Research Article

Survival Data Analysis with Time-Dependent Covariates Using Generalized Additive Models

Masaaki Tsujitani,[1] Yusuke Tanaka,[2] and Masato Sakon[3]

[1] *Department of Engineering Informatics, Osaka Electro-Communication University, Osaka 572-8530, Japan*
[2] *Clinical Information Division, Data Science Center, EPS Corporation, Osaka 532-0003, Japan*
[3] *Nishinomiya Municipal Central Hospital, Hyogo 663-8014, Japan*

Correspondence should be addressed to Masaaki Tsujitani, ekaaf900@ricv.zaq.ne.jp

We discuss a flexible method for modeling survival data using penalized smoothing splines when the values of covariates change for the duration of the study. The Cox proportional hazards model has been widely used for the analysis of treatment and prognostic effects with censored survival data. However, a number of theoretical problems with respect to the baseline survival function remain unsolved. We use the generalized additive models (GAMs) with B splines to estimate the survival function and select the optimum smoothing parameters based on a variant multifold cross-validation (CV) method. The methods are compared with the generalized cross-validation (GCV) method using data from a long-term study of patients with primary biliary cirrhosis (PBC).

1. Introduction

Several prognostic models for PBC data have been developed using the Cox proportional hazards model, and the values of all covariates were determined at the time when the patient entered the study [1]. However, situations may exist in which the values of covariates change for the duration of the study. The time-dependent model uses follow-up data to estimate the effect of the evolution of the covariates during the course of the disease; see, for example, Cox [2], Altman and Stavola [3] and Collett [4].

Let t be a continuous lifetime variable and $\mathbf{x} = (x_1, \ldots, x_I)$ a vector of *time-fixed* covariates. The Cox's proportional hazards model postulates that the hazard at time t is the product of two components [5, 6]

$$h(t; \mathbf{x}) = h_0(t) \exp\left[\sum_{i=1}^{I} b_i x_i\right], \qquad (1)$$

where $\mathbf{b} = (b_1, \ldots, b_I)$ is a vector of coefficients. The proportional hazards assumption is that the baseline hazard $h_0(t)$ is

a function of t but does not involve the values of covariates $\mathbf{x}$ which are measured at the beginning of an interval to predict short-term survival.

We investigate PBC data for 312 patients who were seen at the Mayo Clinic and were monitored for the duration of the study, as described in Murtaugh et al. [7] and Therneau and Grambsch [8]. The Cox proportional hazards model was developed based on the relationship between survival and the patient characteristics observed when the patient entered the study. The precision of time-fixed models used in PBC is rather low, partly because these models are based on data for which the covariates were measured at the time when the patient entered the study.

For the analysis of data with time-dependent covariates, however, the survivor function for any individual depends on time t and the baseline hazard function. This means that the survivor function cannot be expressed as a power of the baseline survivor function and is generally difficult to obtain for any individual; see, for example, Kalbfleish and Prentice [9] and Marubini and Valsecchi [10]. The Mayo updated model (e.g., [7]), and the European new version model (e.g., [3, 11, 12]) have been commonly used to improve the accuracy

TABLE 1: Values of covariates for deceased patient #9.

Time interval l	Midpoint day $a_l^{(9)}$	Age (year) $x_{l1}^{(9)}$	Prothrombin time (sec) $x_{l2}^{(9)}$	Bilirubin (mg/dL) $x_{l3}^{(9)}$	$\delta_l^{(9)}$
1	92.0	42.54	11.0	3.2	0
2	272.5	43.04	12.5	7.0	0
3	542.0	43.53	11.2	4.2	0
4	875.0	44.52	14.1	13.5	0
5	1211.5	45.35	11.5	12.0	0
6	1837.0	46.36	11.5	16.2	0
7	2339.0	48.78	13.0	14.8	1

of survival predictions as a function of covariates measured at any time during the course of the disease. In the present article, we propose the variant multifold CV method for GAM when choosing the optimum smoothing parameters in order to estimate the survival function and predict the short-term survival (say, for the following six months) at any time during the course of the disease.

Another useful idea in our analysis is the concept of competing risk. There is "liver transplantation" in PBC data as competing risk. Competing risk has been treated as censored data. By adding the liver transplantation as one of time-dependent covariate, one can test the significance of liver transplantation.

2. Model Building

By extending the Cox proportional hazard model (1), a flexible survival model has been examined

$$h(t; \mathbf{x}) = h_0(t) \exp\left[\sum_{i=1}^{I} s_i(\mathbf{x}_i)\right], \qquad (2)$$

where $s_i(\mathbf{x}_i)$ is a spline function for the covariate $\mathbf{x}_i$ [13–15]. The proportional hazard model (2) used the time-fixed values of covariates as shown in Dickson et al. [1]. The estimates of hazard ratio by relative survival regression model [16] with time-dependent covariates are compared with that of Cox proportional hazard model. A new approach [17, 18] is proposed with PBC data, aiming to capture nonlinear patterns of bilirubin time courses and their relationship with survival time of patients. However, because most patients with PBC make repeated visits to the clinic, it is natural to ask the optimum timing of liver transplantation by predicting short-term survival at any time in the course of the disease.

The time-dependent covariates $\mathbf{X}_l^{\langle d \rangle} = (a_l^{\langle d \rangle}, x_{l1}^{\langle d \rangle}, \ldots, x_{lI}^{\langle d \rangle})$ are provided for patient #d, where $a_l^{\langle d \rangle}$ is the midpoint of time interval $[t_l^{\langle d \rangle}, t_{l+1}^{\langle d \rangle}]$ for the lth clinic visit. The event times may be subject to the usual random censoring. Then only the minimum of survival and censoring time with censoring indicator

$$\delta_l^{\langle d \rangle} = \begin{cases} 1 : \text{patient } \#d \text{ died at time-interval} \\ \quad \text{for the } l\text{th clinic visit} \\ 0 : \text{otherwise} \end{cases} \qquad (3)$$

are observed. The relative hazard $h(t)/h_0(t)$ then depends on time t, and thus the proportional hazards assumption is no longer satisfied, as described in Altman and Stavola [3] and Arjas [19].

For example, Table 1 shows the values of *age*, *prothrombin time*, and *bilirubin* as time-dependent covariates for dead patient #9; for details, see Table 4 in Murtaugh et al. [7]. Patients were scheduled to return for further observations at six months, 12 months, and yearly. Thus, $n = 312$ patients generate 1945 observations in total. The covariates values for each patient were allowed to vary with the time interval for the lth clinic visit.

A grouped version of Cox's proportional hazard model with time-fixed covariates has been considered in the framework of discrete grouped data for the feed-forward neural network. Given the continuous survivor time, piecewise models arise from the partition of the time axis into disjointed intervals. Biganzoli et al. [20, 21] show that, by treating the time interval as an input variable in a feed forward neural network, it is possible to estimate smoothed discrete hazards as conditional probabilities of failure. Biganzoli et al. [20] also pointed out that an advantage of this kind of data structure is the possibility of straightforward use of time-dependent covariates since each subject is represented, for each observation interval, by one input vector which can change across intervals. In order to apply this neural network approach, which is called partial logistic regression models [20], discretization of one-month or one-week intervals must be applied for the continuous survivor time with time-fixed covariates. We cannot determine which discretization, one-month or one-week intervals, must be applied; that, is the discretization is not originally unique. For the data in Table 1, however, the choice of discretization of the time axis for the partial logistic regression model is generally determined by clinical relevance, possibly according to the scheduled time intervals between follow-up visits.

The primary goal of the present study is to predict short-term survival in patients on the basis of measurements of several characteristics having time-dependent covariates $\mathbf{X}_l^{\langle d \rangle}$ for the purpose of facilitating the decision as to when to undertake liver transplantation. Based on partial logistic model due to Cox [22] and Efron [23] for the grouped data, Tsujitani and Sakon [24] have proposed a partial logistic model with a discrete hazard rate $h_l^{<d>}$ for ungrouped data

having time-dependent covariates

$$\ln\left(\frac{h_l^{\langle d\rangle}}{1-h_l^{\langle d\rangle}}\right) = \beta_0 + \gamma t_l^{\langle d\rangle} + \beta_1 x_{l1}^{\langle d\rangle} + \beta_2 x_{l2}^{\langle d\rangle} + \cdots + \beta_I x_{lI}^{\langle d\rangle},$$

$$d = 1, 2, \ldots, n, \tag{4}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_I)$ is a vector of coefficients. The modeled response is the logit of hazard rate, and the logit is linear in the covariates. However, this assumption is violated when covariate effects are best represented by smooth, nonlinear function. In recent years, a variety of powerful techniques have been developed for exploring the function form of effects. We examine here a flexible survival model GAM that does not require linearity of the covariate function by extending a generalized linear model (GLM); see, for example, Hastie and Tibshirani [13] and McCullagh and Nelder [25]. By identification of nonlinear covariate effects, we can estimate more accurately a patient's prognosis and thus determine a liver transplant based on prediction of short-term survival.

The linear predictor in (4) is specified as a sum of smooth functions $s(\mathbf{x})$ with twice continuous derivatives of some or all of the covariates for the discrete hazard rate $h_l^{<d>}$ of patient #d at the time interval l

$$\ln\left(\frac{h_l^{\langle d\rangle}}{1-h_l^{\langle d\rangle}}\right) = \beta_0 + s_0\left(t_l^{\langle d\rangle}\right) + s_1\left(x_{l1}^{\langle d\rangle}\right)$$

$$+ s_2\left(x_{l2}^{\langle d\rangle}\right) + \cdots + s_I\left(x_{lI}^{\langle d\rangle}\right). \tag{5}$$

The smooth functions in (5) can be represented as

$$s_0(x) = \sum_{j=1}^{q_0} \beta_j b_{0j}(t),$$

$$s_1(x) = \sum_{j=1}^{q_1} \beta_{q_0+j} b_{1j}(x),$$

$$\cdots \tag{6}$$

$$s_I(x) = \sum_{j=1}^{q_I} \beta_{q_{I-1}+j} b_{Ij}(x),$$

where $q_1, q_2, \ldots, q_I$ are the numbers of knots, and

$$\boldsymbol{\beta} = \left(\beta_0, \beta_1, \ldots, \beta_{q_0}, \beta_{q_0+1}, \beta_{q_0+2}, \ldots, \beta_{q_0+q_1},\right.$$

$$\ldots, \beta_{q_0+q_1+\cdots+q_{I-1}+1}, \beta_{q_0+q_1+\cdots+q_{I-1}+2}, \tag{7}$$

$$\left.\ldots, \beta_{q_0+q_1+\cdots+q_I}\right),$$

$$\mathbf{Z} = \begin{bmatrix} 1 & b_{01}(x_{11}) & \ldots & b_{0q_1}(x_{11}) & \cdot & b_{I-1,q_0+q_1+\cdots+q_{I-1}+1}(x_{I-1,1}) & \ldots & b_{I,q_0+q_1+\cdots+q_I}(x_{I1}) \\ 1 & b_{01}(x_{12}) & \ldots & b_{0q_1}(x_{12}) & \cdot & b_{I-1,q_0+q_1+\cdots+q_{I-1}+1}(x_{I-1,2}) & \ldots & b_{I,q_0+q_1+\cdots+q_I}(x_{I2}) \\ \cdot & \cdot & \ldots & \cdot & \cdot & \ldots & \ldots & \cdot \\ 1 & b_{01}(x_{11}) & \ldots & b_{0q_1}(x_{1n}) & \cdot & b_{I-1,q_0+q_1+\cdots+q_{I-1}+1}(x_{I-1,n}) & \ldots & b_{I,q_0+q_1+\cdots+q_I}(x_{In}) \end{bmatrix}. \tag{8}$$

The functions $b_{ij}(x_{ij})$ in the matrix $\mathbf{Z}$ are B-spline basis functions. Thus, (5) can be rewritten as

$$\ln\left(\frac{h_l^{\langle d\rangle}}{1-h_l^{\langle d\rangle}}\right) = \underset{n\times q}{\mathbf{Z}} \underset{q\times 1}{\boldsymbol{\beta}}; \quad q = \sum_{i=0}^{I} q_i, \tag{9}$$

where B-splines with 10 interiors knots will be used for each continuous covariate. The number of knots is arbitrary but appears to have little effect on the results, provided that the number is not too small, as described in Gray [15].

At the time interval for the lth clinic visit of patient #d, we define

$$\delta_l'^{\langle d\rangle} = \begin{cases} 1 : \text{patient } \#d \text{ was censored at} \\ \quad \text{time-interval for the } l\text{th clinic visit} \\ 0 : \text{otherwise,} \end{cases}$$

$$\mathbf{v}_l^{\langle d\rangle} = \left(\delta_1^{\langle d\rangle}, \delta_1'^{\langle d\rangle}, \delta_2^{\langle d\rangle}, \delta_2'^{\langle d\rangle}, \ldots, \delta_{l-1}^{\langle d\rangle}, \delta_{l-1}'^{\langle d\rangle}\right)$$

$$= (0, 0, \ldots, 0),$$

$$\mathbf{v}_l'^{\langle d\rangle} = \left(\mathbf{v}_l^{\langle d\rangle}, \delta_l^{\langle d\rangle}\right) = \left(0, 0, \ldots, 0, \delta_l^{\langle d\rangle}\right), \tag{10}$$

where $\mathbf{v}_l^{\langle d\rangle}$ is the history of dead and censored of time intervals for the first lth clinic visit of patient #d, and $\mathbf{v}_l'^{\langle d\rangle} = (\mathbf{v}_l^{\langle d\rangle}, \delta_l^{\langle d\rangle})$ is the same history extended to include $\delta_l^{<d>}$. Tsujitani and Sakon [24] derived the full log likelihood for all patients

$$\ln L = \ln L(\boldsymbol{\beta}) + \sum_{d=1}^{n} \sum_{l=1}^{l_d} \ln p\left(\delta_l'^{\langle d\rangle} \mid \mathbf{v}_l'^{\langle d\rangle}\right) \tag{11}$$

with partial log likelihood

$$\ln L(\boldsymbol{\beta}) = \sum_{d=1}^{n} \left\{ \sum_{l=1}^{l_d-1} \ln\left(1-h_l^{\langle d\rangle}\right) + \delta_{l_d}^{\langle d\rangle} \ln h_{l_d}^{\langle d\rangle} \right.$$

$$\left. + \left(1 - \delta_{l_d}^{\langle d\rangle}\right) \ln\left(1 - h_{l_d}^{\langle d\rangle}\right) \right\}. \tag{12}$$

The unknown parameters $\boldsymbol{\beta}$ in (9) can thus be estimated by maximizing the partial log likelihood (12), which is the log likelihood for the independent Bernoulli trial. Although $\ln L(\boldsymbol{\beta})$ is not a log likelihood in the usual sense, it possesses the usual asymptotic properties under fairly broad conditions; see, for example, Andelsen and Gill [26].

To avoid overfitting, such models are estimated by penalized maximum likelihood

$$\ln L(\boldsymbol{\beta}) = \sum_{d=1}^{n} \left\{ \sum_{l=1}^{l_d-1} \ln\left(1 - h_l^{\langle d\rangle}\right) + \delta_{l_d}^{\langle d\rangle} \ln h_{l_d}^{\langle d\rangle} \right.$$
$$\left. + \left(1 - \delta_{l_d}^{\langle d\rangle}\right) \ln\left(1 - h_{l_d}^{\langle d\rangle}\right) \right\} \qquad (13)$$
$$+ \frac{1}{2} \sum_{i=1}^{I} \lambda_i \int \{s_i''(t)\}^2 dt,$$

where λ_i are smoothing parameters that control the tradeoff between the fit and the smoothness, and s_i'' is the twice derivative of s_i with respect to t. The advantage of penalized estimates is enlightened in Wood ([27], Section 4.1).

Two model-fitting issues remain. The first concerns the selection of smoothing parameter λ_i in (13). The careful smoothing parameter choice is outweighed by the easy identification of a covariate's functional form, and the applicability of established inferential methods to short-term survival prediction. In order to select the smoothing parameters, the algorithm due to Wood [27–29] can be used by minimizing GCV as an approximation to leaving-one-out CV. For example, however, the dead patient #9 generated seven observations as shown in Table 1. Patients were scheduled to return for further observations at six months, 12 months, and yearly. It should be noted that this patient generated seven observations. Thus, $n = 312$ patients generate 1945 observations in total.

We propose a natural extension of v-fold CV algorithm by "leaving-one-out" CV based on each $n = 312$ patients. The ordinal v-fold CV divides the data randomly in v groups so that their sizes are as nearly equal as possible. The partition should be made to avoid possible biases, as described in Zhang [30]. In many problems, the ordinal v-fold CV is, thus, unsatisfactory in several respects for time-dependent covariates. Applying this kind of data structure to the CV algorithm, we obtain insights into how the partition of data should be done. A natural extension of v-fold CV algorithm by setting $v = n$ is to allow the deletion of the patient with several observations. The variant v-fold CV is given as follows:

Step 1. Split the original sample $\mathbf{X} = \{\mathbf{X}^{\langle 1\rangle}, \ldots, \mathbf{X}^{\langle d\rangle}, \ldots, \mathbf{X}^{\langle n\rangle}\}$ into n parts $\mathbf{X} = \{\mathbf{X}^{\langle 1\rangle} | \ldots | \mathbf{X}^{\langle d\rangle} | \ldots | \mathbf{X}^{\langle n\rangle}\}$, where $\mathbf{X}^{\langle d\rangle} = \{\mathbf{X}_1^{\langle d\rangle}, \ldots, \mathbf{X}_{l_d}^{\langle d\rangle}\}$.

Step 2. Fit the model to $\mathbf{X}^* = \{\mathbf{X}^{\langle 1\rangle} | \ldots | \mathbf{X}^{\langle d-1\rangle} | \mathbf{X}^{\langle d+1\rangle} | \ldots | \mathbf{X}^{\langle n\rangle}\}$ with the d-th subject (i.e., patient) deleted of the data, and predict $\sum_{l=1}^{l_d-1} \ln(1 - \hat{h}_l^{\langle d\rangle}) + \delta_{l_d}^{\langle d\rangle} \ln \hat{h}_{l_d}^{\langle d\rangle} + (1 - \delta_{l_d}^{\langle d\rangle}) \ln(1 - \hat{h}_{l_d}^{\langle d\rangle})$ for the deleted dth sample $\mathbf{X}^{\langle d\rangle}$.

Step 3. Do the above for $d = 1, 2, \ldots, n$ and combine the CV estimates

$$\text{CV} = \sum_{d=1}^{n} \left\{ \sum_{l=1}^{l_d-1} \ln\left(1 - \hat{h}_l^{\langle d\rangle}\right) + \delta_{l_d}^{\langle d\rangle} \ln \hat{h}_{l_d}^{\langle d\rangle} \right.$$
$$\left. + \left(1 - \delta_{l_d}^{\langle d\rangle}\right) \ln\left(1 - \hat{h}_{l_d}^{\langle d\rangle}\right) \right\}. \qquad (14)$$

A second issue is the goodness-of-fit test of the model. After choosing the optimum smoothing parameters via v-fold CV algorithm, the deviance allows us to test the goodness of fit

$$\text{Dev} = 2(\ln L_{\max} - \ln L_c), \qquad (15)$$

where $\ln L_c$ denotes the maximized partial log likelihood under some current GAM, and the log likelihood for the maximum (full) model $\ln L_{\max}$ is zero. The deviance given by (15) is, however, not even approximately a χ^2 distribution for the case in which ungrouped binary responses are available; see, for example, Landwehr et al. [31] and Tsujitani and Koshimizu [32] and Collett [33]. The number of degrees of freedom required for the test for significance using the assumed χ^2 distribution for the deviance is a contentious issue. No adequate distribution theory exists for the deviance. The reason for this is somewhat technical; for details, see Section 3.8 in Collett [33]. Consequently, the deviance on fitting a model to binary response data cannot be used as a summary measure of the goodness-of-fit test of the model.

Based on the above discussion, we employ bootstrapping to the deviance of (15) in order to obtain the goodness-of-fit test due to Efron and Tibshirani [34].

Step 1. Generate B bootstrap samples $\mathbf{X}^* = \{\mathbf{X}_l^{\langle 1\rangle *}, \ldots, \mathbf{X}_l^{\langle n\rangle *}\}$ from the original sample $\mathbf{X}$. Let $\mathbf{X}^*(b)$ denote the bth bootstrap sample.

Step 2. For the bootstrap sample $\mathbf{X}^*(b)$, the deviance of (15) is computed by $\text{Dev}(b) = 2[\ln L_f - \ln(\mathbf{X}^*(b); \hat{\boldsymbol{\beta}}^*(b))]$.

This process is repeated independently B times, and the computed values are arranged in ascending order.

Step 3. The value of the jth order statistic Dev* of the B replications can be taken as an estimate of the quantile of order $j/(B + 1)$.

Step 4. The estimate of the $100(1 - \alpha)$-th percentile (i.e., $\alpha\%$ critical point) of Dev* is used to test the goodness-of-fit of a model having a specified significance level $\alpha = 1 - j/(B + 1)$. The value of deviance of (15) being greater than the estimate of the percentile indicates that the model fits poorly.

3. Results

The survival function for our discretized situation is

$$S(t_l) = \prod_{1 \le i \le l} \left(1 - h_i^{\langle d\rangle}\right). \qquad (16)$$

TABLE 2: Optimum smoothing parameters.

Covariates	Variant v-fold CV	GCV
Time	0.1	0.00045
Age	10	0.000015
Prothrombin time	0.001	0.00039
Bilirubin	0.001	0.00000017

TABLE 3: Test of significance for the covariates (P values).

Covariates	Variant v-fold CV	GCV
Time	0.263	0.0028
Age	<0.0001	<0.0001
Prothrombin time	<0.0001	<0.0001
Bilirubin	<0.0001	<0.0001

The conditional probability $\Pr(t, t + \Delta t)$ of survival over a short-time interval Δt (say, six months) after time t can be estimated as

$$\widehat{\Pr}(t, \Delta t) = \frac{\hat{S}(t + \Delta t)}{\hat{S}(t)}. \tag{17}$$

By using variant v-fold CV and GCV, the optimum smoothing parameters for GAM are determined as shown in Table 2. Table 3 summarizes the P values to test the nonparametric effects of covariates for the model $s(\text{time}) + s(\text{age}) + s(\text{pro}) + s(\text{bili})$ with the optimum smoothing parameters and GCV. From Table 3, all covariates are highly significant for GCV; however, time is not significant for variant v-fold CV. GCV is only the approximation of leaving-one-out CV. Furthermore, the variant v-fold CV is leaving one-out CV based on each $n = 312$ patients. So variant v-fold CV is better than GCV. For the purpose of comparison, we included the results in the case using GCV in Tables 2 and 3. Table 4 shows the results of a number of models that were fit to the data.

The likelihood ratio (LR) statistic based on deviance can be conducted to test whether the spline effect provides a significantly better fit than a linear effect. Table 5 shows the test of significance for spline effects based on the models in Table 4. It is clear from Table 5 that the spline effects of prothrombin time and bilirubin are strongly significant. No spline provides a significantly better fit than a linear model for age. Thus, we accept the final model: age + $s(\text{pro}) + s(\text{bili})$ with the edf(effective degrees of freedom) 3.954 and 4.477 for $s(\text{pro})$ and $s(\text{bili})$, respectively. We used likelihood ratio test instead of information criteria as a valid alternative approach for model selection. From Table 4, however, it is found that the same final model is selected by using AIC.

Figure 1 shows the histogram of the bootstrapped $\text{Dev}(b)$ for the optimum model with $B = 400$. The bootstrap estimate of the 95th percentile (i.e., 5% critical point) Dev^* is $\text{Dev}^* = 734.59$. Comparison to $\text{Dev} = 663.65$ of (15) suggests that the model fits the data.

Figure 2 shows the prediction of the probability of surviving beyond the next six months for dead patient #9. For the purpose of comparison, the results obtained using partial logistic regression, the Mayo updated, and the European new
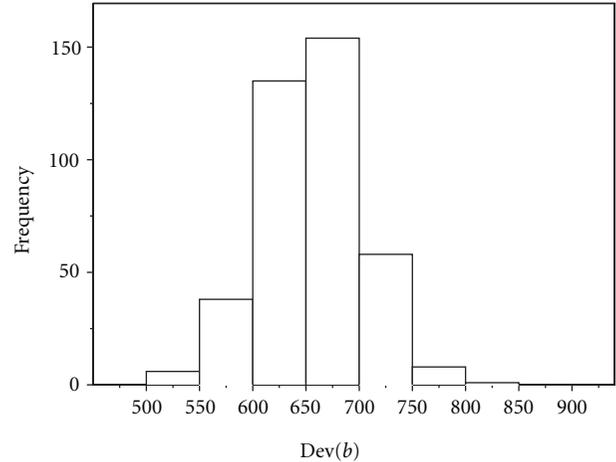
FIGURE 1: Histogram of the bootstrapped $\text{Dev}(b)$ for $B = 400$.

version models are also provided. Figure 2 also indicates that the six-month survival probability predicted by GAM are lower than those predicted by the other models. Because the patient #9 died, the lower predicted probabilities are better. The conditional probability $\Pr(t, \Delta t)$ of survival over a short time interval Δt (say, six months) after time t during the course of the disease can be predicted from data collected for censored and dead data.

For the graphical representation, the individual probabilities for predicted survival are averaged in order to compare the Mayo updated model, the European new version model, the partial logistic regression model, and GAM. We can predict the probability of survival over the following six months using the four models with respect to data and censored data out of 312 patients. For the group ($g = 1$) and the censored group ($g = 2$), the probability of surviving over the next Δt months is denoted by $\Pr_d^{[g]}(l, l + \Delta t)$ for the l-th clinic visit of patient #d. The average probability of survival over the next Δt months for the l-th clinic visit of patient #d in group g can be estimated as

$$S_g(l) = \frac{1}{n_l^{[g]}} \sum_{d=1}^{n_l^{[g]}} \Pr_d^{[g]}(l, l + \Delta t), \quad g = 1, 2, \tag{18}$$
$$l = 1, 2, \ldots, L,$$

where $n_l^{[g]}$ is the total number of patients for the l-th clinic visit in group g, and $\Pr_d^{[g]}(l)$ is the survival function $\Pr(l)$ of patient #d at the time interval l in group g; see, for example, Markus et al. [35], Marubini and Valsecchi [10], and Thomsen et al. [36].

Figure 3 shows a comparison of the probability of survival over the next six months using the four models with respect to dead and censored data among all 312 patients. The figure clarifies that

(i) for the case of dead data, the six-month survival probabilities predicted by GAM are lower than those predicted by the other models, and,

(ii) for the case of censored data, the difference among the four models is very small.

TABLE 4: Ranks and F values for nonparametric effects.

No.	Model	Deviance	d.f.	CV (AIC)
I	s(time) + s(age) + s(pro) + s(bili)	661.77	1933.16	691.99 (685.44)
II	s(age) + s(pro) + s(bili)	663.65	1934.57	690.54 (684.52)
III	s(time) + s(pro) + s(bili)	705.50	1933.93	739.03 (727.64)
IV	Age + s(pro) + s(bili)	663.65	1934.57	690.51 (684.51)
V	Age + s(pro) + bili	686.63	1937.58	708.44 (701.46)
VI	Age + pro + s(bili)	675.80	1938.02	696.53 (689.77)

TABLE 5: Test of significance for spline effects.

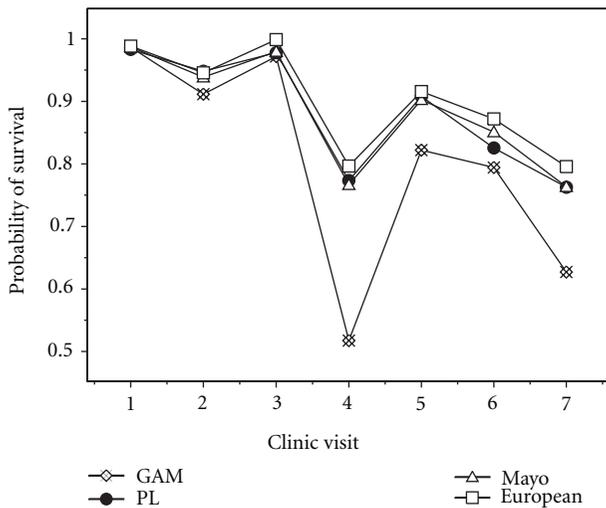| Spline effect	Δ	d.f.
Age	0.0039 (Model IV-II)	0.005
Prothrombin time	12.15 (Model VI-IV)	3.45
Bilirubin	22.97 (Model V-IV)	3.01

FIGURE 2: Probability of survival over the next six months using the four models with respect to dead patient #9.
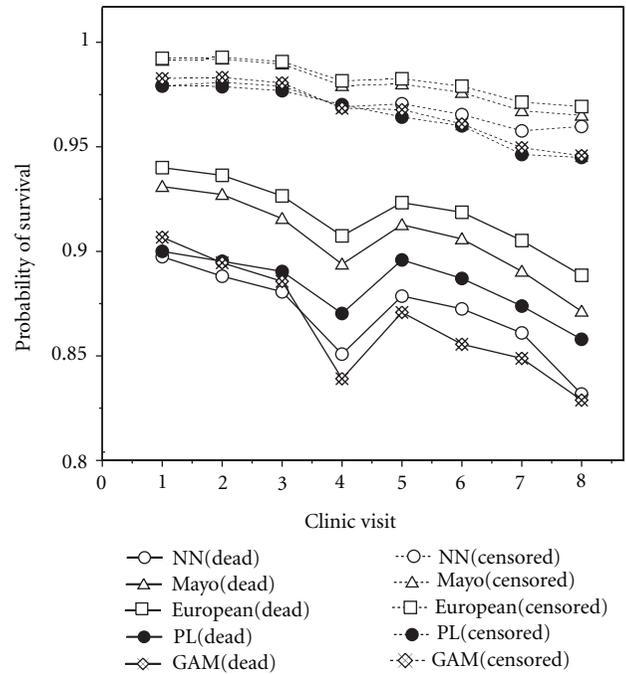

FIGURE 3: Probability of survival over the next six months using the four models with respect to dead and censored data among all 312 patients.

Figure 4 also shows the box and whisker plots of probability of survival over the next six months using GAM with respect to dead data among all 312 patients. It should be noted that the variance of probabilities of survival over the next six months is much higher in the fourth clinic visits than in other clinic visits. Another useful idea in our analysis is the concept of competing risk. There is "liver transplantation" in PBC data as competing risk. Competing risk has been treated as censored data. By adding $x_{l4}^{\langle d \rangle}$ as one of time-dependent covariate for the liver transplantation, one can test the significance of liver transplantation. The covariate for liver transplantation is taken as a binary variable (codes 0 before liver transplantation, 1 at liver transplantation) as shown in Giorgi and Gouvernet [16] and Crowley [37]. Table 6 shows the three types for the combination of "censored" and "liver transplantation." Table 7 shows the values of covariates for liver-transplanted patient #5.
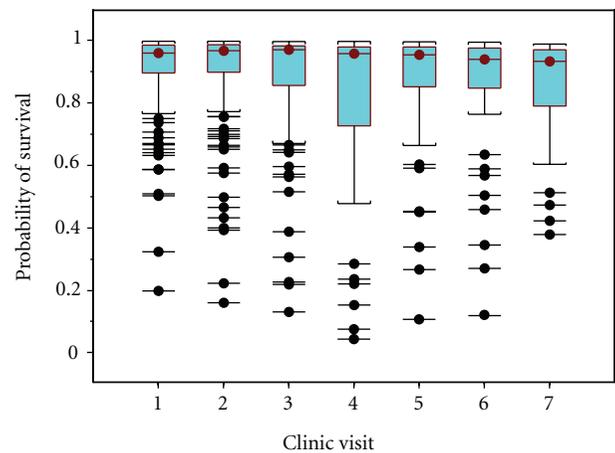

FIGURE 4: Box and whisker plots of probability of survival over the next six months using GAM with respect to dead data among all 312 patients.

TABLE 6: Three types of "censored" and "liver transplantation."

Dead patient		Censored patient		Liver-transplanted patient	
$\delta_l^{\langle d \rangle}$	Liver transplant	$\delta_l^{\langle d \rangle}$	Liver transplant	$\delta_l^{\langle d \rangle}$	Liver transplant
0	0	0	0	0	0
.	.	.	.	.	.
.	.	.	.	.	.
1	0	0	0	0	1

TABLE 7: Values of covariates for dead patient #5.

Time interval l	Midpoint day $a_l^{\langle 9 \rangle}$	Age (year) $x_{l1}^{\langle 9 \rangle}$	Prothrombin time (sec) $x_{l2}^{\langle 9 \rangle}$	Bilirubin (mg/dL) $x_{l3}^{\langle 9 \rangle}$	Liver trans-plantation $x_{l4}^{\langle 9 \rangle}$
1	99.5	38.11	10.9	3.4	0
2	295.0	38.65	10.7	1.9	0
3	580.0	39.18	10.5	2.5	0
4	933.5	40.21	11.4	5.7	0
5	1276.5	41.11	11.3	5.2	0
6	1480.0	42.09	13.9	19.0	1

TABLE 8: Test of significance for covariates using model (19).

Covariate	P-value
Age	<0.0001
Prothrombin time	<0.0001
Bilirubin	<0.0001

TABLE 9: Deviance and d.f.

Model	Deviance	d.f.
IV	663.65	1934.57
IV′	653.00	1934.00

In order to test the significance of "liver transplantation," we consider two models:

Model IV: age + s(pro) + s(bili).

Model IV′: age + s(pro) + s(bili) + liver transplantation. The values of deviance and d.f. are given in Table 8. The reduction in the value of deviance is $10.657 = 663.65 - 653.00$ on 0.973 d.f. This is significant at the 1% level.

For the purpose of the comparison, the hazard of the cumulative incidence function (CIF) may be modeled in the presence of competing risks. The model is based on

$$\gamma(t; \mathbf{x}) = \gamma_0(t) \exp\left[\sum_{i=1}^{I} \beta_i x_i \right], \quad (19)$$

where t is the time of the last observation (not the midpoint at the time interval l), γ is the hazard of the subdistribution, and γ_0 is the baseline hazard of the subdistribution ([38], Section 6.2). The P values are summarized in Table 8 to test the significance for covariates using the model (19). From Tables 3 and 8, there is little difference between our method and the CIF.

4. Discussion

In this paper, we introduced the probabilistic interpretation of GAM and constructed the maximum likelihood principle of GAM for the analysis of survival data having time-dependent covariates. We proposed the information criterion based on the variant v-fold CV when choosing the optimal smoothing parameters in application of GAM. Introducing the maximum likelihood principle into GAM, the deviance allows us to test the goodness-of-fit of GAM. The proposed methods were illustrated by comparing the probability of survival over the next six months using the Mayo-updated model, the European new version model, the partial logistic regression model, and GAM with respect to dead and censored data among PBC data. We expect that flexible methods for modeling survival data with time-dependent covariates using machine learning theory such that support vector machine will be very useful in this real-world contexts; see, for example, Hastie et al. [39]. Furthermore, smoothing spline ANOVA models by Gu [40] will enable us to include the interactions between the covariates.

We assume that there is only one cause of failure; that is, the event is allowed to occur only once for each patient. However, there is increasing interest to apply survival data sets with multiple events per patient [8, 41]. Wei et al. [42] analyzed bladder cancer data by modeling marginal distributions of multivariate failure time with proportional hazards models. The model may violate the proportional hazards assumption, even when the overall data set does not (Table 9). By modifying such as

$$\delta_l^{\langle d \rangle} = \begin{cases} 1 : \text{recurrent event for patient } \#d \\ \quad \text{occurred at the time-interval } l \\ 0 : \text{otherwise.} \end{cases} \quad (20)$$
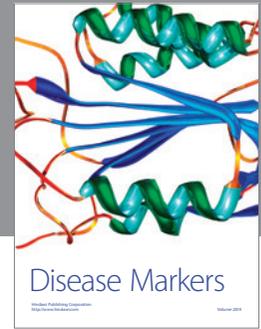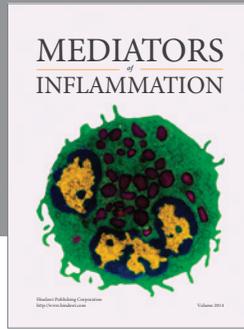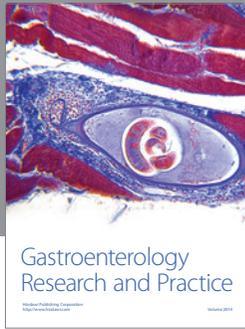
Equation (11) becomes

$$\ln L(\boldsymbol{\beta}) = \sum_{d=1}^{n} \sum_{l=1}^{l_d} \left\{ \delta_{l_d}^{\langle d \rangle} \ln h_{l_d}^{\langle d \rangle} + \left(1 - \delta_{l_d}^{\langle d \rangle}\right) \ln\left(1 - h_{l_d}^{\langle d \rangle}\right) \right\}. \tag{21}$$

Thus, the ideas presented in this paper can be extended to identification of prognostic factors relative to survival time in the case that the same event may recur during a follow-up study, and covariates values change with time.

References

[1] E. R. Dickson, P. M. Grambsch, T. R. Fleming, L. D. Fisher, and A. Langworthy, "Prognosis in primary biliary cirrhosis: model for decision making," *Hepatology*, vol. 10, no. 1, pp. 1–7, 1989.

[2] D. R. Cox, "Regression models and life tables (with discussion)," *Journal of the Royal Statistical Society: Series B*, vol. 34, pp. 187–220, 1972.

[3] D. G. Altman and B. L. de Stavola, "Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates," *Statistics in Medicine*, vol. 13, no. 4, pp. 301–341, 1994.

[4] D. Collett, *Modelling Survival Data in Medical Research*, Chapman & Hall/CRC, London, UK, 2nd edition, 2003.

[5] J. P. Klein and M. L. Moeschberger, *Survival Analysis*, Springer, New York, NY, USA, 2nd edition, 2003.

[6] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley, New York, NY, USA, 2nd edition, 2003.

[7] P. A. Murtaugh, E. R. Dickson, G. M. van Dam et al., "Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits," *Hapatology*, vol. 20, pp. 126–134, 1994.

[8] T. M. Therneau and P. M. Grambsch, *Modeling Survival Data: Extending the Cox Model*, Springer, New York, NY, USA, 2000.

[9] J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, John Wiley, New York, NY, USA, 2nd edition, 2002.

[10] E. Marubini and M. G. Valsecchi, *Analysing Survival Data from Clinical Trials and Observational Studies*, John Wiley, New York, NY, USA, 1995.

[11] E. Christensen, P. Schlichting, P. K. Andersen et al., "Updating prognosis and therapeutic effect evaluation in cirrhosis with Cox's multiple regression model for time-dependent variables," *Scandinavian Journal of Gastroenterology*, vol. 21, no. 163, pp. 174–1986.

[12] E. Christensen, D. G. Altman, J. Neuberger et al., "Updating prognosis in primary biliary cirrhosis using a timedependent Cox regression model," *Gastroenterology*, vol. 105, pp. 1865–1876, 1993.

[13] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*, Chapman & Hall, London, NY, USA, 1990.

[14] L. A. Sleeper and D. P. Harrington, "Regression splines in the Cox model with application to covariate effects in liver disease," *Journal of the American Statistical Association*, vol. 85, pp. 941–949, 1990.

[15] R. J. Gray, "Flexible methods for analyzing survival data using splines,with applications to breast cancer prognosis," *Journal of the American Statistical Association*, vol. 87, pp. 942–951, 1992.

[16] R. Giorgi and J. Gouvernet, "Analysis of time-dependent covariates in a regressive relative survival model," *Statistics in Medicine*, vol. 24, no. 24, pp. 3863–3870, 2005.

[17] J. Ding and J. L. Wang, "Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data," *Biometrics*, vol. 64, no. 2, pp. 546–556, 2008.

[18] R. Schoop, E. Graf, and M. Schumacher, "Quantifying the predictive performance of prognostic models for censored survival data with timedependent covariates," *Biometrics*, vol. 64, pp. 603–610, 2008.

[19] E. Arjas, "A graphical method for assessing goodness of fit in Cox's proportional hazards model," *Journal of the American Statistical Association*, vol. 83, pp. 204–212, 1988.

[20] E. Biganzoli, P. Boracchi, L. Mariani, and E. Marubini, "Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach," *Statistics in Medicine*, vol. 17, no. 10, pp. 1169–1186, 1998.

[21] E. Biganzoli, P. Boracchi, and E. Marubini, "A general framework for neural network models on censored survival data," *Neural Networks*, vol. 15, no. 2, pp. 209–218, 2002.

[22] D. R. Cox, "Partial likelihood," *Biometrika*, vol. 62, no. 2, pp. 269–276, 1975.

[23] B. Efron, "Logistic regression, survival analysis, and Kaplan-Meier curve," *Journal of the American Statistical Association*, vol. 83, pp. 414–425, 1988.

[24] M. Tsujitani and M. Sakon, "Analysis of survival data having time-dependent covariates," *IEEE Transactions on Neural Networks*, vol. 20, no. 3, pp. 389–394, 2009.

[25] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman & Hall, London, UK, 2nd edition, 1989.

[26] P. K. Andelsen and R. D. Gill, "Cox's regression model for counting process: a large sample study," *Annals of Statistics*, vol. 10, pp. 1100–1120, 1982.

[27] S. N. Wood, *Generalized Additive Models: An Introduction with R*, Chapman & Hall, London, UK, 2006.

[28] S. N. Wood, "Stable and efficient multiple smoothing parameter estimation for generalized additive models," *Journal of the American Statistical Association*, vol. 99, pp. 673–686, 2004.

[29] S. N. Wood, "Fast stable direct fitting and smoothness selection for generalized additive models," *Journal of the Royal Statistical Society: Series B*, vol. 70, no. 3, pp. 495–518, 2008.

[30] P. Zhang, "Model selection via multifold cross validation," *Annals of Statistics*, vol. 21, pp. 299–313, 1993.

[31] J. M Landwehr, D. Pregibon, and A. C. Shoemaker, "Graphical methods for assessing logistic regression models," *Journal of the American Statistical Association*, vol. 79, pp. 61–71, 1984.

[32] M. Tsujitani and T. Koshimizu, "Neural discriminant analysis," *IEEE Transactions on Neural Networks*, vol. 11, no. 6, pp. 1394–1401, 2000.

[33] D. Collett, *Modelling Binary Data*, Chapman & Hall/CRC, London, UK, 2nd edition, 2003.

[34] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, London, UK, 1993.

[35] B. H. Markus, E. R. Dickson, P. M. Grambsch et al., "Efficacy of liver transplantation in patients with primary biliary cirrhosis," *The New England Journal of Medicine*, vol. 320, no. 26, pp. 1709–1713, 1989.

[36] B. L. Thomsen, N. Keiding, and D. G. Altman, "A note on the calculation of expected survival, illustrated by the survival of liver transplant patients," *Statistics in Medicine*, vol. 10, no. 5, pp. 733–738, 1991.

[37] J. Crowley and M. Hu, "Covariance analysis of heart transplant survival data," *Journal of the American Statistical Association*, vol. 72, pp. 27–36, 1977.

[38] M. Pintilie, *Competing Risks: A Practical Perspective*, John Wiley, New York, NY, USA, 2006.

[39] T. J. Hastie, R. J. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, NY, USA, 2001.

[40] C. Gu, "Smoothing Spline ANOVA Models," pp. Springer–New York, NY, USA, 2002.

[41] E. T. Lee and J. W. Wang, *Statistical Methods for Survival Data Analysis*, John Wiley, New York, NY, USA, 3rd edition, 2003.

[42] L. J. Wei, D. Y. Lin, and L. Wessfeld, "Regression analysis of multivariate incomplete failure time data by modeling marginal distributions," *Journal of the American Statistical Association*, vol. 84, pp. 1065–1073, 1989.

The Scientific
World Journal

Gastroenterology
Research and Practice

MEDIATORS
of
INFLAMMATION

Journal of
Diabetes Research

Disease Markers

Journal of
Immunology Research

International Journal of
Endocrinology

PPAR Research

BioMed
Research International

Hindawi

Submit your manuscripts at
http://www.hindawi.com

Journal of
Ophthalmology

Stem Cells
International

eCAM

Evidence-Based
Complementary and
Alternative Medicine

Journal of
Obesity

Journal of
Oncology

Computational and
Mathematical Methods
in Medicine

Behavioural
Neurology

Parkinson's
Disease

AIDS
Research and Treatment

Oxidative Medicine and
Cellular Longevity