The Annals of Applied Statistics
2017, Vol. 11, No. 1, 185–201
DOI: 10.1214/16-AOAS999

A PENALIZED COX PROPORTIONAL HAZARDS MODEL WITH MULTIPLE TIME-VARYING EXPOSURES[1]

BY CHENKUN WANG, HAI LIU AND SUJUAN GAO

Indiana University and Gilead Sciences, Inc

In recent pharmacoepidemiology research, the increasing use of electronic medication dispensing data provides an unprecedented opportunity to examine various health outcomes associated with long-term medication usage. Often, patients may take multiple types of medications intended for the same medical condition and the medication exposure status and intensity may vary over time, posing challenges to the statistical modeling of such data. In this article, we propose a penalized Cox proportional hazards (PH) model with multiple functional covariates and potential interaction effects. We also consider constrained coefficient functions to ensure a diminishing medication effect over time. Hypothesis testing of interaction effect and main effect was discussed under the penalized Cox PH model setting. Our simulation studies demonstrate the adequate performance of the proposed methods for both parameter estimation and hypothesis testing. Application to a primary care depression cohort study was also illustrated to examine the effects of two common types of antidepressants on the risk of coronary artery disease.

1. Introduction. The increasing use of electronic medical records (EMR) in health care systems has been routinely capturing medication dispensing data including medication names, dosage and duration. The availability of such data offers an unprecedented opportunity for pharmacoepidemiological studies to examine various health outcomes associated with long-term medication use for monitoring both treatment efficacy and potential side effects. Many medications gain approval in relatively short-term randomized trials with strict exclusion criteria. The use of electronic medication dispensing data will allow the investigation of the associations between long-term medication use and disease treatment, prognosis and side effects in patient populations such as the elderly that may be underrepresented in the original trials for drug approvals.

An important research question that can be addressed using medication dispensing data is the association between medication exposure and time to various clinical events. Existing statistical models, however, face a number of methodological challenges when analyzing medication dispensing data. First, a patient's

Received May 2015; revised October 2016.

[1]Supported in part by National Institutes of Health (NIH) Grants R24 MH080827, K24 AG024078, R01 AG019181 and P30 AG10133.

Key words and phrases. Pharmacoepidemiology, time-varying exposure, interaction, penalized spline.

medication use and dosage vary over time. Thus, any appropriate measure of medication exposure will need to consider both exposure status and its intensity as time-varying variables. Second, for modeling exposure to medication or environmental toxins, it is often assumed, based on the compounds' half life, that exposures occurring a long time before an outcome have negligible influence. Therefore, analysis methods need to reveal not only the association between a clinical event and a particular medication, but also the relevant time window when such exposure is critical. Finally, patients may take multiple types of medications, concurrently or sequentially, intended for the same medical condition. For example, patients with depression may switch antidepressants due to side effects or lack of efficacy. Patients with hypertension, on the other hand, may take more than one class of antihypertensive medications at the same time to reach optimal blood pressure control. Therefore, statistical models need to be able to detect potential interactions among various types of medications.

Breslow et al. (1983) and Thomas (1988) proposed the concept of weighted cumulative exposure (WCE) by assigning predetermined weights to time-varying exposure levels and summing the exposures into a unidimensional variable. Applications using WCE with predefined parametric functions can also be found in Abrahamowicz et al. (2006), Vacek (1997), Langholz et al. (1999) and Richardson (2009). Extending the WCE approach, various approaches have also been proposed to estimate the weight function using nonparametric splines in generalized linear models [Berhane, Hauptmann and Langholz (2008), Hauptmann et al. (2000)] and survival models [Gasparrini (2014), Sylvestre and Abrahamowicz (2009)]. Extension to multiple exposures with WCE functions was made by Gasparrini (2014) under the additivity assumption. However, no studies to date have considered modeling the potential interaction effects of multiple time-varying exposures.

Models for time-varying exposure have also been an active focus in functional data analysis where functional coefficients from functional regression models are equivalent to the weight functions for time-varying exposures in WCE models with a difference of constant ratio. Many functional regression models include a single functional covariate [Zhang, Lin and Sowers (2007), Schipper, Taylor and Lin (2008) and Bhadra et al. (2012)]. Goldsmith et al. (2012) and Ferraty and Vieu (2009) included two or more functional covariates using additive models. Fuchs, Scheipl and Greven (2015) proposed a penalized scalar-on-function regression model with an interaction term, extending the main effect models in Wood (2011). However, there has been little progress in modeling multiple time-varying exposures with potential interactions for time-to-event data.

One characteristic of exposure models is the constraint that coefficient functions should be bounded toward zero at a distant boundary. Previously, Schipper, Taylor and Lin (2008) proposed two constraints in generalized functional models by requiring a zero coefficient at zero dose and imposing monotonicity of the coefficient function so that a higher dose leads to more adverse outcomes. A constrained

model for time-varying exposures for survival outcomes has not been considered previously.

In this article, we propose a penalized Cox proportional hazards (PH) model for multiple time-varying exposures with interactions. In particular, coefficient functions for functional covariates are modeled using penalized cubic B-splines for the main effects and tensor product splines for the interaction effect. We impose ridge penalties on the coefficient functions for the main effect and the coefficient surface for the interaction effect to better handle the issue of overfitting on coefficient functions using regression splines [Sylvestre and Abrahamowicz (2009)]. Alternative penalties can be used for the WCE functions, including the L_1 penalty [Goeman, Meijer and Chaturvedi (2016)], P-splines [Muggeo (2008), Obermeier et al. (2015)] or penalties on the second derivative for cubic regression splines [Wood (2006)]. In particular, P-splines for the Cox PH model have been discussed by Therneau, Grambsch and Pankratz (2003) and were implemented in the "survival" package in R. The coefficient functions and coefficient surface are estimated by maximizing the penalized log partial likelihood [Perperoglou (2014), Therneau, Grambsch and Pankratz (2003)]. The hypothesis testing is carried out using Wald's method [Gray (1992, 1994)]. The performance of the model is evaluated in simulation studies.

Section 2 describes data from a primary care depression study as a motivating example. Section 3 introduces the penalized Cox PH model with an interaction term and describes the parameter estimation methods as well as the hypothesis testing. Section 4 presents results from simulation studies. In Section 5, we apply the proposed model to the depression study data and estimate the association between antidepressant exposures and the risk of coronary artery disease (CAD).

2. A primary care depression screening cohort. From January 1991 to June 1993, 3767 elderly patients, age 65 or older, attending primary care clinics at Wishard Health Services were enrolled into a depression screening study [Callahan et al. (1994)]. Electronic medical records of these patients from their enrollment to December 31, 2010 were extracted from the Regenstrief Medical Record System (RMRS) [McDonald et al. (1999)]. RMRS is one of the first electronic medical record systems in the United States and has been actively used for research purposes. RMRS routinely captures laboratory results, narrative reports, orders, medications, radiology reports, registration information, nursing assessments, vital signs and other clinical data. The medical records include comprehensive medication dispensing information capturing medication names, daily dose and start/end dates for each dispensing record.

Antidepressant is one of the most commonly prescribed medication groups in the United States [Lindsley (2012)]. An older class of antidepressants, tricyclic antidepressants (TCAs), has been shown to have a detrimental effect on the cardiovascular function by inhibiting cardiovascular Na(+), Ca(2+) and K(+) channels often leading to life-threatening arrhythmia [Glassman (1984), Jefferson (1975)].

A newer class of antidepressants, selective serotonin reuptake inhibitors (SSRIs), became the preferred treatment for depression due to its comparable efficacy with TCAs and its superior tolerability. SSRIs were hypothesized to show a different cardiovascular effect from TCAs due to their pharmacologic profile [Bergstrom et al. (1988)]. However, reports of the first-degree atrioventricular block, prolonged QTc interval and orthostatic hypotension in SSRI-treated patients were raising concerns that SSRIs may also have important cardiac and vascular effects [de la Torre et al. (2001), Pacher and Kecskemeti (2004)]. Given that many of the studies on antidepressants were conducted in laboratory settings with brief antidepressant treatment, EMR data with detailed medication dispensing information offer a unique opportunity to examine the effect of the long-term use on cardiac functions in the population of elderly patients, a group often underrepresented in clinical trials. Moreover, none of the previous studies considered the comprehensive information of time-varying antidepressant use. It is widely known that patients on depression treatments receive different dosages and switch medications frequently; it is therefore of interest to examine whether the time-varying exposures to TCAs or SSRIs and their interaction are associated with the risk of CAD.

In the data set, we included patients who were taking TCAs or SSRIs during the follow-up period. To derive comparable dosages among different medications, we first standardized medication doses using each daily dose divided by the recommended minimum dose for the particular medicine [Damush et al. (2008), Gray et al. (2015)]. In this cohort, antidepressant usage exhibits great variations in both dose intensity and duration of treatment (Figure 1). Such exposure patterns demonstrate the need for a statistical method that can account for the time-varying exposure and potential interaction effect on the risk of a disease. We emphasize that the conventional Cox PH model cannot be expected to provide a good fit to such a scenario due to its assumption of time-invariant coefficients.

3. Methods.

3.1. *A Cox PH model with multiple time-varying exposures.* Let $\bar{A}_i(u) = \{D_i^1(t)\}_{t=0}^u$ and $\bar{B}_i(u) = \{D_i^2(s)\}_{s=0}^u$ denote two series of time-varying exposures, A and B, up to time u, and let $\boldsymbol{X}_i$ be a row vector of p covariates for subject i, $i = 1, \ldots, n$, with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)^T$ being the vector of corresponding regression coefficients. The hazard function of a Cox PH model at time u given the history of the exposures through u and covariates $\boldsymbol{X}_i$ can be defined as

$$
\begin{aligned}
&h\big(u|\bar{A}_i(u), \bar{B}_i(u), \boldsymbol{X}_i\big) \\
&\quad = h_0(u)\exp\bigg\{\int_0^u w_1(u-t)D_i^1(t)\,dt + \int_0^u w_2(u-s)D_i^2(s)\,ds \\
&\qquad + \int_0^u\int_0^u w_3(u-t, u-s)D_i^1(t)D_i^2(s)\,dt\,ds + \boldsymbol{X}_i\boldsymbol{\alpha}\bigg\},
\end{aligned} \tag{3.1}
$$

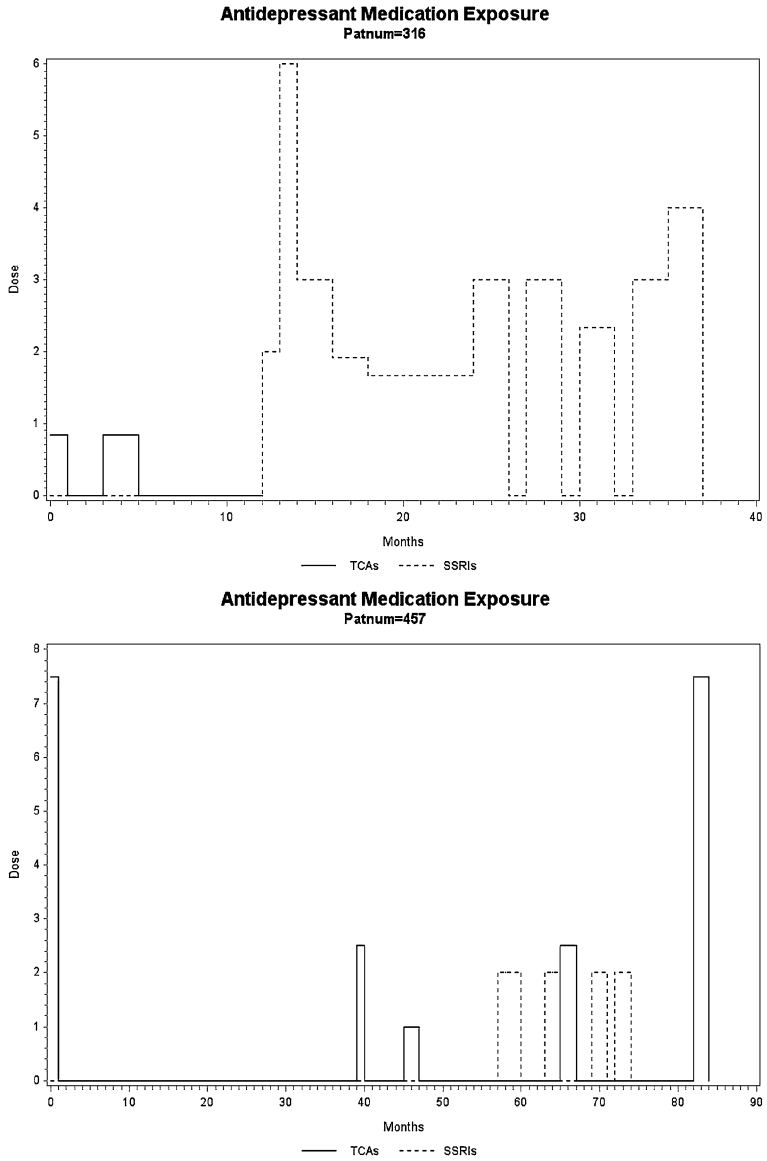


FIG. 1. *Examples of medication exposure from two patients. Top*: *patient who switched from TCAs to SSRIs*; *Bottom*: *patient who took TCAs and SSRIs sequentially.*

where $h_0(u)$ is an unspecified baseline hazard function, $w_1(u-t)$ and $w_2(u-s)$ are unknown smooth coefficient functions for time-varying exposures $D_i^1(t)$ and $D_i^2(s)$, respectively, and $w_3(u-t, u-s)$ is a bivariate coefficient surface function which quantifies potential interaction effect between the two exposures. The term $\int_0^u w_k(u-t)D_i^k(t)\,dt$ summarizes the influence of time-varying exposure

$k, k = 1, 2$, up to time u. The coefficient functions were defined on reversed time axes so that time 0 for coefficient functions corresponds to exposure at time u. We further assume that past exposure occurring at certain time units, for example, $a_k, k = 1, 2$, before time u has a negligible effect, and thus the coefficient functions smoothly go to zero at time a_k. We impose such constraint by defining the coefficient function as an integral $w_k(t) = \int_{-a_k}^{-t} s_k(c)\,dc, t \in [0, a_k]$ so that $w_k(a_k) = 0$. The integrand $s_k(c)$ can be flexibly expressed as a linear combination of some spline basis functions as $s_k(c) = \sum_{l=1}^{N_k} \beta_{kl} B_{kl}(c)$, where N_k is the number of basis functions $B_{kl}(c)$, with coefficients β_{kl}. In this study, we use cubic B-spline bases, and the total number of basis functions is $N_k = L_k + 4$ with L_k interior knots, $k = 1, 2$.

The coefficient surface $w_3(u - t, u - s)$ between two time-varying exposures is modeled by a nonparametric bivariate function using tensor product splines as $w_3(u - t, u - s) = \sum_{l,r} \gamma_{lr} B_{lr}(u - t, u - s)$, where $B_{lr}(u - t, u - s)$ is obtained by forming all pairwise products between the cubic B-spline basis functions as $B_{lr}(u - t, u - s) = B_l(u - t)B_r(u - s), l, r = 1, \ldots, N_3$, resulting in a total of $N_3 \times N_3$ basis functions, assuming the same number of interior knots along both directions for simplicity.

Define row vector $\boldsymbol{M}_i^k(u) = (M_{i1}^k(u), \ldots, M_{i,N_k}^k(u)), k = 1, 2$, where $M_{il}^k(u) = \int_0^u D_i^k(t) \int_{-a_k}^{-(u-t)} B_{kl}(c)\,dc\,dt, l = 1, \ldots, N_k$. And let $\boldsymbol{Z}_i(u) = (Z_{i11}(u), \ldots, Z_{i,1,N_3}(u), \ldots, Z_{i,N_3,N_3}(u))$, where $Z_{ilr}(u) = \int_0^u \int_0^u B_{lr}(u - t, u - s) D_i^1(t) \times D_i^2(s)\,dt\,ds, l, r = 1, \ldots, N_3$. Denote $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{k,N_k})^T, k = 1, 2$, and $\boldsymbol{\gamma} = (\gamma_{11}, \ldots, \gamma_{1,N_3}, \ldots, \gamma_{N_3,N_3})^T$ as the coefficient vectors for the B-spline basis and tensor product basis, respectively. It follows that model (3.1) can be rewritten in a more condensed form with regression parameter vector $\boldsymbol{\eta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\gamma}^T)^T$:

$$
\begin{aligned}
&h\big(u|\bar{A}_i(u), \bar{B}_i(u), \boldsymbol{X}_i\big) \\
&\qquad = h_0(u) \exp\big\{\boldsymbol{M}_i^1(u)\boldsymbol{\beta}_1 + \boldsymbol{M}_i^2(u)\boldsymbol{\beta}_2 + \boldsymbol{Z}_i(u)\boldsymbol{\gamma} + \boldsymbol{X}_i\boldsymbol{\alpha}\big\}.
\end{aligned} \tag{3.2}
$$

3.2. *Penalized partial likelihood.* The estimation of the Cox PH model (3.2) can be implemented by maximizing the partial likelihood using a standard statistical software. The estimated coefficient functions will be represented by regression splines, which, however, can be unstable and sensitive to the number and location of knots [Hastie and Tibshirani (1986)] and suffer from overfitting issues [Gray (1992)]. Therefore, we adopted the penalized spline method to estimate the coefficient functions for the time-varying exposures. The use of the penalized approach is well known to avoid identifiability problems [Ramsay and Silverman (2005), Scheipl and Greven (2015)]. In particular, a ridge penalty of the L_2 norm of the regression coefficients was imposed, and the resulting penalized log partial likelihood [Perperoglou (2014)] has the form of

$$
\ell_p(\boldsymbol{\eta}, \boldsymbol{\lambda}) = \ell(\boldsymbol{\eta}) - \lambda_1 \boldsymbol{\beta}_1^T \boldsymbol{\beta}_1 - \lambda_2 \boldsymbol{\beta}_2^T \boldsymbol{\beta}_2 - \lambda_3 \boldsymbol{\gamma}^T \boldsymbol{\gamma}, \tag{3.3}
$$

where $\ell(\boldsymbol{\eta})$ is the log partial likelihood, and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)^T$ are the smoothing parameters. The parameter estimates are obtained by maximizing the penalized log partial likelihood in (3.3) for any given non-negative smoothing parameters $\boldsymbol{\lambda}$ via the Newton–Raphson procedure.

In the penalized Cox PH model above, it is crucial to determine the appropriate smoothing parameters $\boldsymbol{\lambda}$, which balance the trade-off between the goodness of fit and the smoothness of the coefficient functions. One way to optimize the penalty weight is to conduct a grid search over some prespecified range of $\boldsymbol{\lambda}$. The models are fitted with different values of smoothing parameters, and the optimal choice maximizes a chosen criterion such as the Akaike information criterion (AIC) with $AIC = 2\ell(\boldsymbol{\eta}) - 2df$ [Perperoglou (2014)]. The degrees of freedom (df) of the penalized Cox model can be calculated as $df = \text{trace}[\boldsymbol{I}(\boldsymbol{\eta})\boldsymbol{H}^{-1}(\boldsymbol{\eta}, \boldsymbol{\lambda})]$, where $\boldsymbol{I}(\boldsymbol{\eta}) = -\partial^2\ell(\boldsymbol{\eta})/\partial\boldsymbol{\eta}^2$ is the observed information matrix for the partial likelihood, $\boldsymbol{H}(\boldsymbol{\eta}, \boldsymbol{\lambda}) = \boldsymbol{I}(\boldsymbol{\eta}) + \boldsymbol{P}(\boldsymbol{\lambda})$, and $\boldsymbol{P}(\boldsymbol{\lambda})$ is a block diagonal matrix $\text{diag}(\lambda_1\boldsymbol{I}_1, \lambda_2\boldsymbol{I}_2, \lambda_3\boldsymbol{I}_3)$ with $\boldsymbol{I}_k$ being identity matrix of dimension N_k, $k = 1, 2, 3$ [Gray (1992)]. Alternatively, Hurvich, Simonoff and Tsai (1998) suggested a corrected AIC (AICc), using $n(df + 1)/(n - (df + 2))$ (n is the total number of events in the Cox PH model) as the correction term in place of df in AIC, based on the argument that AIC can lead to models with an excessive number of degrees of freedom. We consider both AIC and AICc and compare their performance in simulation studies described in Section 4.

In practice, the time-varying exposures are usually measured at discrete time points, for example, daily medication intake, and the integrals in $M_{il}^k(u)$ and $Z_{ilr}(u)$ can be replaced by corresponding summations, which are straightforward and hence omitted here.

3.3. *Inference*. Several important scientific questions can be addressed by hypothesis tests. For example, one hypothesis of interest is whether the two exposures have interacting effects on the risk of a survival outcome, with a null hypothesis of $H_0 : w_3 = 0$ on the coefficient surface w_3. In the penalized spline model setup (3.2), this hypothesis is equivalent to $H_0 : \boldsymbol{\gamma} = \mathbf{0}$, that is, coefficient vector $\boldsymbol{\gamma}$ is a zero vector. If the null hypothesis on the interaction term is not rejected, the hazard function becomes additive of the two exposure effects on the logarithm scale. Then one would want to know if the two exposures individually contribute significant influence to the risk of the outcome, leading to the null hypothesis of $H_0 : w_k = 0$, or $\boldsymbol{\beta}_k = \mathbf{0}$, $k = 1, 2$.

The above hypotheses follow a general form of $\boldsymbol{C}\boldsymbol{\eta} = \mathbf{0}$ with a properly defined matrix $\boldsymbol{C}$ of full row rank. Following Gray (1992), a Wald-type test statistic can be defined as $(\boldsymbol{C}\boldsymbol{\eta})^T(\boldsymbol{C}\boldsymbol{H}^{-1}\boldsymbol{C}^T)^{-1}(\boldsymbol{C}\boldsymbol{\eta})$. In the penalized Cox model, however, this test will be conservative if a standard Chi-square statistic is assumed. Gray (1992) showed that the expectation of the test statistic under the null hypothesis is approximately equal to $df_{\text{Wald}} = \text{trace}[(\boldsymbol{C}\boldsymbol{H}^{-1}\boldsymbol{C}^T)^{-1}(\boldsymbol{C}\boldsymbol{H}^{-1}\boldsymbol{I}\boldsymbol{H}^{-1}\boldsymbol{C}^T)]$, which will be considered as the generalized degree of freedom of the Wald test in our

proposed model. Simulation studies were conducted to examine the performance of the hypothesis testing procedure.

To measure the variability of estimated coefficient functions, sample-based confidence bands are provided. In simulations, the confidence bands are calculated from simulated data sets. In applications, we relied on nonparametric bootstrap resampling. For each of the samples, model (3.1) is estimated. Next, the empirical distribution of the point estimates of coefficient functions is constructed. The percentile method can then be used to compute 95% pointwise confidence bands.

4. Simulations.

4.1. *Simulation setup*. To evaluate the performance of our proposed model, simulation studies were conducted. Three patient groups were considered based on the types of exposure. Subjects in the first group took medication A exclusively, subjects in the second group took medication B exclusively and subjects in the third group took both medications A and B over the follow-up period, sequentially or concurrently. Assuming equal sample size in the three groups, three levels of total sample size were investigated: 300, 450 and 600.

The timing of the exposure, measured by days taking medication or off medication in blocks of 7 days, was generated from a lognormal distribution with a mean of 0.5 and a standard deviation of 0.8, that is, $\log(\text{duration}) \sim N(0.5, 0.8^2)$, and rounded up to the nearest integer. The intensity of the time-varying exposure was assumed to be constant over each treatment period (in a 7-day unit), and it could take values of 0.005, 0.01, 0.015, 0.02, 0.025 or 0.03 during the follow-up period with equal probabilities.

Several coefficient functions with a fixed time window of 90 days were considered: (1) zero function ($w(t) = 0, t \in [0, 90]$) indicating no exposure effect over the time window, (2) decreasing function ($w(t) = \sin(\frac{\pi}{2}\frac{t}{90} + \frac{\pi}{2}), t \in [0, 90]$) indicating a greater effect closer to the event time and a lesser effect for distant exposures, and (3) U-shape function ($w(t) = \sin(\pi\frac{t}{90} + \pi), t \in [0, 90]$) suggesting a greater negative effect for mid-point exposures and lesser effects for recent and distant exposures. All true coefficient functions smoothly go to zero at the distant boundary of the exposure interval. The coefficient functions were defined on reversed time axes so that day zero corresponds to a given time u when the hazard function is measured. The assumption that exposures 90 days before have negligible effects is equivalent to the assumption that coefficient functions are flat zero lines beyond 90 days. We generated 200 data sets under each of the four scenarios as shown in Figure 2.

The bivariate function in Figure 3 was used to generate the interaction term between the two types of exposures with a time window of 30 days. Higher values are along the diagonal direction, indicating that medications have a larger interaction effect when taken simultaneously, and this effect decreases closer to the distant boundary. Similarly, the interaction surface was also defined on reversed time

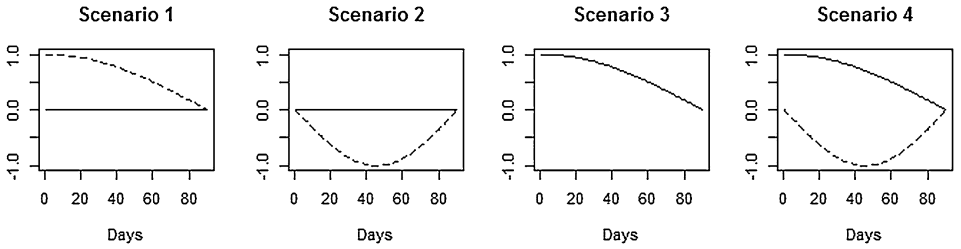
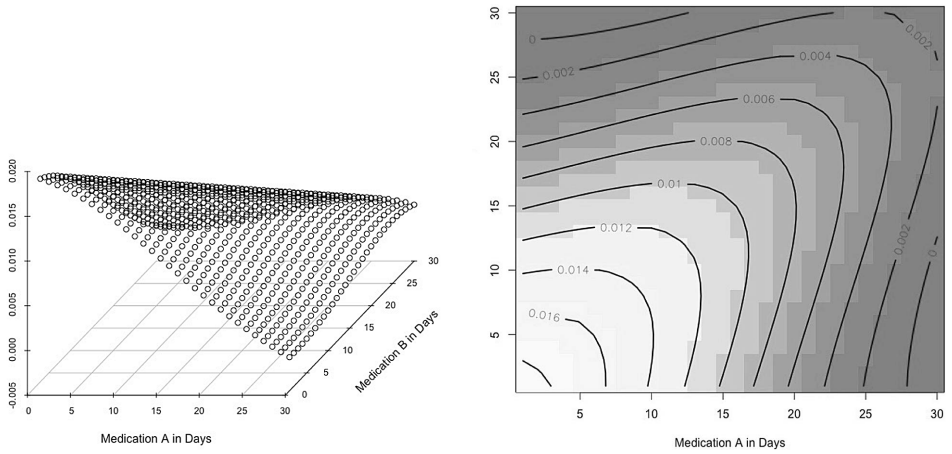


FIG. 2. *Four scenarios of coefficient functions used in simulations*: $w_1(t)$ *is indicated by solid lines*; $w_2(t)$ *is indicated by dashed lines. The coefficient functions were defined on reversed time axes so that day zero corresponds to the time of outcome measures and increasing days correspond to the more distant past. Scenario* 3 *has identical coefficient functions.*

axes. The x-axis represents the exposure time for medication A and the y-axis for medication B. For example, point (a, b) on the surface measures the interaction effect between exposure to medication A that happened a days ago and exposure to medication B that happened b days ago.

Survival outcomes were generated based on various combinations of corresponding coefficient functions using the permutation algorithm, designed to simulate survival times with time-dependent covariates [Sylvestre and Abrahamowicz (2008)].

4.2. *Simulation results*. For each of the 200 simulated data sets under each scenario, we fitted the proposed model (3.1) with six equally spaced interior knots



Interaction surface (3D) Interaction surface (Contour)

FIG. 3. *Interaction surface in* 3*D and contour plots under a prespecified time window of* 30 *days. Surfaces were defined on reversed time axes. The* x*-axis indicates the exposure time for medication A and the* y*-axis represents medication B.*

TABLE 1
Average mean squared error (AMSE) ($\times 10^{-5}$) of coefficient functions and coefficient surface estimated by penalized Cox PH model over 200 replications

Coefficient function	Sample size	Selection method	Scenarios			
			S1	S2	S3	S4
$w_1(t)$	300	(AIC)	4.40	4.76	9.11	9.06
	300	(AICc)	3.58	3.86	8.84	8.26
	450	(AIC)	3.08	3.66	5.72	5.43
	450	(AICc)	2.96	3.33	5.44	5.24
	600	(AIC)	2.26	2.30	5.05	4.46
	600	(AICc)	2.16	2.14	4.77	4.07
$w_2(s)$	300	(AIC)	7.91	12.4	7.97	14.7
	300	(AICc)	6.93	12.1	7.59	14.0
	450	(AIC)	5.60	9.60	5.82	9.39
	450	(AICc)	5.31	9.41	5.68	9.15
	600	(AIC)	4.44	8.08	4.67	7.71
	600	(AICc)	4.37	8.07	4.50	7.61
$w_3(t,s)$	300	(AIC)	2.45	2.28	2.37	1.84
	300	(AICc)	1.83	1.89	2.00	1.52
	450	(AIC)	1.29	1.20	1.88	1.17
	450	(AICc)	1.19	1.08	1.68	1.07
	600	(AIC)	0.81	0.93	1.24	0.88
	600	(AICc)	0.77	0.89	1.16	0.80

for the coefficient functions, leading to $N_1 = N_2 = 10$ basis functions used for each main effect. One interior knot was used for each direction of coefficient surface, and in total there are 25 tensor product basis functions ($N_3 = 5$) used for the interaction effect. Both AIC and AICc were used for the selection of smoothing parameters.

Estimates of coefficient functions $\hat{w}_1$, $\hat{w}_2$ and $\hat{w}_3$ were obtained under each scenario. The average mean squared error (AMSE) for the interaction effect is calculated as

$$AMSE(w_3) = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{30\times 30}\sum_{t=1}^{30}\sum_{s=1}^{30}\big(\hat{w}_3(t,s) - w_3(t,s)\big)^2, \tag{4.1}$$

where R represents the total number of replicates. The AMSE for w_1 and w_2 can be similarly defined.

Results of the penalized spline estimates are presented in Table 1. The AMSE of estimated coefficient functions and coefficient surface decreases as the sample size increases. Models that use the corrected AIC (AICc) to select smoothing parameters have better performance when compared with models using AIC. Both methods show satisfactory performance in terms of the mean squared error.

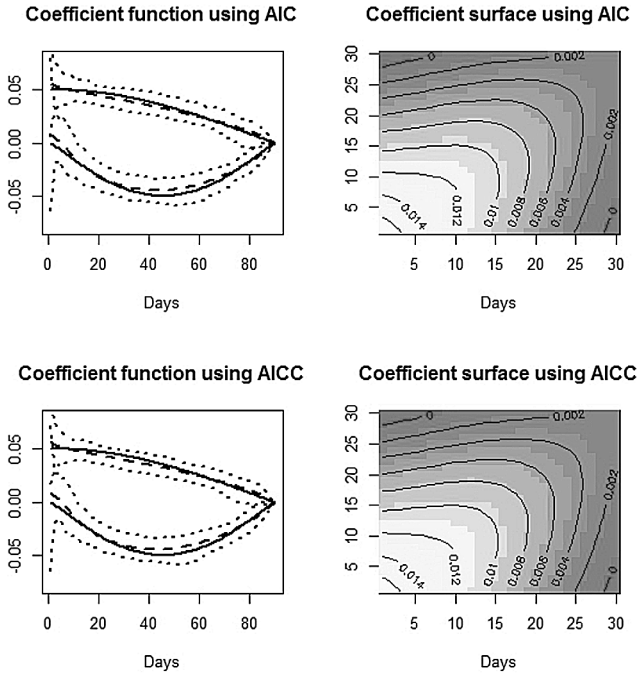


FIG. 4. *Mean estimated coefficient functions and mean estimated coefficient surface in Scenario* 4 *with a sample size of* 600. *True coefficient functions are denoted as solid lines. Dashed lines are mean coefficient functions. Dotted lines are sample-based* 95% *point-wise confidence intervals*

Figure 4 shows mean estimated coefficient functions and the mean estimated coefficient surface under Scenario 4 with a sample size of 600. In the left panels, the true coefficient functions are denoted by solid lines and the sample mean coefficient functions are denoted by dashed lines. Dotted lines are sample-based 95% point-wise confidence intervals. The estimated functions and surfaces are able to capture the shape of the true functions and surfaces. Similar results were obtained in the other three scenarios.

Another simulation study was conducted to assess the performance of the Wald tests on the interaction effect and the coefficient functions discussed in Section 3.3. A similar simulation setup was used except that no interaction term was included when generating the data under all four scenarios. The empirical Type I error rates of testing the null hypothesis of no interaction effect, that is, $w_3 = 0$ or $\boldsymbol{\gamma} = \mathbf{0}$, were summarized in Table 2 with a sample size of 450 across 200 replications. In Scenario 1 and Scenario 2, w_1's are flat zero lines, indicating no main effect from exposure A. Therefore, hypothesis tests for the main effect of exposure A were also performed under Scenario 1 and Scenario 2 by testing $w_1 = 0$ or $\boldsymbol{\beta}_1 = \mathbf{0}$. The results show that all empirical Type I error rates are close to the 0.05 nominal level and the selection method using AIC has comparable performance to those using AICc.

TABLE 2
Empirical Type I errors of Wald tests with a sample size of 450 *across* 200 *replications*

	Scenarios			
Null hypothesis (Selection method)	**S1**	**S2**	**S3**	**S4**
$w_3 = 0$ (AIC)	0.06	0.07	0.05	0.03
$w_3 = 0$ (AICc)	0.05	0.07	0.05	0.03
$w_1 = 0$ (AIC)	0.06	0.04	–	–
$w_1 = 0$ (AICc)	0.06	0.04	–	–

5. Application. In this section, we revisit the motivating example described in Section 2 and apply the proposed model to the data set. Our primary interest is to examine the effects of two classes of antidepressants (TCA and SSRI) on the risk of CAD. We included only patients who started antidepressants after study enrollment. For patients who started taking TCA or SSRI during follow-up, we redefined study baseline as the time of the first antidepressant dispensing. Patients with diagnoses of existing CAD [indicated by the International Classification of Diseases, Ninth Revision (ICD-9) codes contained in inpatient, outpatient and emergency room records] before initiating antidepressants were excluded. Incident CAD was defined as any of the following events during the follow-up period: (a) acute myocardial infarction (MI) indicated by ICD-10 codes as the cause of death (from National Death Index); (b) first diagnosis of CAD by ICD-9 codes in EMR; or (c) the first laboratory evidence of acute MI (Creatine kinase-myocardial band isoenzyme value > 3.0 ng/ml or troponin value > 0.3 ug/L) in EMR. Time to event was calculated as time to the first of the three CAD events defined above. Censoring time was the last clinic visit time for those who were alive or time of death for deceased patients without CAD. The analysis data set contained 297 subjects in total, with 160 patients taking TCAs exclusively, 93 patients taking SSRIs exclusively and 44 patients taking both. A total of 91 patients (30%) developed CAD during the follow up.

We present patient demographic information in the three medication groups in Table 3. Age at first dispensing is significantly different among the three patient groups, and those in the SSRIs only group are on average older than the other two groups. The group taking both TCAs and SSRIs included significantly fewer African–American patients than the other two groups. Univariate Cox PH models show that age is a risk factor for CAD. Therefore, we included both age at first dispensing and race as covariates in our models. AICc was used as the criterion to select smoothing parameters as suggested by simulation studies. A variety of prespecified exposure time windows were examined. Penalized cubic B-splines with 10 basis functions and cubic tensor product splines with 25 basis functions were used for the estimation of coefficient functions (w_1 and w_2) and coefficient surface (w_3), respectively.

TABLE 3
Comparison of demographic information among three groups. P-value is calculated using ANOVA for continuous variables and Chi-square test for binary variables

	SSRIs ($n = 93$)	**TCAs ($n = 160$)**	**TCAs&SSRIs ($n = 44$)**	***P*-value**
Age at first serve Mean (SD)	73.62 (6.15)	70.21 (6.88)	69.73 (6.39)	0.0001
Years of education Mean (SD)	8.91 (2.67)	8.76 (2.83)	8.48 (3.45)	0.7094
Female (%)	80 (86.02)	120 (75)	35 (79.55)	0.1147
African–American (%)	69 (74.19)	115 (71.88)	24 (54.55)	0.0484
Smoke at any time (%)	31 (33.33)	69 (43.13)	16 (36.36)	0.2827
CAD Events (%)	27 (29.03)	53 (33.13)	11 (25)	0.5389

Table 4 presents coefficient estimates for the covariates and p-values obtained from hypothesis tests for interaction and main effects under various prespecified time windows (in months). For all the time windows considered, results from Wald tests indicate that there is no interaction between TCAs and SSRIs. We also obtained similar results with longer time intervals (results not shown). In all models, we observed a marginally significant main effect for TCAs, but not for SSRIs. Figure 5 shows estimated coefficient functions for TCAs under various exposure time intervals and under 12 months with a 95% bootstrap confidence band. All models indicate that exposure to TCAs increases the risk of CAD. The largest influence from TCAs exposure was consistently seen at around 3 months within event time, implying the necessity of close monitoring after first initiate TCAs treatment.

6. Discussion. In this article, we proposed a penalized Cox PH model with multiple functional covariates, interaction term and constraints on the coefficient functions. In our proposed model framework, the coefficient function takes a flexible form through the use of penalized splines, which provides a parsimonious summary measure of both exposure time and intensity. The use of a penalized partial likelihood method stabilizes parameter estimates and is less sensitive to the

TABLE 4
Estimates and P-values of hypothesis test of no interaction effect or no main effect under various prespecified time intervals

Time interval (in month)	**Covariates estimates (*P*-value)**		***P*-values of hypothesis test**		
	African–American	**Age**	**Interaction**	**TCAs**	**SSRIs**
12	0.06 (0.79)	0.04 (0.01)	0.97	0.06	0.67
18	0.05 (0.83)	0.04 (0.01)	0.94	0.07	0.75
24	0.06 (0.80)	0.04 (0.01)	0.83	0.07	0.81
30	0.05 (0.83)	0.04 (0.01)	0.88	0.07	0.76

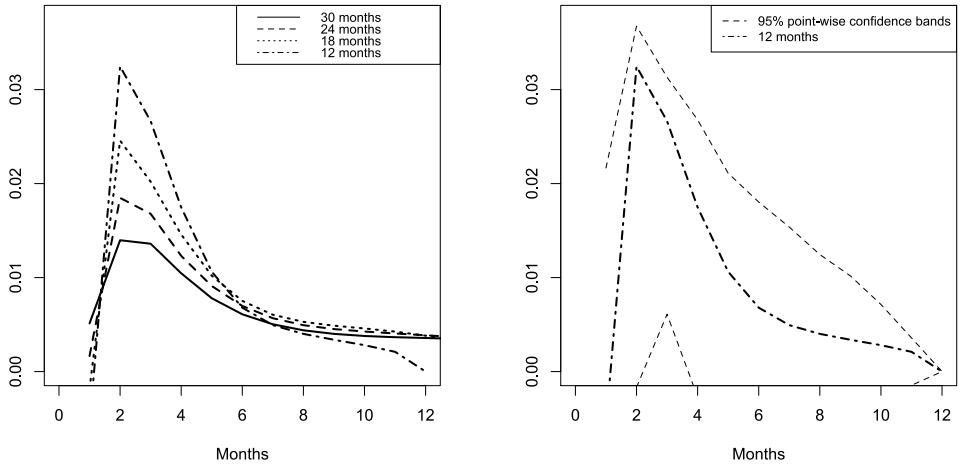


FIG. 5. *Left*: *estimated coefficient functions for TCAs under various prespecified time interval. Right*: *estimated coefficient function for TCAs under* 12 *months with* 95% *bootstrap confidence band.*

selection of knots. The inclusion of interactions in the models allows estimation and inference on potential interactive effects from multiple classes of medications. Potential interactions among multiple medications will be of particular interest for the elderly population, as they may suffer from multiple chronic conditions and many of these medication combinations had not been studied in randomized clinical trials in the elderly. Our simulation studies demonstrate the adequate performance of the proposed methods for parameter estimation as well as hypothesis testing. Our proposed models can be readily applied to medication dispensing data or other time-varying exposures to determine the association between multiple exposures and time-to-event outcomes.

The models considered here have constraints at the right end of coefficient functions for both scientific and model identification purposes. In a WCE model for a single exposure, Sylvestre and Abrahamowicz (2009) imposed a similar constraint by forcing the last two coefficients of basis functions to zero, leading to a loss of degrees of freedom. Our proposed approach imposed the constraint through a functional form, preserving the model's degrees of freedom. In addition, with data manipulation and formatting, our proposed models can be implemented in standard software packages, making the approach readily applicable to other studies. Another significant development in our methods is the inclusion of potential interactions among different medications. In aging studies involving patients with multiple chronic conditions, most patients were found to be on multiple medications at any given period. Thus, it is important that statistical models take into account all medications to which patients have been exposed so that potential risk of adverse events can be detected.

For simplicity, we discussed only two time-varying exposures in this study. When considering more than two medications, the model can be extended to

higher-order interactions analogously, such as three-way interactions with an interaction surface $w(s,t,u)$. It is likely that larger sample size will be needed to estimate the interaction term and computational burden will increase due to the increases in dimension [Fuchs, Scheipl and Greven (2015)]. Further study is required to implement the higher-order models more efficiently. A natural connection between penalized the Cox PH model and the frailty model has been observed [Therneau, Grambsch and Pankratz (2003)]. In our proposed model, we used a grid search to select smoothing parameters based on AIC and AICc. Thus, our approach is computationally demanding, and the results depend on the intensity of the grid search. An alternative approach is to consider the penalized parameters as random effects in a frailty model framework, which may result in higher computational efficiency. A limitation in our models is that the models force the coefficient function to 0 at the end of the exposure window, regardless of whether this is true or not. Obermeier et al. (2015) applied additional ridge penalties to the spline coefficients, where smoothing parameters can be separately estimated to very small values if the assumption of the coefficients approaching a null value in the right tail is not met. It may be worthwhile to explore this new method in modeling cumulative exposure data.

In summary, we proposed a penalized Cox PH model with multiple time-varying exposures with potential interactions. The proposed methods performed adequately in simulation studies. Electronic medication dispensing data have provided an ideal platform to apply our proposed methods to determine whether multiple long-term medication usages are associated with time-to-event outcomes.

Acknowledgments. We would like to thank the Editor, the Associate Editor and the reviewers for their constructive and insightful comments that led to a significantly improved article.

REFERENCES

ABRAHAMOWICZ, M., BARTLETT, G., TAMBLYN, R. and DU BERGER, R. (2006). Modeling cumulative dose and exposure duration provided insights regarding the associations between benzodiazepines and injuries. *J. Clin. Epidemiol.* **59** 393–403.

BERGSTROM, R. F., LEMBERGER, L., FARID, N. A. and WOLEN, R. L. (1988). Clinical pharmacology and pharmacokinetics of fluoxetine: A review. *Br. J. Psychiatry Suppl.* **3** 47–50.

BERHANE, K., HAUPTMANN, M. and LANGHOLZ, B. (2008). Using tensor product splines in modeling exposure-time-response relationships: Application to the Colorado Plateau Uranium Miners cohort. *Stat. Med.* **27** 5484–5496. MR2542365

BHADRA, D., DANIELS, M. J., KIM, S., GHOSH, M. and MUKHERJEE, B. (2012). A Bayesian semiparametric approach for incorporating longitudinal information on exposure history for inference in case–control studies. *Biometrics* **68** 361–370. MR2959602

BRESLOW, N. E., LUBIN, J. H., MAREK, P. and LANGHOLZ, B. (1983). Multiplicative models and cohort analysis. *J. Amer. Statist. Assoc.* **78** 1–12.

CALLAHAN, C. M., HUI, S. L., NIENABER, N. A. and MUSICK, B. S. (1994). Longitudinal study of depression and health services use among elderly primary care patients. *Journal of the American Geriatrics Society* **42** 833–838.

DAMUSH, T. M., JIA, H., RIED, L. D., QIN, H., CAMEON, R., PLUE, L. and WILLIAMS, L. S. (2008). Case-finding algorithm for post-stroke depression in the veterans health administration. *International Journal of Geriatric Psychiatry* **23** 517–522.

DE LA TORRE, B. R., DREHER, J., MALEVANY, I., BAGLI, M., KOLBINGER, M., OMRAN, H., LÜDERITZ, B. and RAO, M. L. (2001). Serum levels and cardiovascular effects of tricyclic antidepressants and selective serotonin reuptake inhibitors in depressed patients. *Therapeutic Drug Monitoring* **23** 435–440.

FERRATY, F. and VIEU, P. (2009). Additive prediction and boosting for functional data. *Comput. Statist. Data Anal.* **53** 1400–1413. MR2657100

FUCHS, K., SCHEIPL, F. and GREVEN, S. (2015). Penalized scalar-on-functions regression with interaction. *Comput. Statist. Data Anal.* **81** 38–51. MR3257399

GASPARRINI, A. (2014). Modeling exposure-lag-response associations with distributed lag nonlinear models. *Stat. Med.* **33** 881–899. MR3249094

GLASSMAN, A. H. (1984). Cardiovascular effects of tricyclic antidepressants. *Annual Review of Medicine* **35** 503–511.

GOEMAN, J., MEIJER, R. and CHATURVEDI, N. (2016). L1 and L2 penalized regression models. Available at http://cran.r-project.org/web/packages/penalized/vignettes/penalized.pdf.

GOLDSMITH, J., CRAINICEANU, C. M., CAFFO, B. and REICH, D. (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **61** 453–469. MR2914521

GRAY, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *J. Amer. Statist. Assoc.* **87** 942–951.

GRAY, R. J. (1994). Spline-based tests in survival analysis. *Biometrics* **50** 640–652. MR1309310

GRAY, S. L., ANDERSON, M. L., DUBLIN, S., HANLON, J. T., HUBBARD, R., WALKER, R., YU, O., CRANE, P. K. and LARSON, E. B. (2015). Cumulative use of strong anticholinergics and incident dementia: A prospective cohort study. *JAMA Internal Medicine* **175** 401–407.

HASTIE, T. and TIBSHIRANI, R. (1986). Generalized additive models. *Statist. Sci.* **1** 297–318. MR0858512

HAUPTMANN, M., WELLMANN, J., LUBIN, J. H., ROSENBERG, P. S. and KREIENBROCK, L. (2000). Analysis of exposure-time–response relationships using a spline weight function. *Biometrics* **56** 1105–1108. MR1815589

HURVICH, C. M., SIMONOFF, J. S. and TSAI, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 271–293. MR1616041

JEFFERSON, J. W. (1975). A review of the cardiovascular effects and toxicity of tricyclic antidepressants. *Psychosomatic Medicine* **37** 160–179.

LANGHOLZ, B., THOMAS, D., XIANG, A. and STRAM, D. (1999). Latency analysis in epidemiologic studies of occupational exposures: Application to the Colorado Plateau uranium miners cohort. *American Journal of Industrial Medicine* **35** 246–256.

LINDSLEY, C. W. (2012). The top prescription drugs of 2011 in the United States: Antipsychotics and antidepressants once again lead CNS therapeutics. *ACS Chemical Neuroscience* **3** 630–631.

MCDONALD, C. J., OVERHAGE, J. M., TIERNEY, W. M., DEXTER, P. R., MARTIN, D. K., SUICO, J. G., ZAFAR, A., SCHADOW, G., BLEVINS, L., GLAZENER, T. et al. (1999). The Regenstrief medical record system: A quarter century experience. *International Journal of Medical Informatics* **54** 225–253.

MUGGEO, V. M. (2008). Modeling temperature effects on mortality: Multiple segmented relationships with common break points. *Biostatistics* **9** 613–620.

OBERMEIER, V., SCHEIPL, F., HEUMANN, C., WASSERMANN, J. and KÜCHENHOFF, H. (2015). Flexible distributed lags for modelling earthquake data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **64** 395–412. MR3302306

PACHER, P. and KECSKEMETI, V. (2004). Cardiovascular side effects of new antidepressants and antipsychotics: New drugs, old concerns? *Curr. Pharm. Des.* **10** 2463–2475.

PERPEROGLOU, A. (2014). Cox models with dynamic ridge penalties on time-varying effects of the covariates. *Stat. Med.* **33** 170–180. MR3141561

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

RICHARDSON, D. B. (2009). Latency models for analyses of protracted exposures. *Epidemiology* (*Cambridge*, *Mass.*) **20** 395.

SCHEIPL, F. and GREVEN, S. (2015). Identifiability in penalized function-on-function regression models. Preprint. Available at arXiv:1506.03627.

SCHIPPER, M., TAYLOR, J. M. G. and LIN, X. (2008). Generalized monotonic functional mixed models with application to modelling normal tissue complications. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **57** 149–163. MR2420434

SYLVESTRE, M.-P. and ABRAHAMOWICZ, M. (2008). Comparison of algorithms to generate event times conditional on time-dependent covariates. *Stat. Med.* **27** 2618–2634. MR2440055

SYLVESTRE, M.-P. and ABRAHAMOWICZ, M. (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Stat. Med.* **28** 3437–3453. MR2744373

THERNEAU, T. M., GRAMBSCH, P. M. and PANKRATZ, V. S. (2003). Penalized survival models and frailty. *J. Comput. Graph. Statist.* **12** 156–175. MR1965213

THOMAS, D. C. (1988). Models for exposure-time-response relationships with applications to cancer epidemiology. *Annual Review of Public Health* **9** 451–482.

VACEK, P. M. (1997). Assessing the effect of intensity when exposure varies over time. *Stat. Med.* **16** 505–513.

WOOD, S. N. (2006). *Generalized Additive Models*: *An Introduction with R*. Chapman & Hall/CRC, Boca Raton, FL. MR2206355

WOOD, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 3–36. MR2797734

ZHANG, D., LIN, X. and SOWERS, M. (2007). Two-stage functional mixed models for evaluating the effect of longitudinal covariate profiles on a scalar outcome. *Biometrics* **63** 351–362. MR2370793

C. WANG
S. GAO
DEPARTMENT OF BIOSTATISTICS
INDIANA UNIVERSITY SCHOOL OF MEDICINE
AND RICHARD M. FAIRBANK SCHOOL OF PUBLIC HEALTH
INDIANA UNIVERSITY
410 W. 10TH STREET, SUITE 3000
INDIANAPOLIS, INDIANA 46202-3002
USA
E-MAIL: wang280@umail.iu.edu
sgao@iu.edu

H. LIU
GILEAD SCIENCES, INC
333 LAKESIDE DRIVE
FOSTER CITY, CALIFORNIA 94404
USA
E-MAIL: hai.liu@gilead.com