allowing time-varying coefficient using the two proposed estimators. We illustrate that some controversial risk factors in previous studies may have effects in a certain age span and that well-established risk factors may not have a strong impact in a certain age span.

The remainder of this paper is organized as follows. In Section 2, we describe the probability model, which is followed by discussion of estimation and computational details in Section 3. Section 4 will introduce a formal test for time-constancy of effect of a covariate, so that it can be used to determine which covariates are to be included as time-varying. In Section 5, we report simulation results evaluating finite sample performance of the proposed estimators. We present the motivating example, the NOMAS in Section 6 and conduct analysis using age at stroke onset as time scale by accommodating left truncation with time-varying and time-constant hazard ratios. We offer a few concluding remarks in Section 7. The R codes for necessary computations of the proposed profile estimator and the backfitting estimator are presented in the Supporting Information. For the asymptotic results, we presented in the Supplementary Material.

2 Model

Let T^* , L , and C be age at onset, left truncation and censoring time, respectively. Also let $\mathbf{Z}$ and $\mathbf{X}$ be $q \times 1$ and $p \times 1$ covariates corresponding to time-varying and time-constant coefficients. For notational simplicity, we assume that $q = 1$.

We observe (L_i, T_i, X_i, Z_i) for the i^{th} subject, where $T_i = \min(T_i^*, C_i) > L_i$, $\Delta_i = I(T_i^* \leq C_i)$ is an uncensoring indicator ($i = 1, 2, \dots, n$) and $I(\cdot)$ be the indicator function. We assume that (T^*, L) and C are conditionally independent given $\mathbf{X}$ and $\mathbf{Z}$. We also assume quasi-independence between T^* and L , that is, T^* and L are independent given $(\mathbf{X}, \mathbf{Z})$ in the observed region. Let $N_i(t) = I(T_i \leq t, \Delta_i = 1)$. As in the right-censored data, $dN_i(t) = I(t- < T_i \leq t, \Delta_i = 1)$, and $dN_i(t)$ indicates whether the i^{th} subject fails at t . We denote the at-risk process by $Y_i(t) = I(L_i < t \leq T_i)$. Note that the at-risk process

accommodates left truncation and is different from the case of the right-censored data. Under quasi-independence assumption, analysis for left truncation data requires minor modification of the at-risk process, consequently the risk sets, from the method for the right-censored data. Majority of the methods developed for the LTRC data assume quasi-independence including Andersen et al. (1993) that extended the Cox PH model to the LTRC data. Let $h(t|X_i, Z_i)$ be the hazard function of T^* given X and Z . We assume the hazard function of the form,

$$h(t|X_i, Z_i) = h_0(t) \exp\{Z_i\beta(t) + X_i^T \gamma\}, \quad (1)$$

where $h_0(t)$ is an unspecified baseline hazard function and $\beta(t)$ is a smooth function. In the Cox PH model, it is assumed

$$h(t|X_i, Z_i) = h_0(t) \exp\{Z_i\eta + X_i^T \xi\}, \quad (2)$$

therefore, if $\beta(t)$ is constant and equals η , the model reduces to the Cox PH model. When $\beta(t)$ is a step function with known jump points, the model reduces to the Cox PH model with properly defined time-varying covariates. Under left truncation with time-constant coefficients only, Andersen et al. (1993) showed that a consistent coefficient estimate can be obtained by solving

$$0 = n^{1/2} U_0(\eta, \xi) = \sum_{i=1}^n U_{0i}(\eta, \xi) = \sum_{i=1}^n \int_{\tau_L}^{\tau_U} [Z_i^* - E(\eta, \xi, u)] dN_i(u),$$

where

$$E(\eta, \xi, u) = \frac{\sum_{j=1}^n Y_j(u) Z_j^* \exp(Z_j \eta + X_j^T \xi)}{\sum_{j=1}^n Y_j(u) \exp(Z_j \eta + X_j^T \xi)},$$

τ_L and τ_U are prespecified constants, $\Pr(T_i > \tau_U) > 0$, $\Pr(L_i < \tau_L) > 0$ and $Z_i^{*T} = (Z_i^T, X_i^T)$. Due to quasi-independence assumption, we have $h(t|X_i, Z_i, L_i) = h(t|X_i, Z_i)$. Under quasi-independence assumption, further adjustment of left truncation time as covariate is not necessary once left truncation is adjusted via at-risk process (Cheng and Wang, 2015). Gail et al. (2009) made this point from a practical standpoint.

3 Estimation and computation

When *all* covariates have time-varying coefficients and data are right-censored, Cai and Sun (2003) and Tian et al. (2005) proposed the kernel smoothing partial likelihood method. The main idea of kernel smoothing partial likelihood is to approximate $\beta(t)$ by a linear function in a window around each time point t using the observed failure times and the risk sets within the window. The estimated linear function at t is the estimate of the time-varying coefficient function at t . When covariates are mixed with time-varying and time-constant coefficients, Tian et al. (2005) proposed to estimate first as time-varying coefficients and

then integrate over the time-varying coefficient estimate to obtain time-constant coefficient estimate, γ . Yu and Lin (2010) proposed profile likelihood approach for γ and alternate estimating $\beta(t)$ and γ given that the other is fixed at the current estimate. We propose to adopt the approach of the kernel profile partial likelihood method by Yu and Lin (2010) after modifying for left truncation. Having time-constant coefficients in the model adds complexity since the estimate of the time-varying coefficient estimates depends on the time-constant estimates.

Let $\tilde{Z}_i(u, t) = (1, u - t)^T \otimes Z_i$, where $\otimes$ is the Kronecker product. The kernel profile estimation procedure is as follows:

- I. For given γ and fixed time point t , we maximize the kernel log-partial likelihood with respect to $\mathbf{b} = (b_0, b_1)^T$,

$$pl_1(\mathbf{b}, t; \gamma) = (nh_n)^{-1} \sum_{i=1}^n \int_{\tau_L}^{\tau_U} K\{(u-t)/h_n\} \left[\tilde{Z}_i(u, t)^T \mathbf{b} + X_i^T \gamma + \log \left\{ \sum_{j=1}^n Y_j(u) \exp(\tilde{Z}_j(u, t)^T \mathbf{b} + X_j^T \gamma) \right\} \right] dN_i(u), \quad (3)$$

where $h_n = O(n^{-\nu})$ with $\nu > 0$ is the bandwidth that controls the size of the neighborhood, and the kernel function $K(\cdot)$ is a symmetric probability function with support $[-1, 1]$, mean 0, and bounded first derivative. Specifically, for given γ and fixed time point t , we solve the following kernel estimating equation with respect to $\mathbf{b}$,

$$0 = U_1(\mathbf{b}, t; \gamma) = (nh_n)^{-1/2} \sum_{i=1}^n U_{1i}(\mathbf{b}, t; \gamma),$$

where

$$U_{1i}(\mathbf{b}, t; \gamma) = \int_{\tau_L}^{\tau_U} K\{(u-t)/h_n\} [\tilde{Z}_i(u, t) - E(\mathbf{b}, u; \gamma)] dN_i(u)$$

and

$$E(\mathbf{b}, u; \gamma) = \frac{\sum_{j=1}^n Y_j(u) \tilde{Z}_j(u, t) \exp(\tilde{Z}_j(u, t)^T \mathbf{b} + X_j^T \gamma)}{\sum_{j=1}^n Y_j(u) \exp(\tilde{Z}_j(u, t)^T \mathbf{b} + X_j^T \gamma)}.$$

For $t \in [\tau_L + h_n, \tau_U - h_n]$, the resulting estimator of $\beta(t)$ is $\hat{\beta}(t, \gamma) = \hat{b}_0$. Similarly to Tian et al. (2005), for $t < \tau_L + h_n$ we let $\hat{\beta}(t) = \hat{\beta}(\tau_L + h_n)$, and $t > \tau_U - h_n$, $\hat{\beta}(t) = \hat{\beta}(\tau_U - h_n)$, so that the asymptotic properties can be derived over $[\tau_L, \tau_U]$.

- I. For given the kernel estimator $\hat{\beta}(t, \gamma)$, we maximize the profile log-partial likelihood with respect to γ ,

$$pl_2(\gamma, \hat{\beta}(u, \gamma)) = n^{-1} \sum_{i=1}^n \int_{\tau_L}^{\tau_U} \left[Z_i \hat{\beta}(u, \gamma) + X_i^T \gamma + \right. \\ \left. - \log \left\{ \sum_{j=1}^n Y_j(u) \exp(Z_j \hat{\beta}(u, \gamma) + X_j^T \gamma) \right\} \right] dN_i(u), \quad (4)$$

that is, to solve the following profile estimating equation with respect to γ ,

$$0 = U_2(\gamma, \hat{\beta}(u, \gamma), \hat{\beta}_\gamma(u, \gamma)) = n^{-1/2} \sum_{i=1}^n \int_{\tau_L}^{\tau_U} [Z_i \hat{\beta}_\gamma(u, \gamma) + X_i - E(\gamma, u)] dN_i(u),$$

where

$$E(\gamma, u) = \frac{\sum_{j=1}^n Y_j(u) (Z_j \hat{\beta}_\gamma(u, \gamma) + X_j) \exp(Z_j \hat{\beta}(u, \gamma) + X_j^T \gamma)}{\sum_{j=1}^n Y_j(u) \exp(Z_j \hat{\beta}(u, \gamma) + X_j^T \gamma)}$$

and $\hat{\beta}_\gamma(u, \gamma)$ is the first derivative of $\hat{\beta}(u, \gamma)$ with respect to γ .

We iterate above two steps until convergence and obtain the kernel estimator $\hat{\beta}(t, \hat{\gamma})$ and the profile estimator $\hat{\gamma}$ for a fixed bandwidth h . To solve these estimating equations, we can use the Newton–Raphson algorithm. Note that the partial likelihood functions used in two steps are similar but different from Yu and Lin (2010) in the definition of $Y_A(u)$. Yu and Lin (2010) also pointed that $pl_2(\cdot)$ is concave function in a small neighborhood of γ_0 asymptotically using the theorems in Andersen and Gill (1982).

Under the assumptions given in the Supplementary Material, it can be shown that $\hat{\beta}(t)$ is consistent for $\beta_0(t)$, $(nh_n)^{1/2} \{ \hat{\beta}(t) - \beta_0(t) \} \approx I^{-1}(\hat{\beta}(t), t, \gamma_0) U(\beta_0(t), t, \gamma_0)$, where

$$U(\beta(t), t, \gamma) = (nh_n)^{-1/2} \sum_{i=1}^n \int_{\tau_L}^{\tau_U} K\{(u-t)/h_n\} [Z_i - E(\beta(t), u, \gamma)] dN_i(u),$$

$$E(\beta(t), u, \gamma) = \frac{S^{(1)}(\beta(t), u, \gamma)}{S^{(0)}(\beta(t), u, \gamma)},$$

$$S^{(k)}(\beta(t), u, \gamma) = n^{-1} \sum_{j=1}^n Y_j(u) Z_j^k \exp\{Z_j \beta(t) + X_j^T \gamma\},$$

$k = 0, 1, 2$ and

$$I(\hat{\beta}(t), t, \gamma) = (nh_n)^{-1} \sum_{i=1}^n \int_{\tau_L}^{\tau_U} \left[\frac{S^{(2)}(\hat{\beta}(t), s, \gamma)}{S^{(0)}(\hat{\beta}(t), s, \gamma)} - \left\{ \frac{S^{(1)}(\hat{\beta}(t), s, \gamma)}{S^{(0)}(\hat{\beta}(t), s, \gamma)} \right\}^2 \right] K\{(s-t)/h_n\} dN_i(s).$$

A consistent estimator of the variance of $(nh_n)^{1/2}\hat{\beta}(t)$ can be obtained by $I(\hat{\beta}(t), t, \hat{\gamma})^{-1} \int K^2(s) ds$. Also, under the assumptions given in the Supplementary Material it can be shown that $\hat{\gamma}$ converges in probability to γ_0 . Evaluation of $\left\{ -\frac{\partial^2}{\partial \gamma \partial \gamma^T} pl_2(\gamma) \right\}$ requires computing $\hat{\beta}_{\gamma\gamma}(t, \gamma)$, the second derivative of $\hat{\beta}(t, \gamma)$ with respect to γ . A consistent estimator of the variance of $\hat{\gamma}$ can be obtained by $\left\{ -\frac{\partial^2}{\partial \gamma \partial \gamma^T} pl_2(\gamma) \right\}^{-1}$ evaluated at $\hat{\gamma}$, $\hat{\beta}(t, \hat{\gamma})$, $\hat{\beta}_{\gamma}(t, \hat{\gamma})$, and $\hat{\beta}_{\gamma\gamma}(t, \hat{\gamma})$. Note that the estimator of γ requires computation of $\hat{\beta}_{\gamma}(t, \gamma)$, and its estimator of the variance needs additional computation of $\hat{\beta}_{\gamma\gamma}(t, \gamma)$. We provide R codes in the Supporting Information.

We also present a backfitting estimator of γ that solves

$$0 = n^{-1/2} \sum_{i=1}^n \int_{\tau_L}^{\tau_U} \left[X_i - \frac{\sum_{j=1}^n Y_j(u) X_j \exp(Z_j \hat{\beta}(u, \gamma) + X_j^T \gamma)}{\sum_{j=1}^n Y_j(u) \exp(Z_j \hat{\beta}(u, \gamma) + X_j^T \gamma)} \right] dN_i(u),$$

while $\hat{\beta}(t)$ is estimated the same as the profile estimator. Unlike the profile method, the backfitting method does not require $\hat{\beta}_{\gamma}(u, \gamma)$ and $\hat{\beta}_{\gamma\gamma}(u, \gamma)$ in computation. We provide R codes in the Supporting Information to compute the backfitting estimator for the NOMAS data. This computation can be carried out by built-in R functions `survSplit` and `coxph` using kernel functions as weight, treating terms with time-varying coefficients as offset in computing $\hat{\gamma}$, and terms with $\hat{\gamma}$ as offsets in computing $\hat{\beta}(t)$. Theoretical comparisons of the two methods in partially linear models are studied when units are independent (Opsomer and Ruppert, 1999; Van Keilegom and Carroll, 2007) and correlated (Hu et al., 2004). In case of independent units, Van Keilegom and Carroll (2007) shows that the two methods yield the estimators with the same limiting distribution under broad conditions when undersmoothing of $\hat{\beta}(t)$ is employed. Section 5 reports a numerical study to investigate the performance of the two methods.

For the bandwidth selection, Cai and Sun (2003) showed the analytic form of the mean integrated squared error and obtained the optimal bandwidth, say h_{opt} , as its minimizer. Since the mean integrated squared error is a function of unknown quantities, the bandwidth should be chosen empirically. Tian et al. (2005) proposed to use a $\mathcal{K}$ -fold cross-validation method and we adopt this procedure. The data are divided into $\mathcal{K}$ equal-sized parts and the estimates are obtained by deleting the k^{th} part at a time where $k = 1, 2, \dots, \mathcal{K}$. Since the predicted partial likelihood is contributed by uncensored data only, we recommend to divide the data stratified by uncensoring indicator. Then we fit the model using the other $(\mathcal{K} - 1)$

parts with a fixed bandwidth h and calculate the minus predicted partial likelihood denoted by $\text{PPL}_k(h)$, where

$$\text{PPL}_k(h) = - \sum_{l \in I_k} \int_{\tau_L}^{\tau_U} \left[Z_l \hat{\beta}_{-k}(u) + X_l^T \hat{\gamma}_{-k} - \log \left(\sum_{r \in I_k} Y_r(u) \exp(Z_r \hat{\beta}_{-k}(u) + X_r^T \hat{\gamma}_{-k}) \right) \right] dN_l(u),$$

I_k is the index set of the k^{th} part, and the subscript “ $-k$ ” denotes the estimate computed from the data without the k^{th} part. Argument h emphasizes that $\text{PPL}_k(h)$ is a function of bandwidth h . The bandwidth, $\hat{h}_{opt}$ is selected that minimizes $\text{PPL}(h) = \sum_{k=1}^{\mathcal{K}} \text{PPL}_k(h)$. In addition, we consider the following modified version of $\text{PPL}(h)$ criteria. A rational is that asymptotically $\text{PPL}(h)$ is equivalent to a sum of equally weighted Mahalanobis distance between the observed versus expected covariate of the failed unit over the risk set in the k^{th} part of the data. Each term of $\text{PPL}(h)$ represents goodness of fit at each observed time point. With left-truncated data, a large deviation could occur at early time after τ_L . To avoid penalizing poor fit where data are sparse, we propose to weigh each observation proportional to the risk set size as follows:

$$\text{PPL}_k^*(h) = - \sum_{l \in I_k} \int_{\tau_L}^{\tau_U} \frac{Y_l(u)}{Y(u^*)} \left[Z_l \hat{\beta}_{-k}(u) + X_l^T \hat{\gamma}_{-k} - \log \left(\sum_{r \in I_k} Y_r(u) \exp(Z_r \hat{\beta}_{-k}(u) + X_r^T \hat{\gamma}_{-k}) \right) \right] dN_l(u),$$

where $Y(u) = \sum_{i=1}^n Y_i(u)$ and u^* is the median follow-up time. If the size of the risk set is very small due to truncation, the weight is close to zero, thus a large deviation at that time point would be ignored in $\text{PPL}_k^*(h)$. We computed the bandwidth $\hat{h}_{opt}^*$ that minimizes $\text{PPL}^*(h) = \sum_{k=1}^{\mathcal{K}} \text{PPL}_k^*(h)$. Computation of $\text{PPL}_k^*(h)$ and $\text{PPL}_k(h)$ could be demanding for the proposed profile estimator with R, and in Section 6, we implemented it in MATLAB, reducing the computational time by 20%.

For the estimation of the cumulative baseline hazard function $H_0(t) = \int_{\tau_L}^t h_0(s) ds$, we propose Breslow–Aalen type estimator as in Yu and Lin (2010) for the model (1) adjusting for left truncation,

$$\hat{H}_0(t) = \sum_{i=1}^n \int_{\tau_L}^t \frac{dN_i(u)}{\sum_{j=1}^n Y_j(u) \exp\{Z_j \hat{\beta}(u) + X_j^T \hat{\gamma}\}},$$

where $\hat{\beta}(t)$ and $\hat{\gamma}$ are the kernel and profile estimator, respectively.

4 Testing proportionality

In this section, we propose a numerical method to test whether the model with time-varying coefficient is the correct model, that is $\beta(t) = \eta$. One may check whether the confidence bands of $\hat{\beta}(t)$ exclude horizontal line $\hat{\eta}$, but this procedure would not be powerful due to slow convergence of $\hat{\beta}(t)$. The proposed method below is extended from Tian et al. (2005) to

accommodate the LTRC data. Adjustment for the presence of time-constant coefficients is not necessary, due to the result by Yu and Lin (2010) that $(nh_n)^{1/2}\{\hat{\beta}(t, \hat{\gamma}) - \beta(t, \gamma_0)\} = (nh_n)^{1/2}\{\hat{\beta}(t, \gamma_0) - \beta(t, \gamma_0)\} + o_p(1)$, where $\beta(t, \gamma_0) = \beta(t)$. Tian et al. (2005) examined the validity of the PH assumption for each covariate by using the $\hat{B}(t)$, where $\hat{B}(t) = \int_0^t \hat{\beta}(s) ds$. They showed that the process $\sqrt{n}\{\hat{B}(t) - B_0(t)\}$ converges weakly to a mean-zero Gaussian process under some regularity conditions, where $B_0(t) = \int_0^t \beta_0(s) ds$. The test by Tian et al. (2005) assumes the rest of the covariates have time-varying effects. We consider a test assuming the rest of the covariates have time-constant effects. Let $\Gamma(t) = n^{1/2} \int_{\tau_L}^t + h_n \{\hat{\beta}(s, \hat{\gamma}) - \hat{\eta}\} ds$, where $t \in [b_1, b_2] \subseteq [\tau_L + h_n, \tau_U - h_n]$. Following the arguments of Tian et al. (2005), we can verify that $\Gamma(t)$ converges weakly to a mean zero Gaussian process under H_0 . Using (A.5) of Tian et al. (2005) (Eubank 1988, p.128) and the stochastic perturbation technique one can show that the distribution of limiting process can be approximated by the distribution of

$$\begin{aligned} \hat{\Gamma}(t) = & n^{-1/2} \sum_{i=1}^n \int_{\tau_L}^{\tau_U} \left[h_n^{-1} \left[\int_{\tau_L + h_n}^t K\{(s-u)/h_n\} du \right] I^{-1}(\hat{\beta}(s), s, \hat{\gamma}) \right. \\ & \left. \{Z_i - E(\hat{\beta}(s), s, \hat{\gamma})\} + \right. \\ & \left. - \{I_0^{-1}(\hat{\eta}, \hat{\xi})(Z_i^* - E(\hat{\eta}, \hat{\xi}, s))\}_1 (t - \tau_L - h_n) \right] dN_i(s) G_i, \end{aligned} \quad (5)$$

where $\{v\}_1$ is the first component of vector v , $I_0(\eta, \xi) = n^{-1} \sum_{i=1}^n 1 - \partial U_{0i}(\eta, \xi) / \partial(\eta, \xi^T)$, and $\{G_i\}_{i=1}^n$ is a random sample from the standard normal distribution, and is independent of $\{(L_i, T_i, X_i, Z_i, \Delta_i)\}_{i=1}^n$. Using the fact that (5) is a sum of independent mean zero random variables, a consistent estimate of standard error $\hat{\sigma}(t)$ of $\Gamma(t)$ can be obtained.

Let K be a number of resampling in Bootstrap. At the k^{th} replication, we define $\hat{\Gamma}^{(k)}$ by replacing G_i with $G_i^{(k)}$, where $\{G_i^{(k)}: i = 1, \dots, n\}$ are random samples from the standard Gaussian. Also let $S_n = \sup_t |\Gamma(t)/\hat{\sigma}(t)|$, $\hat{S}_n = \sup_t |\hat{\Gamma}(t)/\hat{\sigma}(t)|$ and s_n is the realization of S_n . The p -value, $\Pr(S_n \leq s_n)$, can be approximated by $\Pr(\hat{S}_n \geq s_n)$. We can estimate $\Pr(\hat{S}_n \geq s_n)$ by $\sum_{k=1}^K I(\hat{S}_n^{(k)} \geq s_n) / K$, where $\hat{S}_n^{(k)} = \sup_t |\hat{\Gamma}^{(k)}(t)/\hat{\sigma}(t)|$. A general discussion on model checking and selection technique was given by Lin et al. (2002).

5 Simulation study

We conducted a small simulation study to evaluate the finite sample properties of the profile and the backfitting estimators. The key difference is that the profile estimate takes account of $\hat{\beta}_{\gamma}(t, \hat{\gamma})$, in estimating γ , while the backfitting estimate ignores it. The supplementary material of Yu and Lin (2010) showed that $\hat{\beta}_{\gamma}(t, \hat{\gamma})$ can be expressed as the covariance between Z and X given the follow-up time and the censoring indicator. We considered three different values of $\hat{\beta}_{\gamma}(t, \hat{\gamma})$, through difference covariance between Z and X . We set

$$h(t|\mathbf{X}, Z) = h_0(t) \exp\{Z\beta(t) + \mathbf{X}^T\boldsymbol{\gamma}\},$$

where $\mathbf{X}^T = (X_1, X_2)$, $\boldsymbol{\gamma} = (1, 0.5)^T$ and $h_0(t) = 1$. The censoring time (C) and the left truncation time (L) were generated independently, and L was from the standard exponential distribution. The covariates X_1 , X_2 , and Z were generated from $N(0, 0.5)$, Bernoulli(0.5), and Uniform(0, 1), respectively, and X_1 and (X_2, Z) were set to be independent. Three different values of $\text{Corr}(X_2, Z)$ were set. For the case $\text{Corr}(X_2, Z) = 0$, we considered different functions of $\beta(t)$. The scenarios we considered are as follows:

$$(S1.1) \text{Corr}(X_2, Z) = 0, C \sim U(0, 2), \beta(t) = t.$$

$$(S1.2) \text{Corr}(X_2, Z) = 0, C \sim U(0, 3), \beta(t) = \pi^{-1/2} \exp\{-16(t-1)^2/2\}.$$

$$(S2) \text{Corr}(X_2, Z) = 0.3, C \sim U(0, 2.5), \beta(t) = t.$$

$$(S3) \text{Corr}(X_2, Z) = 0.7, C \sim U(0, 2.5), \beta(t) = t.$$

The percentages of left truncation from initially generated data for (S1.1), (S1.2), (S2), and (S3) were 40.2%, 43.0%, 39.9%, and 39.9%, respectively. Two hundred replications were conducted with a sample size of $n = 500$. The percentages of censored observations were around 37% for (S1.1), 32% for (S1.2) and 30% for (S2) and (S3). We used the Epanechnikov kernel. To reduce computational demand, we used an efficient iterative algorithm proposed by Cai et al. (2000). Computational time for the profile estimator took about 2.5 times longer than the backfitting algorithm using the computer Intel(R) Core(TM) i5-4460 CPU with 3.20GHz. We considered bandwidths ($h = 0.3, 0.5, 0.7$) for backfitting and profile estimates for scenario (S3). For other scenarios, we used bandwidths ($h = 0.3, 0.5, 0.7$) for backfitting and $h = 0.5$ for profile estimate, and we report the case $h = 0.5$. The results with $h = 0.3, 0.7$ were similar to those with $h = 0.5$ for both backfitting and profile estimates.

For $\hat{\gamma}$, we computed empirical bias, empirical standard errors (SE), estimated standard errors (eSE), and the associated 95% confidence coverage rate (cov. rate). For $\hat{\beta}(t)$, we evaluated the mean absolute deviation (MAD), where

$$\text{MAD} = n_{\text{grid}}^{-1} \sum_{l=1}^{n_{\text{grid}}} |\hat{\beta}(t_l) - \beta_0(t_l)|,$$

and $\{t_l, l = 1, \dots, n_{\text{grid}}\}$ are the grid points at which $\beta(\cdot)$ were estimated. The empirical coverage probabilities of the pointwise 95% confidence intervals were computed.

The simulation results are summarized in Tables 1 and 2. The two methods show similar MAD and coverage probabilities for the estimate of $\beta(t)$. Table 1 shows that results of the two methods for $\boldsymbol{\gamma}$ are similar when the covariates are independent or the size of correlation is 0.3. For scenario (S3) when $\text{Corr}(X_2, Z) = 0.7$, the profile estimate of $\boldsymbol{\gamma}_2$ had slightly smaller SE than the backfitting counterpart. More importantly, the coverage probability was

nominal for the profile estimate, but for the backfitting estimator, was significantly lower than the nominal value. This may suggest that when the correlation between the covariates with time-varying coefficient and time-constant coefficient is small, the backfitting estimator can be satisfactory, but when the correlation is high, the profile estimator for γ may be needed to obtain proper interval estimation.

6 Analysis of the NOMAS data

The NOMAS is a population-based prospective cohort study designed to evaluate the effects of medical, socio-economic, and other risk factors on the incidence of vascular disease in a stroke-free multiethnic community cohort. Participants were identified by dual-frame random digit dialing in the Northern Manhattan community and were eligible if they met the following criteria: (1) had never been diagnosed with a stroke; (2) were over the age of 39 years; and (3) resided in Northern Manhattan for >3 months in a household with a telephone. The cohort was collected from 1993 to 2001, and a total of 3298 participants were recruited. The average age was 69.2 years, and 62.9% were women; 54.2% of the cohort was Hispanic, 25.0% non-Hispanic black, and 20.8% non-Hispanic white. At the time of the analysis, out of 3298 subjects, 269 had a stroke of all types. The total number of deaths was 1069 and among those 158 were death due to stroke. Other causes of death were treated as censored and the analysis represents a cause-specific Cox regression model.

The NOMAS has identified risk factors for stroke using Cox PH models with time-on-study as time scale. We fitted the Cox PH model with age at onset of all type of stroke as time scale and the entry age to the study as left truncation time. The lowest truncation time was 40 (τ_L) and the highest censoring time was 109.1 (τ_U). Table 3 shows coefficient estimates for the “final” model in Willey et al. (2010) using age as time scale alongside with the estimates from the analysis using time-on-study as time scale.

Covariates included in the model are Moderate to heavy physical activity (actmodeheavy), Moderate alcohol consumption (etmod), Any cardiac disease (NCAD), Black race-ethnicity (BLACK), Hispanic race-ethnicity (HISP), Male gender (MALE), Systolic blood pressure (SYSTOLIC), Baseline diabetes mellitus (DIABETES), Former smoker (SMOKER1), Current smoker (SMOKER2), and logarithms of HDL (LHDL) and LDL (LLDL). These two time-scale analyses yield similar results. Physical activity has been controversial in terms of whether it is protective for stroke using time-on-study as time scale. Since stroke patients tend to be old, a question arises whether actmodeheavy is beneficial at this age. To investigate this question, we used actmodeheavy as time-varying effect and other variables in the Table 3 as time-constant.

For bandwidth selection, we conducted the 10-fold cross-validation analysis and found $\hat{h}_{opt} = 15$ and $\hat{h}_{opt}^* = 12.5$ based on the PPL and PPL* criteria, respectively. Figure 1 shows the effect of physical activity (actmodheavy) as a function of age with bandwidths $\hat{h}_{opt}^* = 12.5$ and $\hat{h}_{opt} = 15$, and the effect was protective in age interval (68.9, 73.5) and (69.8, 72.1), respectively. The confidence interval was somewhat wider for $\hat{h}_{opt}^* = 12.5$ than for $\hat{h}_{opt} = 15$ near τ_L and τ_U . As shown in Table 3 we found the effect of physical activity to

be not significant when the variable was forced to be time-constant; however, after allowing the effect to be time-varying, we identified the pattern of the effect over age and the age span where the effect is protective.

Table 4 shows the estimates of the time-constant coefficients, where the effect of actmodheavy is admitted to be time-varying with $\hat{h}_{opt} = 15$. In the left half of the Table 4 we present the profile estimator for γ , and on the right half, the backfitting estimator. In this dataset, the two estimators and their standard errors were very close.

Diabetes was identified as a strong significant factor when all the effect was assumed to be constant as shown in Table 3. Figure 3 shows the time-varying coefficient estimates for diabetes as a function of age for time interval $[\tau_L + h_m, \tau_U - h_n]$ using bandwidths $h = 12.5, 15, 17.5$, and 20 while other effects of covariates in Table 3 as time-constant. The effects obtained from all bandwidths show a similar decreasing trend with the advance of age and the estimated time-varying coefficients were insensitive with respect to the choice of bandwidth. Although not reported, the estimates of time-constant coefficients were similar when different bandwidths were selected.

We tested whether each of the variables has constant effect over age. All plots in Fig. 4 describe $\{\hat{\Gamma}(t)/\hat{\sigma}(t)\}$ when $q = 1$ with $h = 15$. The background curves represent 50 curves out of 1000 generated curves as described in Section 4 and the bold curve is the curve from the observed data. The figures show that when the effect is not time-constant, the curves from the observed data exhibit unusual patterns. The approximated p -value can be calculated as we mentioned in Section 4. The time-constant effects were rejected at 0.05 level for actmodheavy ($p = 0.026$), DIABETES ($p = 0.02$), etmod ($p = 0.03$), MALE ($p < 0.001$), NCAD ($p = 0.004$), SMOKER1 ($p = 0.046$), while the effects of LHDL ($p = 0.189$), SMOKER2 ($p = 0.067$), and SYSTOLIC ($p = 0.688$) did not display a strong evidence to be time-varying.

Figure 2 shows that the effects of selected variables as a function of age with $h = 15$ when we put each variable as time-varying and the rest as time-constant ($q = 1$). We can verify that the effects of Systolic blood pressure (SYSTOLIC), Current smoker (SMOKER2), and HDL (LHDL) did not change much over age. Among the variables where the effects are rejected to be time-constant, DIABETES, etmod, MALE, and NCAD were significant as time-constant effect in Table 3. The figure suggests that even for well-known potent risk factors such as DIABETES, NCAD, and MALE may not have significant effects after mid-80's. On the contrary, the effect of etmod may become more protective in later years. It is noteworthy that former smokers (SMOKER1) do not have a significantly elevated risk when the effect is specified as time-constant, but the confidence interval for the log hazard ratio excludes zero between ages (62.55, 72.64) when allowed to be time-varying. The estimated risk of current smoking (SMOKER2) stays high for a prolonged period of age.

We finally fitted a model specifying six covariates identified via hypothesis testing as time-varying with $h = 15$ and the rest of the variable as time-constant ($q = 6$). Both estimates of time-varying and time-constant estimates were similar to those obtained when each of the variable was specified as time-varying.

7 Discussion

We have conducted analysis of the NOMAS data using age as time scale with time-varying coefficient using backfitting and profile methods. While the backfitting approach yields a biased estimator of time-constant coefficient in general, it has a substantial computational advantage of carrying out using built-in R functions as in Supporting Information. When the covariance between the covariates of time-varying versus time-constant coefficient is substantial, bias can be nonnegligible, and the profile approach is preferred.

Our analysis has demonstrated that the effect of moderate to heavy physical activity could be protective in a certain age span and the detrimental effect of a well-established risk factor, such as diabetes, could diminish at an older age. When coupled with age at onset as time scale of analysis adjusting for left truncation, the Cox time-varying coefficient model can serve as a useful analytic tool providing new clinical insights in longitudinal cohort studies. In our analysis, we tested a single time-varying coefficient at a time in a model where other covariates have time-constant coefficients. Alternatively, one of the referees suggested conducting the same test where other covariates have time-varying coefficients, and comparing these two, which could be a future topic for further research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) for Paik (No. 2013R1A2A2A01067262 and No. 2011-0030810), and National Institute of Health of United States for Sacco, Elkind, Willey, and Cheung (R01 NS 29993), for Cheung (R01 HL111195), and for Willey (NINDS K23 NS 073104).

References

- Andersen PK, Borgan O, Gill RD and Keiding N (1993). Statistical Models Based on Counting Processes. Springer, New York, NY.
- Andersen PK and Gill RD (1982). Cox's regression model for counting processes: a large sample study. The annals of statistics 10, 1100–1120.
- Cai Z, Fan J and Li R (2000). Efficient estimation and inferences for varying-coefficient models. Journal of the American Statistical Association 95, 888–902.
- Cai Z and Sun Y (2003). Local linear estimation for time-dependent coefficients in cox regression models. The Scandinavian Journal of Statistics 30, 93–111.
- Cheng YJ and Wang MC (2015). Causal estimation using semiparametric transformation models under prevalent sampling. Biometrics 71, 302–312. [PubMed: 25715045]
- Cox DR (1972). Regression models and life-tables. Journal of the Royal Statistical Society Series B 34, 187–220.
- Gail MH, Graubard B, Williamson DF and Flegal KM (2009). Comment on choice of time scale and its effect on significance of predictors in longitudinal studies. Statistics in Medicine 28, 1315–1317. [PubMed: 19288532]
- Gray RJ (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. Journal of the American Statistical Association 87, 942–951.
- Gray RJ (1994). Spline-based tests in survival analysis. Biometrics 50, 640–652. [PubMed: 7981391]

- Hennerfeind A, Brezger A and Fahrmeir L (2006). Geoadditive survival models. *Journal of the American Statistical Association* 101, 1065–1075.
- Hu Z, Wang N and Carroll RJ (2004). Profile-kernel versus backfitting in the partially linear models for longitudinal/clustered data. *Biometrika* 91, 251–262.
- Kauermann G and Khomski P (2006). Additive two-way hazards model with varying coefficients. *Computational statistics and data analysis* 51, 1944–1956.
- Kneib T and Fahrmeir L (2007). A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics* 34, 207–228.
- Korn EL, Graubard BI and Midthune D (1997). Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *American Journal of Epidemiology* 145, 72–80. [PubMed: 8982025]
- Lin D, Wei L and Ying Z (2002). Model-checking techniques based on cumulative residuals. *Biometrics* 58, 1–12. [PubMed: 11890304]
- Martinussen T, Scheike TH and Skovgaard IM (2002). Efficient estimation of fixed and timevarying covariate effects in multiplicative intensity models. *The Scandinavian Journal of Statistics* 29, 57–74.
- Murphy S and Sen P (1991). Time-dependent coefficients in a cox-type regression. *Stochastic Processes and their Applications* 39, 153–180.
- Opsomer JD and Ruppert D (1999). A root-n consistent backfitting estimator for semiparametric additive modeling. *Journal of Computational and Graphical Statistics* 8, 715–732.
- Pencina MJ, Larson MG and D'Agostino RB (2007). Choice of time scale and its effect on significance of predictors in longitudinal studies. *Statistics in Medicine* 26, 1343–1359. [PubMed: 16955538]
- Sauerbrei W, Royston P and Look M (2007). A new proposal for multivariable modeling of time-varying effects in survival data based on fractional polynomial time-transformation. *Biometrical Journal* 49, 453–473. [PubMed: 17623349]
- Tian L, Zucker D and Wei LJ (2005). On the cox model with time-varying regression coefficients. *Journal of the American Statistical Association* 100, 172–183.
- Tsai WY (1990). Testing the assumption of independence of truncation time and failure time. *Biometrika* 77, 169–177.
- Tsai WY, Jewell NP and Wang MC (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* 74, 883–886.
- Van Keilegom I and Carroll RJ (2007). Backfitting versus profiling in general criterion functions. *Statistica Sinica* 17, 797–816.
- Wang MC, Jewell NP and Tsai WY (1986). Asymptotic properties of the product limit estimate under random truncation. *The Annals of Statistics* 14, 1597–1605.
- Willey JZ, Paik MC, Sacco R, Elkind MS and Boden-Albala B (2010). Social determinants of physical inactivity in the northern manhattan study (nomas). *Journal of Community Health* 35, 602–608. [PubMed: 20574777]
- Yu Z and Lin X (2010). Semiparametric regression with time-dependent coefficients for failure time data analysis. *Statistica Sinica* 20, 853–869. [PubMed: 20514365]
- Zucker DM and Karr AF (1990). Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *The Annals of Statistics* 18, 329–353.

Figure 1.

Estimates of time-varying coefficient for moderate to heavy physical activity as a function of age and 95% pointwise confidence interval. The blue solid and red dotted are for $h = 15$. And the solid and dotted black are for $h = 12.5$.

Figure 2. Estimates of time-varying coefficients as a function of age and 95% pointwise confidence interval when $h = 15$. The dotdash line indicates the estimates under Cox PH model (2).

Figure 3.
Estimates of time-varying coefficient for diabetes as a function of age using different bandwidths.

Figure 4.

Testing proportionality of the coefficients. The solid-curves are the observed processes of the standardized $\Gamma(t)$ and gray curves are the generated counterpart.

Table 1Simulation results for γ .

Scenarios	γ	Backfitting				Profiling			
		BIAS	eSE	SE	cov. rate	BIAS	eSE	SE	cov. rate
(S1.1)	X_1	-0.046	0.093	0.077	0.955	-0.048	0.093	0.077	0.955
	X_2	-0.021	0.115	0.118	0.955	-0.022	0.115	0.118	0.955
(S1.2)	X_1	0.010	0.122	0.111	0.985	0.012	0.123	0.111	0.985
	X_2	0.010	0.111	0.115	0.945	0.012	0.111	0.116	0.945
(S2)	X_1	0.010	0.121	0.116	0.960	0.012	0.123	0.117	0.960
	X_2	0.002	0.110	0.115	0.955	0.014	0.113	0.113	0.960
(S3)	X_1	0.009	0.122	0.126	0.945	0.008	0.123	0.125	0.950
	X_2	-0.012	0.112	0.133	0.915	0.017	0.133	0.131	0.955

Table 2Simulation results for $\beta(t)$.

Scenarios	Backfitting		Profiling	
	MAD	cov. rate	MAD	cov. rate
(S1.1)	0.300	0.933	0.300	0.934
(S1.2)	0.304	0.953	0.304	0.953
(S2)	0.414	0.945	0.414	0.945
(S3)	0.442	0.915	0.438	0.917

Table 3

Age at onset versus time-on-study as time scale.

Variable	Age at onset as time scale			Time-on-study as time scale		
	Estimate	s.e.	p-value	Estimate	s.e.	p-value
Age	–	–	–	0.050	0.007	<0.001
actmodheavy	–0.212	0.247	0.391	–0.194	0.247	0.432
etmod	–0.283	0.144	0.049	–0.289	0.144	0.045
NCAD	0.399	0.133	0.003	0.384	0.133	0.004
BLACK	–0.360	1.303	0.782	–0.430	1.303	0.742
HISP	–2.267	1.189	0.057	–2.337	1.187	0.049
MALE	0.449	0.140	0.001	0.453	0.140	0.001
SYSTOLIC	0.001	0.007	0.840	0.001	0.007	0.878
DIABETES	0.815	0.130	<0.001	0.830	0.130	<0.001
SMOKER1	0.148	0.144	0.303	0.150	0.143	0.295
SMOKER2	0.480	0.177	0.007	0.473	0.177	0.008
LHDL	0.004	0.004	0.399	0.004	0.004	0.404
LLDL	–0.003	0.002	0.089	–0.003	0.002	0.094
SYSTOLIC*BLACK	0.004	0.009	0.636	0.005	0.009	0.601
SYSTOLIC*HISP	0.016	0.008	0.046	0.017	0.008	0.037

Table 4

Estimates of γ coefficients with $h = 15$ for profiling and backfitting methods.

Variable	Profiling			Backfitting		
	Estimate	s.e.	p-value	Estimate	s.e.	p-value
etmod	-0.279	0.142	0.049	-0.282	0.143	0.049
NCAD	0.398	0.132	0.002	0.401	0.132	0.002
BLACK	-0.338	1.296	0.793	-0.324	1.303	0.803
HISP	-2.286	1.186	0.053	-2.272	1.190	0.056
MALE	0.449	0.138	0.001	0.442	0.139	0.001
SYSTOLIC	0.001	0.007	0.848	0.001	0.007	0.842
DIABETES	0.813	0.130	<0.001	0.813	0.130	<0.001
SMOKER1	0.150	0.143	0.294	0.149	0.143	0.296
SMOKER2	0.478	0.176	0.006	0.480	0.176	0.006
LHDL	0.003	0.004	0.402	0.003	0.004	0.405
LLDL	-0.002	0.001	0.092	-0.002	0.001	0.088
SYSTOLIC*BLACK	0.004	0.008	0.649	0.003	0.008	0.657
SYSTOLIC*HISP	0.016	0.008	0.044	0.016	0.008	0.046