Biometrika (2012), **99**, 3, *pp.* 615–630
© 2012 Biometrika Trust
Printed in Great Britain

doi: 10.1093/biomet/ass018
Advance Access publication 29 April 2012

Predictive accuracy of covariates for event times

By LI CHEN

Markey Cancer Center and Department of Biostatistics, University of Kentucky, Lexington, Kentucky 40536, U.S.A.
lichenuky@uky.edu

D. Y. LIN and DONGLIN ZENG

Department of Biostatistics, CB# 7420, University of North Carolina, Chapel Hill, North Carolina 27599-7420, U.S.A.
lin@bios.unc.edu dzeng@bios.unc.edu

Summary

We propose a graphical measure, the generalized negative predictive function, to quantify the predictive accuracy of covariates for survival time or recurrent event times. This new measure characterizes the event-free probabilities over time conditional on a thresholded linear combination of covariates and has direct clinical utility. We show that this function is maximized at the set of covariates truly related to event times and thus can be used to compare the predictive accuracy of different sets of covariates. We construct nonparametric estimators for this function under right censoring and prove that the proposed estimators, upon proper normalization, converge weakly to zero-mean Gaussian processes. To bypass the estimation of complex density functions involved in the asymptotic variances, we adopt the bootstrap approach and establish its validity. Simulation studies demonstrate that the proposed methods perform well in practical situations. Two clinical studies are presented.

Some key words: Censoring; Negative predictive value; Positive predictive value; Prognostic accuracy; Receiver operating characteristic curve; Recurrent event; Survival data; Transformation model.

1. Introduction

In cohort studies, potentially censored times until the occurrence of a particular event, such as death or disease recurrence, are often recorded, along with data on certain covariates, including demographic, clinical and genetic factors. An important objective of such studies is to determine which set of covariates is most predictive of event times and thus can be used to select patients for targeted therapy or to refine disease prevention strategies. For example, the gene expression breast cancer study of van de Vijver et al. (2002) was designed to assess whether a 70-gene signature can improve the prediction of survival times and the selection of patients for adjuvant chemotherapy over standard clinical factors. Clinicians are extremely interested in the percentage of patients who can be spared adjuvant chemotherapy, and in the long-term survival probabilities for those patients. To address this type of question, a statistical measure directly related to survival probabilities is needed to quantify the predictive accuracy of covariates and to compare the predictive accuracy of different sets of covariates.

Several methods exist for assessing the predictive accuracy of covariates for survival time or a single event. One approach extends the proportion of variation explained by covariates, R^2, for the linear regression model to the proportional hazards model (Korn & Simon, 1990; Schemper, 1990; Schemper & Henderson, 2000), and a second approach is to use a suitable loss function, such as the Brier score, to measure residual variation (Graf et al., 1999). These measures, however, may lack clinical relevance. A more recent approach is to extend the receiver operating characteristic curve for binary response to survival time (Heagerty & Zheng, 2005; Uno et al., 2007). This measure is focused on classification rather than prediction and is not directly related to survival probabilities.

For a binary disease status D and a binary covariate Z, each taking values 0 and 1, standard measures of predictive accuracy include the positive predictive value and negative predictive value, which are defined as $\text{PPV} = \text{pr}(D = 1 \mid Z = 1)$ and $\text{NPV} = \text{pr}(D = 0 \mid Z = 0)$, respectively. The predictive values pertain to the probability of disease or nondisease given a positive or negative result for a dichotomous test and are thus of direct interest to the end users of the test. When Z is continuous and a larger value of Z corresponds to a higher risk of disease, Moskowitz & Pepe (2004) defined the positive and negative predictive curves $\text{PPV}(v) = \text{pr}\{D = 1 \mid F_Z(Z) > v\}$ and $\text{NPV}(v) = \text{pr}\{D = 0 \mid F_Z(Z) \leqslant v\}$ $(0 \leqslant v \leqslant 1)$, where $F_Z(\cdot)$ is the distribution function of Z. Using $F_Z(Z)$ instead of Z yields two appealing properties. First, the positive and negative predictive curves do not depend on the scales upon which the continuous covariates are measured and are thus comparable among different covariates. Second, it can be shown that $\text{PPV}(v) = 1 - \{\text{pr}(D = 0) - v\text{NPV}(v)\}(1 - v)^{-1}$, which means that a higher negative predictive curve corresponds to a higher positive predictive curve, so either curve is sufficient to quantify predictive accuracy. These measures were recently extended to survival data by Zheng et al. (2008), who defined time-dependent predictive curves as $\text{PPV}(t, v) = \text{pr}\{T \leqslant t \mid F_Z(Z) > v\}$ and $\text{NPV}(t, v) = \text{pr}\{T > t \mid F_Z(Z) \leqslant v\}$, where T is the survival time; the extension pertains to a single event with a single covariate. Uno et al. (2007) considered the predictive curves at a fixed time-point for a single event with multiple covariates by standardizing the linear predictor with a known transformation function and proposed inverse probability weighted estimators under completely-at-random censoring.

In the present paper, we propose a new concept, the generalized negative predictive function, to quantify the predictive accuracy of multiple covariates for both survival time and recurrent event times. In the special case of a single event with a single covariate, our function reduces to that of Zheng et al. (2008). We show that the proposed function is maximized at the set of covariates that is truly related to event times and thus can be used to quantify and compare the predictive accuracy of covariates. No such theoretical results are available even for a binary outcome or a single event with a single covariate. We develop nonparametric estimators for the proposed function, allowing censoring to depend on covariates, and establish their large-sample properties. Because a linear combination of covariates is obtained from a possibly misspecified working model to predict event times, the asymptotic distributions of the estimators for the case of multiple covariates are considerably more complicated than the case of a single covariate.

2. Generalized negative predictive function

Let $N^*(t)$ be the counting process which records the number of events the subject has experienced by time t. Let X be the vector of all potential covariates, and Z be a $p \times 1$ subset of X. We wish to quantify the ability of Z in predicting $N^*(\cdot)$.

We first consider the case where Z is a single covariate. Suppose that a larger value of Z corresponds to a higher event rate. Then at each time-point $u \in (0, t]$, we can define a negative predictive curve, $\text{pr}\{\text{d}N^*(u) = 0 \mid Y^*(u) = 1, F_Z(Z) \leqslant v\}$, for the subjects who are at risk at that time, where $Y^*(u)$ is a process indicating whether the subject is at risk for experiencing an event at time u. By taking the product integral of these curves in the time interval $(0, t]$, we define the generalized negative predictive function of Z as

$$G(t, v) = \prod_{0 \leqslant u \leqslant t} \mathrm{pr}\{\mathrm{d}N^*(u) = 0 \mid Y^*(u) = 1, F_Z(Z) \leqslant v\}. \tag{1}$$

For single-event data, $N^*(\cdot)$ has a single jump at the event time T and $Y^*(t) = I(T \geqslant t)$, so that equation (1) reduces to $\textsc{npv}(t, v) = \mathrm{pr}\{T > t \mid F_Z(Z) \leqslant v\}$ of Zheng et al. (2008). For recurrent event data, equation (1) is equivalent to

$$G(t, v) = \exp\left[-\int_0^t E\{\mathrm{d}N^*(u) \mid Y^*(u) = 1, F_Z(Z) \leqslant v\}\right],$$

which can be rewritten as $\exp[-E\{N^*(t) \mid F_Z(Z) \leqslant v\}]$ because $Y^*(\cdot) = 1$; therefore, $-\log\{G(t, v)\}$ can be interpreted as the predictive number of events by time t in the subgroup with $F_Z(Z) \leqslant v$. If the counting process has a Poisson structure such that the occurrence of an event is independent of the prior event history conditional on Z, then (1) can also be interpreted as the probability of no event by time t in the subgroup with $F_Z(Z) \leqslant v$.

We now consider the case that Z includes multiple covariates. To use Z to predict $N^*(\cdot)$, a common approach is to obtain a linear combination of Z under a regression model. We consider a semiparametric working model in the form of

$$E\{N^*(t) \mid Z\} = g(t, \beta^{\mathrm{T}} Z), \tag{2}$$

where β is a $p \times 1$ vector of unknown regression parameters, and $g(t, v)$ is a strictly increasing function of v for all $t > 0$. For single-event data, we can use the class of semiparametric transformation models (Dabrowska & Doksum, 1988; Zeng & Lin, 2006), which includes the proportional hazards and proportional odds models as special cases. For recurrent event data, we can use the class of semiparametric transformation mean models (Lin et al., 2001).

Let $\hat{\beta}$ be the estimator of β described in Zeng & Lin (2006). For single-event data and recurrent events with a Poisson structure, $\hat{\beta}$ is the nonparametric maximum likelihood estimator. For recurrent events without a Poisson structure, $\hat{\beta}$ is viewed as a maximum pseudo-likelihood estimator. It can be shown that $\hat{\beta}$ converges to some constant vector β^* under mild regularity conditions even when model (2) is misspecified. The predictive accuracy of Z can be quantified by the generalized negative predictive function of $\beta^{*\mathrm{T}} Z$,

$$G_{Z,\beta^*}(t, v) = \prod_{0 \leqslant u \leqslant t} \mathrm{pr}\{\mathrm{d}N^*(u) = 0 \mid Y^*(u) = 1, F_{\beta^{*\mathrm{T}} Z}(\beta^{*\mathrm{T}} Z) \leqslant v\}, \tag{3}$$

which is equal to $\mathrm{pr}\{T > t \mid F_{\beta^{*\mathrm{T}} Z}(\beta^{*\mathrm{T}} Z) \leqslant v\}$ for single-event data, and to $\exp[-E\{N^*(t) \mid F_{\beta^{*\mathrm{T}} Z}(\beta^{*\mathrm{T}} Z) \leqslant v\}]$ for recurrent event data.

As a function of both t and v, the generalized negative predictive function can be plotted against the cut-off percentage v for a specific t. By comparing such plots for two sets of covariates, one can tell the difference in the event-free probability at time t between the two groups that include v percentage of low-risk subjects as defined by the two sets of covariates. The low-risk group defined by the more predictive set of covariates has a higher survival probability. As in the case of the receiving operating characteristic curve, the area under the curve can be used as a summary measure. One can set a cut-off of survival probability to define low risk. For breast cancer patients, the low-risk group is usually defined to have a 5-year survival probability of at least 0·95. Then drawing a horizontal line at 0·95 would yield the percentage of patients who can be defined as low risk. For this type of application, the function is required to be monotone. The proposed function can also be plotted against t for a specific v. By comparing such plots for two sets of covariates, one can tell how the difference in the event-free probability changes over time for the two low-risk groups defined by the two sets of covariates.

We now provide a formal justification for using the generalized negative predictive function to compare the predictive accuracy of different sets of covariates. Using Lemma A1 of Appendix 1, we can obtain an optimality property of $G_{Z,\beta^*}(t,v)$ that this function is maximized at the true linear combination of covariates and that the maximization holds uniformly for all t and v. Specifically, suppose that Z_0 is the subset of X truly related to $N^*(\cdot)$ in that $E\{N^*(t) \mid X\}$ depends on Z_0 only and that β_0 is the true value of β associated with Z_0 under model (2). Then $G_{Z_0,\beta_0}(t,v) \geqslant G_{Z,\beta}(t,v)$ for all $t > 0$ and $v \in (0,1)$.

3. Inference procedures

3·1. *Preliminaries*

Let C denote the censoring time. In the presence of censoring, the observed counting process is $N(t) = N^*(t \wedge C)$ and the at-risk process is $Y(t) = Y^*(t)I(C \geqslant t)$, where $a \wedge b = \min(a,b)$, and $I(\cdot)$ is the indicator function. The data consist of n independent replicates $\{N_i(t), Y_i(t), X_i : t \in [0,\tau]\}$ $(i = 1, \ldots, n)$, where τ is the endpoint of the study. We use the estimator $\hat\beta$ from the working model (2) to construct nonparametric estimators for $G_{Z,\beta^*}(t,v)$. We consider the situation that C is independent of $N^*(\cdot)$ and X, as well as the situation that C is independent of $N^*(\cdot)$ conditional on X. We refer to these two censoring mechanisms as completely-at-random censoring and covariate-dependent censoring, respectively.

3·2. *Single-event data under completely-at-random censoring*

Under completely-at-random censoring, $\text{pr}\{\text{d}N^*(u) = 0 \mid Y^*(u) = 1, F_{\beta^{*\text{T}}Z}(\beta^{*\text{T}}Z) \leqslant v\} = \text{pr}\{\text{d}N(u) = 0 \mid Y(u) = 1, F_{\beta^{*\text{T}}Z}(\beta^{*\text{T}}Z) \leqslant v\}$. Thus, we estimate $G_{Z,\beta^*}(t,v)$ by

$$\hat G(t,v) = \prod_{0<u\leqslant t}\left[1 - \frac{\sum_{i=1}^n I\{F_n(\hat\beta^{\text{T}}Z_i) \leqslant v\}\text{d}N_i(u)}{\sum_{i=1}^n I\{F_n(\hat\beta^{\text{T}}Z_i) \leqslant v\}Y_i(u)}\right],$$

where $F_n(\cdot)$ is the empirical distribution of $\hat\beta^{\text{T}}Z$. Note that $\hat G(t,v)$ is the Kaplan–Meier estimator of the survival function in the subgroup with $F_n(\hat\beta^{\text{T}}Z) \leqslant v$.

In Appendix 2, we show that the process $n^{1/2}\{\hat G(t,v) - G_{Z,\beta^*}(t,v)\}$ converges weakly to a zero-mean Gaussian process and is asymptotically equivalent to the process

$$n^{-1/2}\sum_{i=1}^n \xi_{1i}(t,v) + n^{1/2}\{\hat G(t;\hat c_v,\hat\beta) - \hat G(t;c_v,\beta^*)\}, \tag{4}$$

where

$$\xi_{1i}(t,v) = -G_{Z,\beta^*}(t,v)I(\beta^{*\text{T}}Z_i \leqslant c_v)\int_0^t \frac{\text{d}N_i(u) - Y_i(u)\,\text{d}\Lambda(t;c_v,\beta^*)}{E\{I(\beta^{*\text{T}}Z \leqslant c_v)Y(u)\}},$$

$$G(t; c, \beta) = \prod_{0<u\leqslant t} \left[1 - \frac{E\{I(\beta^{\mathrm{T}} Z \leqslant c)\, \mathrm{d}N^*(u)\}}{E\{I(\beta^{\mathrm{T}} Z \leqslant c) Y^*(u)\}}\right],$$

$$\hat{G}(t; c, \beta) = \prod_{0<u\leqslant t} \left\{1 - \frac{\sum_{i=1}^{n} I(\beta^{\mathrm{T}} Z_i \leqslant c)\, \mathrm{d}N_i(u)}{\sum_{i=1}^{n} I(\beta^{\mathrm{T}} Z_i \leqslant c) Y_i(u)}\right\},$$

$c_v = F^{-1}_{\beta^{*\mathrm{T}} Z}(v)$, $\hat{c}_v = F_n^{-1}(v)$, and $\mathrm{d}\Lambda(t; c, \beta) = E\{\mathrm{d}N^*(t) \mid Y^*(t) = 1, \beta^{\mathrm{T}} Z \leqslant c\}$. The second term in (4) is asymptotically equivalent to $n^{-1/2} \sum_{i=1}^{n} \xi_{2i}(t, v)$, where $\xi_{2i}(t, v)$ is the value of $\xi_2(t, v)$, defined in equation (A5) of Appendix 2, for the ith subject. Note that $\xi_2(t, v)$ involves conditional density functions given multi-dimensional covariates, which are difficult to estimate directly. To avoid this difficulty, we approximate the asymptotic distribution by bootstrapping the observations $\{N_i(t), Y_i(t), X_i : t \in [0, \tau]\}$ $(i = 1, \ldots, n)$. We recommend using 1000 bootstrap samples. The validity of the bootstrap is proved in Appendix 4.

3·3. *Single-event data under covariate-dependent censoring*

When censoring is not completely at random, the Kaplan–Meier type estimator for the generalized negative predictive function presented in § 3·2 is no longer consistent because $\mathrm{pr}\{\mathrm{d}N^*(u) = 0 \mid Y^*(u) = 1, F_{\beta^{*\mathrm{T}} Z}(\beta^{*\mathrm{T}} Z) \leqslant v\} \neq \mathrm{pr}\{\mathrm{d}N(u) = 0 \mid Y(u) = 1, F_{\beta^{*\mathrm{T}} Z}(\beta^{*\mathrm{T}} Z) \leqslant v\}$. We decompose the set of all potential covariates X into two parts, U and W, where U is known to be independent of C, and C is independent of $N^*(\cdot)$ conditional on W. Let q be the number of continuous components in W. Note that $\mathrm{pr}\{T > t \mid F_{\beta^{*\mathrm{T}} Z}(\beta^{*\mathrm{T}} Z) \leqslant v\} = E[E\{I(T > t, \beta^{*\mathrm{T}} Z \leqslant c_v) \mid W\}]/v = E\{S(t; c_v, \beta^*, W) I(\beta^{*\mathrm{T}} Z \leqslant c_v)\}/v$, where $S(t; c, \beta, w) = \mathrm{pr}(T > t \mid \beta^{\mathrm{T}} Z \leqslant c, W = w)$. Since the cumulative hazard function corresponding to $S(t; c, \beta, w)$, denoted by $\Lambda(t; c, \beta, w)$, is equal to $\int_0^t E\{\mathrm{d}N(u) I(\beta^{\mathrm{T}} Z \leqslant c) \mid W = w\}/E\{I(\beta^{\mathrm{T}} Z \leqslant c) Y(u) \mid W = w\}$, we propose to estimate $S(t; c, \beta, w)$ by a kernel-smoothed Kaplan–Meier type estimator

$$\hat{S}(t; c, \beta, w) = \prod_{0<u\leqslant t} \left[1 - \frac{\sum_{i=1}^{n} K\{(W_i - w)/h\} I(\beta^{\mathrm{T}} Z_i \leqslant c)\, \mathrm{d}N_i(u)}{\sum_{i=1}^{n} K\{(W_i - w)/h\} I(\beta^{\mathrm{T}} Z_i \leqslant c) Y_i(u)}\right],$$

where $K(x) = \tilde{K}(\| x_c \|) I(x_d = 0)$, x_d and x_c are the discrete and continuous components of x, respectively, $\| \cdot \|$ is the Euclidean norm, $\tilde{K}(\cdot)$ is a kernel function for which there exists some integer l such that $\int y^j \tilde{K}(y)\, \mathrm{d}y = 0$ $(j = 1, \ldots, l-1)$, $\int y^l \tilde{K}(y)\, \mathrm{d}y \neq 0$ and $2l - q > 0$, and h is a bandwidth such that $nh^q \to \infty$ and $nh^{2l} \to 0$ as $n \to \infty$. When $q \leqslant 3$, we can use a symmetric smooth probability density function as the kernel function. In light of the relationship between $S(t; c_v, \beta^*, W)$ and $G_{Z,\beta^*}(t, v)$, we can then estimate $G_{Z,\beta^*}(t, v)$ by

$$\tilde{G}(t, v) = v^{-1} n^{-1} \sum_{i=1}^{n} \hat{S}(t; \hat{c}_v, \hat{\beta}, W_i) I(\hat{\beta}^{\mathrm{T}} Z_i \leqslant \hat{c}_v).$$

In Appendix 3, we show that the process $n^{1/2}\{\tilde{G}(t, v) - G_{Z,\beta^*}(t, v)\}$ converges weakly to a zero-mean Gaussian process. As in § 3·2, the asymptotic distribution can be approximated by the bootstrap method.

Due to the curse of dimensionality and the difficulty in choosing an appropriate kernel function for $q > 3$, the proposed method may not be very useful when q is large. For $q > 3$, we suggest obtaining a linear combination of W under a proportional hazards model for the censoring time and then applying the proposed method to the linear combination of W.

3·4. *Recurrent event data*

Under completely-at-random censoring, the inference procedures developed in § 3·2 can be readily applied to recurrent event data because the formulae do not involve specific forms of single-event data. Under covariate-dependent censoring, the fact that $Y^*(\cdot) = 1$ for recurrent event data yields

$$G_{Z,\beta^*}(t, v) = \exp[-v^{-1}E\{\mu(t; c_v, \beta^*, W)I(\beta^{*\mathrm{T}}Z \leqslant c_v)\}],$$

where $\mu(t; c, \beta, w) = \int_0^t E\{\mathrm{d}N^*(u) \mid W = w, \beta^{\mathrm{T}}Z \leqslant c\}$, which can be consistently estimated by

$$\hat{\mu}(t; c, \beta, w) = \int_{u\leqslant t} \frac{\sum_{i=1}^n K\{(W_i - w)/h\}I(\beta^{\mathrm{T}}Z_i \leqslant c)\,\mathrm{d}N_i(u)}{\sum_{i=1}^n K\{(W_i - w)/h\}I(\beta^{\mathrm{T}}Z_i \leqslant c)Y_i(u)}.$$

Thus, we propose to estimate $G_{Z,\beta^*}(t, v)$ by

$$\tilde{G}(t, v) = \exp\left\{-v^{-1}n^{-1}\sum_{i=1}^{n} \hat{\mu}(t; \hat{c}_v, \hat{\beta}, W_i)I(\hat{\beta}^{\mathrm{T}}Z_i \leqslant \hat{c}_v)\right\}.$$

We show in Appendix 3 that the process $n^{1/2}\{\tilde{G}(t, v) - G_{Z,\beta^*}(t, v)\}$ converges weakly to a zero-mean Gaussian process. The asymptotic distribution can again be approximated by the bootstrap method.

4. SIMULATION STUDIES

The first set of simulation studies concerns the performance of the proposed estimators under completely-at-random censoring. To mimic the breast cancer data of § 5·1, we generated survival times from the Weibull regression model $\lambda(t \mid X) = k\lambda^{-1}(t/\lambda)^{k-1}\exp(\beta^{\mathrm{T}}X)$, where $\lambda(t \mid X)$ is the conditional hazard function of T given X, X is zero-mean trivariate normal with a covariance matrix that is the sample covariance matrix of age, tumor size and gene score of the breast cancer data, and the parameters k, λ, β are estimated from the breast cancer data. We also generated recurrent event times from the random-effect intensity model: $\lambda(t \mid X, \xi) = \xi k\lambda^{-1}(t/\lambda)^{k-1}\exp(\beta^{\mathrm{T}}X)$, where $\lambda(t \mid X, \xi)$ is the conditional intensity function of $N^*(t)$ given X and ξ, and ξ is a gamma random variable with mean 1 and standard deviation 0·5. For completely-at-random censoring, we generated censoring times from a Weibull distribution whose parameters are estimated from the breast cancer data. Table 1 summarizes the results for the estimator $\hat{G}(t, v)$ at $v = 0{\cdot}25$, 0·5, 0·75 and 1·0 for $t = 5$ years with $n = 300$. The parameter estimator has little bias. The standard error estimator reflects the true variability very well, and the confidence intervals have proper coverage probabilities. We also evaluated $\tilde{G}(t, v)$, the estimator allowing covariate-dependent censoring, and found it to be nearly as efficient as $\hat{G}(t, v)$. For survival data, we also evaluated the inverse probability weighted estimator of Uno et al. (2007). The relative efficiencies of $\hat{G}(t, v)$ to Uno et al.'s estimator were found to be 3·80, 2·02, 1·34 and 1·0 for $v = 0{\cdot}25$, 0·5, 0·75 and 1·0, respectively.

To assess the performance of the proposed estimators under covariate-dependent censoring, we adopted the above set-up but generated censoring times from the transformation model $\Lambda(t \mid X) = H\{0{\cdot}1t\exp(\beta_c\sum_{i=1}^m X_i)\}$, where $\Lambda(t \mid X)$ is the cumulative hazard function of C conditional on X, $H(\cdot)$ is a specific increasing function, and $X_1, \ldots, X_m$ are standard normal variables with 0·3 correlations with the gene score. To compare the estimators under completely-at-random censoring and covariate-dependent censoring, we chose $H(x) = x$, $m = 3$ and $\beta_c =$

Table 1. *Simulation results for the estimator $\hat{G}(t, v)$ for data under completely-at-random censoring*

	Survival data					Recurrent event data				
v	True	Bias	SE	SEE	CP	True	Bias	SE	SEE	CP
0·25	97·0	0·1	2·0	2·0	93	97·0	−0·0	2·1	2·0	91
0·50	94·9	0·1	1·8	1·9	96	94·8	−0·0	1·9	1·9	96
0·75	92·0	0·1	1·8	1·9	96	91·9	−0·0	1·9	2·0	96
1·0	85·6	−0·0	2·1	2·1	94	83·8	−0·0	2·5	2·5	94

True, the true value (×100) of $G(t = 5 \text{ years}, v)$; Bias, the sampling bias (×100); SE, the sampling standard error (×100); SEE, the sampling mean (×100) of the standard error estimator; CP, the coverage probability (×100) of the 95% confidence interval. Each entry is based on 2000 replicates.

Table 2. *Simulation results for the estimators $\tilde{G}(t, v)$ and $\hat{G}(t, v)$ for survival data under covariate-dependent censoring*

			$\tilde{G}(t, v)$				$\hat{G}(t, v)$			
n	v	True	Bias	SE	SEE	CP	Bias	SE	SEE	CP
300	0·25	97·0	−0·1	3·3	2·9	96	−0·1	2·7	2·6	87
	0·50	94·9	−0·0	2·7	2·6	96	−0·2	2·4	2·4	96
	0·75	92·0	−0·1	2·5	2·5	96	−0·4	2·3	2·4	96
	1·0	85·6	−0·2	2·6	2·5	94	−1·5	2·5	2·5	92
500	0·25	97·0	−0·0	2·5	2·2	95	−0·1	2·1	2·1	93
	0·50	94·9	−0·0	2·1	2·0	95	−0·2	1·8	1·9	96
	0·75	92·0	0·0	2·0	1·9	95	−0·5	1·8	1·8	96
	1·0	85·6	−0·1	2·1	2·0	94	−1·4	2·0	2·0	90

True, the true value (×100) of $G(t = 5 \text{ years}, v)$; Bias, the sampling bias (×100); SE, the sampling standard error (×100); SEE, the sampling mean (×100) of the standard error estimator; CP, the coverage probability (×100) of the 95% confidence interval. Each entry is based on 2000 replicates.

0·5 to produce a censoring rate of 70·2%, which is close to that of the breast cancer data. We used the 3-dimensional independent Gaussian kernel density and set the bandwidth to $\sigma n^{-7/24}$, where σ is the maximum of the standard deviations of X_1, X_2 and X_3. As shown in Table 2, the estimator $\tilde{G}(t, v)$ has very small bias; the variance estimator is accurate and the confidence intervals have proper coverage probabilities. In contrast, the estimator $\hat{G}(t, v)$ has severe bias and improper confidence intervals. We also evaluated the inverse probability weighted estimator of Uno et al. (2007), which turned out to be severely biased and inefficient, the bias being −0·239, −0·138, −0·066, and −0·009 for $v = 0{\cdot}25, 0{\cdot}5, 0{\cdot}75$ and 1·0, respectively, and the corresponding standard errors being 0·086, 0·052, 0·033 and 0·025.

Finally, we considered the situation that censoring depends on a large number of covariates. We increased the number of covariates in the above censoring model to 10 or 20 and set $H(x) = x$ or $H(x) = \log(1 + x)$. We used the one-dimensional Gaussian kernel density and set the bandwidth to $\sigma n^{-1/3}$, where σ is the standard deviation of the linear predictor in the proportional hazards model for the censoring time. The results for $\beta_c = 0{\cdot}1$ and $m = 10$ and 20 are summarized in the Supplementary Material. The proposed method performs well even when the censoring model is incorrectly specified.

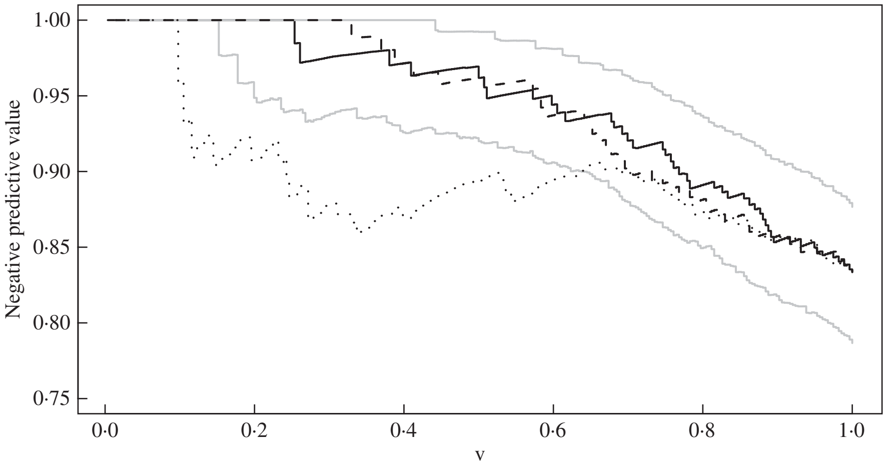

Fig. 1. Estimation of $G(t, v)$ at $t = 5$ years in the breast cancer data. The black solid, dashed and dotted curves pertain to the combination of the gene signature, age and tumor size, gene signature alone, and the combination of age and tumor size, respectively. The grey curves are the 95% pointwise confidence limits corresponding to the black solid curve.

5. Examples

5·1. *Gene expressions in breast cancer patients*

van de Vijver et al. (2002) developed a database of 295 breast cancer patients. In this dataset, the time to death along with a 70-gene signature and some clinical factors were recorded. A total of 276 patients had complete information on the gene signature and clinical factors, and the survival times for 73·6% of them were censored. The main purpose of this study was to determine whether the 70-gene signature can improve the prediction of survival times above the clinical factors and thus be used to guide the selection of patients for adjuvant chemotherapy. We use this dataset to illustrate the generalized negative predictive function.

For our illustration, we compare three sets of covariates: the combination of age and tumor size, the 70-gene signature alone, and the combination of the gene signature, age and tumor size. We consider the proportional hazards models with these three sets of covariates. Because the censoring time depends on tumor size, the p-value being 0·0009, we allow censoring to depend on tumor size in our analysis. Figure 1 displays the 5-year negative predictive curves for the three sets. The gene signature indeed improves predictive accuracy above clinical factors and the gene signature alone is enough for prediction. If clinicians wish to spare 50% of the breast cancer patients adjuvant chemotheraphy, then those patients selected by the combination of the gene signature, age and tumor size will have a 5-year survival probability of 0·962, with a 95% confidence interval of (0·920, 0·992), whereas those patients selected by age and tumor size will have a 5-year survival probability of only 0·89. The area under the curve for the combination of the gene signature, age and tumor size is 0·945, which is significantly larger than that of the combination of age and tumor size and is very close to that of the gene signature alone, the two differences being 0·050 and 0·001, respectively, and the corresponding 95% confidence intervals being (0·020, 0·083) and (−0·008, 0·013). Because the effect of dependent censoring is fairly small in this case, the curves under completely-at-random censoring are very similar to their counterparts in Fig. 1 and are omitted.

We now demonstrate how to use the negative predictive curve to guide the selection of patients for adjuvant chemotherapy. Because of the toxicity of chemotherapy, low-risk patients should be

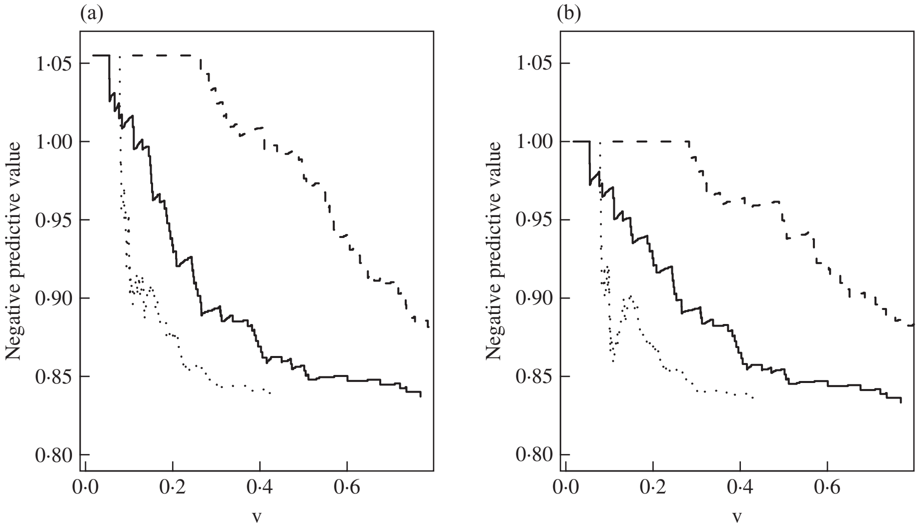

Fig. 2. Estimates of the NPV ($t = 5$ years, v) of Uno et al. (2007) in the breast cancer data. The black solid, dashed and dotted curves pertain to the combination of the gene signature, age and tumor size, gene signature alone, and the combination of age and tumor size, respectively. (a) is based on the inverse probability weighted estimator of Uno et al. (2007); (b) is based on the Kaplan–Meier type estimator.

spared adjuvant chemotherapy. Suppose that we define low risk to be a 5-year survival probability of at least 95%. Then what percentage of patients will be assigned to the low-risk group and thus spared adjuvant chemotherapy based on the 70-gene signature? The answer is given by G^{-1} ($t = 5$, 0·95) = 58·0%, with a 95% confidence interval of (34·4%, 70·7%). In contrast, this percentage is only 10·5%, with a 95% confidence interval of (1·1%, 31·2%) based on age and tumor size. Thus, the use of the 70-gene signature can spare substantially more patients adjuvant chemotherapy.

Finally, we compare our negative predictive curves with those of Uno et al. (2007). Figure 2 displays the estimates of the curves of Uno et al. (2007), for which the linear predictor is standardized by the transformation function $g(y) = 1 - \exp\{-\exp(y)\}$. Figure 2 shows that the x-values of the curves for different linear predictors have different ranges, although the transformation function restricts the linear predictors to the interval (0, 1). In contrast, the x-values of our curves, as shown in Fig. 1, always have the range (0, 1). In Fig. 2, the curve for the combination of the gene signature, age and tumor size is considerably lower than that of the gene signature alone. This is not sensible because the combination of the gene signature, age and tumor size should at least have the same predictive accuracy as the gene signature alone. Therefore, the negative predictive curves standardized by a known transformation function cannot be used to directly compare the predictive accuracy of different covariates. Figure 2 also shows that the inverse probability weighted estimator can yield inappropriate estimates for small cut-off values.

5·2. *Exacerbations of respiratory symptoms in patients with cystic fibrosis*

Patients with cystic fibrosis often suffer from repeated exacerbations of respiratory symptoms. A randomized clinical trial was conducted to evaluate the efficacy of rhDNase, a highly purified recombinant enzyme, in reducing the rate of of exacerbations (Therneau & Hamilton, 1997). By the end of the trial, 139 of 324 untreated patients and 104 of 321 treated patients had experienced at least one exacerbation; 42 untreated patients and 39 treated patients had at least two

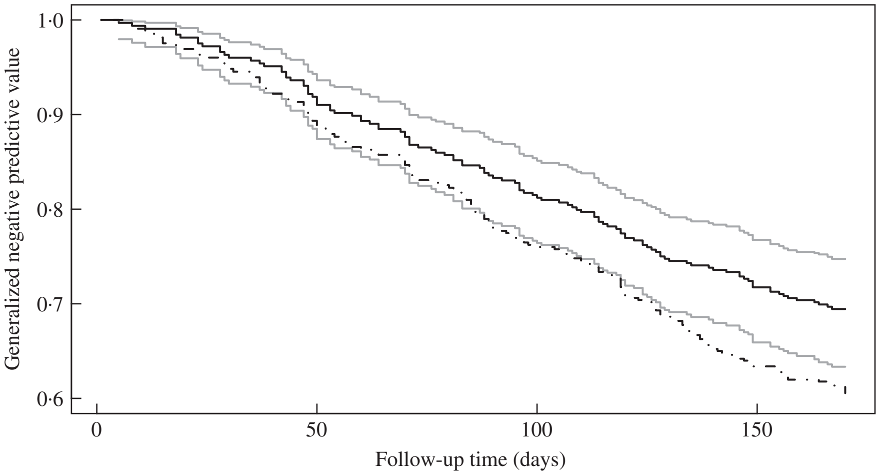

Fig. 3. Estimates of $G(t, v)$ at $v = 0{\cdot}498$ in the rhDNase study. The black solid and black dash-dotted curves pertain to the point estimates for the model with forced expiratory volume and treatment and the model with treatment alone, respectively. The grey curves pertain to the 95% pointwise confidence limits associated with the black solid curve.

exacerbations. In addition to the times to exacerbations, the baseline level of forced expiratory volume in 1 second was recorded for each patient.

We use the proposed methods to evaluate the importance of forced expiratory volume in predicting the number of exacerbations. We compare the predictive accuracy of forced expiratory volume and treatment to that of treatment alone. We consider the class of transformation mean models for these two sets of covariates. The choices of the Box–Cox transformations $H(x) = \{(1+x)^{\rho} - 1\}/\rho$ with $\rho = 0{\cdot}5, 1, 2$ and logarithmic transformations $H(x) = \log(1+rx)/r$ with $r = 0{\cdot}5, 1, 2$ yield highly similar estimates for $G(t, v)$. Since treatment is binary, $G(t, v)$ for the two sets of covariates are comparable only at the threshold $v = 0{\cdot}498$, the proportion of treated patients. Figure 3 displays the estimated $G(t, v)$ versus t at $v = 0{\cdot}498$ based on the proposed estimator under completely-at-random censoring. The model with forced expiratory volume and treatment has much higher generalized negative predictive values than the model with treatment alone, especially for large t, indicating that forced expiratory volume plays an important role in predicting the number of exacerbations. The results based on the estimator allowing the censoring to depend on forced expiratory volume and treatment are virtually the same and omitted.

6. REMARKS

In the proposed methods, the working model is used only to obtain the coefficients for the linear combination of multiple covariates. The estimation of the generalized negative predictive function itself is purely nonparametric, so the predictive accuracy of different sets of covariates can be compared in a fair manner. Additional simulation studies revealed that the performance of the proposed methods is robust to misspecification of the transformation function. In real studies, the estimates of the proposed function, especially the relative magnitudes of the estimates for different sets of covariates, have been found to be similar under different transformation functions.

We have focused on the class of transformation models and its nonparametric maximum likelihood estimators (Zeng & Lin, 2006). We can also consider other survival models and other estimators provided that the influence functions for the regression parameters exist. Indeed, our methods are designed to assess the predictive accuracy of any model, whether it holds or not. If there is only a single covariate or the regression parameters are estimated from a previous study, then the influence functions for the regression parameters are set to zero in the proposed methods. Our work was partly motivated by the desire to compare the predictive accuracy of five gene-expression signatures for breast cancer (Fan et al., 2006), in which case the coefficients for each linear predictor were determined by a previous study and thus considered fixed in our estimation of the generalized negative predictive function; the results will be reported elsewhere.

Uno et al. (2007) considered the positive and negative predictive curves by thresholding the linear predictor with a known transformation function. As shown in § 5·1, such curves cannot be used to directly compare the predictive accuracy of different covariates. In addition, the inverse probability weighted estimator requires completely-at-random censoring and can yield severe bias under covariate-dependent censoring. Even under completely-at-random censoring, the inverse probability weighted estimator is less efficient than our Kaplan–Meier type estimator, especially for small cut-off values. Our proposed function thresholds the linear predictor by its distribution function and can be used to compare the predictive accuracy of different covariates. The use of the distribution function of the linear predictor creates additional challenges in deriving the asymptotic properties since the distribution function needs to be estimated and the additional variation needs to be accounted for.

The proposed function has nice connections with the receiving operating characteristic curve and overall misspecification rate. It can be shown that, for a given time t if the predictive curve of $\beta_1^{\mathrm{T}} Z_1$ is uniformly higher than that of $\beta_2^{\mathrm{T}} Z_2$, then the corresponding receiving operating characteristic curve is also uniformly higher, and vice versa. In addition, a uniformly higher predictive curve implies a lower overall misspecification rate. The key difference of the predictive curve from the receiving operating characteristic curve is that it depends on the patient's survival probabilities and is thus more relevant for prediction. A main purpose of developing a prediction rule is to divide patients with low versus high survival probabilities. The prediction rule based on the receiving operating characteristic curve is not directly connected to survival probabilities, so it is difficult to determine whether the low/high risk groups determined by the prediction rule have high/low survival probabilities. As for the overall misspecification rate, it has the same problem as the receiving operating characteristic curve and does not distinguish between false positive and false negative errors.

There is a growing interest in assessing the predictive accuracy of time-dependent covariates. Henderson et al. (2002) proposed a modification of the explained variation measure of Schemper & Henderson (2000), Schoop et al. (2008) extended the Brier score, and Zheng & Heagerty (2007) extended the receiving operating characteristic method. All these extensions are confined to single-event data. The generalized negative predictive function proposed in this paper can easily incorporate time-dependent covariates by replacing Z in (3) by $Z(u)$. It is straightforward to extend the two estimators for recurrent event data and the estimator for single-event data under completely-at-random censoring to the setting of time-dependent covariates, but the extension is difficult for single-event data under dependent censoring because our estimator relies on the equation $G_{Z,\beta^*}(t, v) = \mathrm{pr}\{T > t \mid F_{\beta^{*\mathrm{T}} Z}(\beta^{*\mathrm{T}} Z) \leqslant v\}$, which does not hold for time-dependent covariates.

Acknowledgement

This research was supported by the National Institutes of Health, U.S.A. The authors thank two referees and an associate editor for their helpful comments.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes additional simulation results and Lemma S1 of Appendix 4.

APPENDIX 1

Optimality of the generalized negative predictive function

LEMMA A1. *Let $\xi(\cdot)$ be a stochastic process, and Z_0 and Z be subvectors of the random vector X. Assume that the linear combinations $\beta_0^{\mathrm{T}} Z_0$ and $\beta^{\mathrm{T}} Z$ are continuous and that $E\{\xi(t) \mid X\} = g_0(t, \beta_0^{\mathrm{T}} Z_0)$, where $g_0(t, v)$ is a strictly decreasing function of v for all $t > 0$. Then $E\{\xi(t) \mid F_{\beta_0^{\mathrm{T}} Z_0}(\beta_0^{\mathrm{T}} Z_0) \leqslant v\} \geqslant E\{\xi(t) \mid F_{\beta^{\mathrm{T}} Z}(\beta^{\mathrm{T}} Z) \leqslant v\}$ for all $t > 0$ and $v \in (0, 1)$. In addition, if there exists some t such that the equality holds for all v, then* $\mathrm{pr}\{F_{\beta_0^{\mathrm{T}} Z_0}(\beta_0^{\mathrm{T}} Z_0) = F_{\beta^{\mathrm{T}} Z}(\beta^{\mathrm{T}} Z)\} = 1$.

Proof. Let $Y_0 = F_{\beta_0^{\mathrm{T}} Z_0}(\beta_0^{\mathrm{T}} Z_0)$, $Y = F_{\beta^{\mathrm{T}} Z}(\beta^{\mathrm{T}} Z)$, and $f(t, y) = g_0\{t, F^{-1}_{\beta_0^{\mathrm{T}} Z_0}(y)\}$. Then $E\{I(Y_0 \leqslant v)\} = E\{I(Y \leqslant v)\} = v$, $E\{\xi(t) \mid X\} = f(t, Y_0)$, and $f(t, y)$ is strictly decreasing in y for all $t > 0$. Clearly, $vE\{\xi(t) \mid F_{\beta_0^{\mathrm{T}} Z_0}(\beta_0^{\mathrm{T}} Z_0) \leqslant v\} = E\{f(t, Y_0) I(Y_0 \leqslant v)\}$, which can be written as

$$E\{f(t, Y_0) I(Y_0 \leqslant v, Y \leqslant v)\} + E\{f(t, Y_0) I(Y_0 \leqslant v, Y > v)\}. \tag{A1}$$

Since $f(t, y)$ is strictly decreasing in y for all $t > 0$, the second term in (A1) is greater than or equal to $f(t, v)E\{I(Y_0 \leqslant v, Y > v)\}$, where the equality holds if and only if $\mathrm{pr}(Y_0 \leqslant v, Y > v) = 0$. Since $E\{I(Y_0 \leqslant v)\} = E\{I(Y \leqslant v)\} = v$, we have $E\{I(Y_0 \leqslant v, Y > v)\} = E\{I(Y_0 > v, Y \leqslant v)\}$. This equality, together with the fact that $f(t, y)$ is strictly decreasing in y for all $t > 0$, implies that the second term in (A1) is greater than or equal to $E\{f(t, Y_0) I(Y_0 > v, Y \leqslant v)\}$. Thus, (A1) is greater than or equal to $E\{f(t, Y_0) I(Y \leqslant v)\}$ or $vE\{\xi(t) \mid F_{\beta^{\mathrm{T}} Z}(\beta^{\mathrm{T}} Z) \leqslant v\}$. Hence, the results in Lemma A1 hold. □

Remark 1. If $\beta_0^{\mathrm{T}} Z_0$ or $\beta^{\mathrm{T}} Z$ is not continuous, then $E\{\xi(t) \mid F_{\beta_0^{\mathrm{T}} Z_0}(\beta_0^{\mathrm{T}} Z_0) \leqslant v\} \geqslant E\{\xi(t) \mid F_{\beta^{\mathrm{T}} Z}(\beta^{\mathrm{T}} Z) \leqslant v\}$ for all $t > 0$ and v in the common range of the distribution functions of $\beta_0^{\mathrm{T}} Z_0$ and $\beta^{\mathrm{T}} Z$.

APPENDIX 2

Weak convergence of $n^{1/2}\{\hat{G}(t, v) - G_{Z,\beta^*}(t, v)\}$

In this section, we prove the weak convergence of $n^{1/2}\{\hat{G}(t, v) - G_{Z,\beta^*}(t, v)\}$ for single-event data. The proof for recurrent events is similar and is omitted.

Let $\mathcal{P}_n$ and $\mathcal{P}$ denote the empirical measure and the distribution under the true model, respectively. For a measurable function f and measure Q, the integral $\int f\,\mathrm{d}Q$ is abbreviated as Qf. We assume that there exists a continuous covariate in Z, denoted as $Z^{(k)}$. Let $Z^{(-k)}$ be the remaining components of Z, and β_k and $\beta_{(-k)}$ be the components of β corresponding to $Z^{(k)}$ and $Z^{(-k)}$, respectively. Let $h_1(t, x; y)$ and $h_2(x; y)$ be, respectively, the conditional densities of $(T, Z^{(k)})$ and $Z^{(k)}$ given $Z^{(-k)} = y$. We require that $h_1(t, x; y)$ and $h_2(x; y)$ are continuously differentiable. Let ψ be the influence function of $\hat{\beta}$ in that $n^{1/2}(\hat{\beta} - \beta^*) = n^{1/2}\mathcal{P}_n \psi + o_p(1)$.

Clearly,

$$n^{1/2}\{\hat{G}(t, v) - G_{Z,\beta^*}(t, v)\} = n^{1/2}\{\hat{G}(t; c_v, \beta^*) - G(t; c_v, \beta^*)\} + n^{1/2}\{\hat{G}(t; \hat{c}_v, \hat{\beta}) - \hat{G}(t; c_v, \beta^*)\}. \tag{A2}$$

The first term in (A2) pertains to the Kaplan–Meier estimator among the subjects with $\beta^{*\mathrm{T}} Z \leqslant c_v$ and is asymptotically equivalent to $n^{1/2}(\mathcal{P}_n - \mathcal{P})\xi(t; c_v, \beta^*)$, where

$$\xi(t; c, \beta) = -\left[G(t; c, \beta) I(\beta^{\mathrm{T}} Z \leqslant c) \int_0^t \frac{\mathrm{d}N(u) - Y(u)\,\mathrm{d}\Lambda(u; c, \beta)}{E\{I(\beta^{\mathrm{T}} Z \leqslant c) Y(u)\}}\right].$$

By the result of Exercise 14 of van der Vaart & Wellner (1996, p. 152), the class of functions $x \mapsto I(\beta^{\mathrm{T}} x \leqslant c)$ with β and c ranging over $\mathcal{R}^p$ and $\mathcal{R}$, respectively, is VC-subgraph and is thus $\mathcal{P}$-Donsker by Theorems 2.6.7 and 2.5.2 of van der Vaart & Wellner (1996). In addition, the classes of functions $\{N(t) : t \in [0, \tau]\}$, $\{Y(t) : t \in [0, \tau]\}$, and $\{\Lambda(t; c, \beta) : t \in [0, \tau]\}$ are $\mathcal{P}$-Donsker by Lemma 4.1 of Kosorok (2008). Thus, $\{\xi(t; c, \beta) : t \in [0, \tau], c \in \mathcal{R}, \beta \in \mathcal{R}^p\}$ is a $\mathcal{P}$-Donsker class by the preservation properties of $\mathcal{P}$-Donsker classes (van der Vaart & Wellner, 1996, § 2.10). This result, together with the fact that $n^{1/2}\{\hat{G}(t; \hat{c}_v, \hat{\beta}) - \hat{G}(t; c_v, \beta^*)\} = n^{1/2}\{G(t; \hat{c}_v, \hat{\beta}) - G(t; c_v, \beta^*)\} + n^{1/2}\{\hat{G}(t; \hat{c}_v, \hat{\beta}) - G(t; \hat{c}_v, \hat{\beta})\} + n^{1/2}\{\hat{G}(t; c_v, \beta^*) - G(t; c_v, \beta^*)\}$, implies that the second term in (A2) is asymptotically equivalent to

$$n^{1/2}\{G(t; \hat{c}_v, \hat{\beta}) - G(t; c_v, \beta^*)\} + n^{1/2}(\mathcal{P}_n - \mathcal{P})\{\xi(t; \hat{c}_v, \hat{\beta}) - \xi(t; c_v, \beta^*)\}. \tag{A3}$$

As to be shown later, $n^{1/2}(\hat{c}_v - c_v)$ is weakly convergent. Thus, the second term in (A3) converges uniformly to zero in probability by Lemma 19.24 of van der Vaart (1998). The first term in (A3) is asymptotically equivalent to

$$\left.\frac{\partial G(t; c, \beta^*)}{\partial c}\right|_{c=c_v} n^{1/2}(\hat{c}_v - c_v) + \left.\frac{\partial G(t; c_v, \beta)}{\partial \beta}\right|_{\beta=\beta^*} n^{1/2}(\hat{\beta} - \beta^*),$$

where $\partial G(t; c, \beta)/\partial \beta$ can be obtained as follows, and $\partial G(t; c, \beta)/\partial c$ in the same manner. Let $g_1(t, \beta, c) = E\{I(T > t) I(\beta^{\mathrm{T}} Z \leqslant c)\}$, and $g_2(\beta, c) = E\{I(\beta^{\mathrm{T}} Z \leqslant c)\}$. Then for $\beta_k > 0$, $g_1(t, \beta, c) = E[E\{I(T > t) I(\beta^{\mathrm{T}} Z \leqslant c) | Z^{(-k)}\}] = E \int_{-\infty}^{\beta_k^{-1}(c - \beta_{(-k)} Z^{(-k)})} \int_t^{\infty} h_1(s, x; Z^{(-k)})\, \mathrm{d}s \mathrm{d}x$. Simple algebraic manipulations yield

$$\frac{\partial G(t; c, \beta)}{\partial \beta} = \frac{\partial g_1(t, \beta, c)/\partial \beta - E\{I(T > t) | \beta^{\mathrm{T}} Z \leqslant c\} \partial g_2(\beta, c)/\partial \beta}{E\{I(\beta^{\mathrm{T}} Z \leqslant c)\}},$$

where

$$\begin{aligned}
\frac{\partial g_1(t, \beta, c)}{\partial \beta_k} &= -c\beta_k^{-2} E \int_t^{\infty} h_1\{s, \beta_k^{-1}(c - \beta_{(-k)} Z^{(-k)}); Z^{(-k)}\}\, \mathrm{d}s, \\
\frac{\partial g_1(t, \beta, c)}{\partial \beta_{(-k)}} &= -\beta_k^{-1} E \left[Z^{(-k)} \int_t^{\infty} h_1\{s, \beta_k^{-1}(c - \beta_{(-k)} Z^{(-k)}); Z^{(-k)}\}\, \mathrm{d}s \right], \\
\frac{\partial g_2(\beta, c)}{\partial \beta_k} &= -c\beta_k^{-2} E \left[h_2\{\beta_k^{-1}(c - \beta_{(-k)} Z^{(-k)}); Z^{(-k)}\} \right], \\
\frac{\partial g_2(\beta, c)}{\partial \beta_{(-k)}} &= -\beta_k^{-1} E[Z^{(-k)} h_2\{\beta_k^{-1}(c - \beta_{(-k)} Z^{(-k)}); Z^{(-k)}\}].
\end{aligned}$$

We now establish the weak convergence of $n^{1/2}(\hat{c}_v - c_v)$. Clearly,

$$\begin{aligned}
n^{1/2}\{F_n(c) - F_{\beta^{*\mathrm{T}} Z}(c)\} = {} & n^{1/2}(\mathcal{P}_n - \mathcal{P}) I(\beta^{*\mathrm{T}} Z \leqslant c) + n^{1/2}\mathcal{P}\{I(\hat{\beta}^{\mathrm{T}} Z \leqslant c) - I(\beta^{*\mathrm{T}} Z \leqslant c)\} \\
& + n^{1/2}(\mathcal{P}_n - \mathcal{P})\{I(\hat{\beta}^{\mathrm{T}} Z \leqslant c) - I(\beta^{*\mathrm{T}} Z \leqslant c)\}.
\end{aligned} \tag{A4}$$

The second term on the right-hand side of (A4) is asymptotically equivalent to $\partial g_2(\beta^*, c)/\partial \beta\, n^{1/2}(\hat{\beta} - \beta^*)$, and the third term converges to zero in probability. It then follows from Lemma 12.8 of Kosorok (2008) that $n^{1/2}(\hat{c}_v - c_v)$ is asymptotically equivalent to $-\{\mathrm{d}F_{\beta^{*\mathrm{T}} Z}(c)/\mathrm{d}c|_{c=c_v}\}^{-1} n^{1/2}(\mathcal{P}_n - \mathcal{P})\{I(\beta^{*\mathrm{T}} Z \leqslant c_v) + \partial g_2(\beta^*, c_v)/\partial \beta \psi\}$.

Combining the above results, we conclude that $n^{1/2}\{\hat{G}(t,v)-G_{Z,\beta^*}(t,v)\}$ is asymptotically equivalent to $n^{1/2}\mathcal{P}_n\{\xi_1(t,v)+\xi_2(t,v)\}$, where $\xi_1(t,v)=\xi(t;c_v,\beta^*)$ and

$$\xi_2(t,v)=-\{I(\beta^{*\mathrm{T}}Z\leqslant c_v)-v\}\left.\frac{\partial G(t;c,\beta^*)/\partial c}{\mathrm{d}F_{\beta^{*\mathrm{T}}Z}(c)/\mathrm{d}c}\right|_{c=c_v}$$
$$-\left\{\left.\frac{\partial G(t;c,\beta^*)/\partial c}{\mathrm{d}F_{\beta^{*\mathrm{T}}Z}(c)/\mathrm{d}c}\right|_{c=c_v}\left.\frac{\partial g_2(\beta,c_v)}{\partial\beta}\right|_{\beta=\beta^*}-\left.\frac{\partial G(t;c_v,\beta)}{\partial\beta}\right|_{\beta=\beta^*}\right\}\psi. \qquad \text{(A5)}$$

APPENDIX 3

Weak convergence of $n^{1/2}\{\tilde{G}(t,v)-G_{Z,\beta^*}(t,v)\}$

In this section, we prove the weak convergence of $n^{1/2}\{\tilde{G}(t,v)-G_{Z,\beta^*}(t,v)\}$ for single-event data. The proof for recurrent events is similar and is omitted. We express $vn^{1/2}\{\tilde{G}(t,v)-G_{Z,\beta^*}(t,v)\}$ as

$$\begin{aligned}
&n^{1/2}(\mathcal{P}_n-\mathcal{P})\{S(t;c_v,\beta^*,W)I(\beta^{*\mathrm{T}}Z\leqslant c_v)\}\\
&\quad+n^{1/2}\mathcal{P}[\{\hat{S}(t;c_v,\beta^*,W)-S(t;c_v,\beta^*,W)\}I(\beta^{*\mathrm{T}}Z\leqslant c_v)]\\
&\quad+n^{1/2}\mathcal{P}\{\hat{S}(t;\hat{c}_v,\hat{\beta},W)I(\hat{\beta}^{\mathrm{T}}Z\leqslant \hat{c}_v)-\hat{S}(t;c_v,\beta^*,W)I(\beta^{*\mathrm{T}}Z\leqslant c_v)\}\\
&\quad+n^{1/2}(\mathcal{P}_n-\mathcal{P})[\{\hat{S}(t;\hat{c}_v,\hat{\beta},W)-S(t;\hat{c}_v,\hat{\beta},W)\}I(\hat{\beta}^{\mathrm{T}}Z\leqslant \hat{c}_v)]\\
&\quad+n^{1/2}(\mathcal{P}_n-\mathcal{P})\{S(t;\hat{c}_v,\hat{\beta},W)I(\hat{\beta}^{\mathrm{T}}Z\leqslant \hat{c}_v)-S(t;c_v,\beta^*,W)I(\beta^{*\mathrm{T}}Z\leqslant c_v)\}. \qquad \text{(A6)}
\end{aligned}$$

To study the second to the fourth terms in (A6), we first study

$$\hat{\Lambda}(t;c,\beta,w)=\int_0^t\frac{\sum_{i=1}^n K\{(W_i-w)/h\}I(\beta^{\mathrm{T}}Z_i\leqslant c)\,\mathrm{d}N_i(s)}{\sum_{i=1}^n K\{(W_i-w)/h\}I(\beta^{\mathrm{T}}Z_i\leqslant c)Y_i(s)}.$$

Write $\tilde{T}=T\wedge C$ and $\Delta=I(T\leqslant C)$. Then $\hat{\Lambda}(t;c,\beta,w)-\Lambda(t;c,\beta,w)$ can be written as

$$\begin{aligned}
&\mathcal{P}_n\left[\frac{K(\frac{W-w}{h})I(\beta^{\mathrm{T}}Z\leqslant c)\Delta I(\tilde{T}\leqslant t)}{\mathcal{P}_n\{K(\frac{W-w}{h})I(\beta^{\mathrm{T}}Z\leqslant c)Y(u)\}|_{u=\tilde{T}}}\right]-\mathcal{P}\left[\left.\frac{I(\beta^{\mathrm{T}}Z\leqslant c)I(\tilde{T}\leqslant t)}{\mathcal{P}\{I(\beta^{\mathrm{T}}Z\leqslant c)Y^*(u)\mid W\}|_{u=\tilde{T}}}\right| W\right]\\
&=(\mathcal{P}_n-\mathcal{P})\left[\frac{K(\frac{W-w}{h})I(\beta^{\mathrm{T}}Z\leqslant c)\Delta I(\tilde{T}\leqslant t)}{\mathcal{P}_n\{K(\frac{W-w}{h})I(\beta^{\mathrm{T}}Z\leqslant c)Y(u)\}|_{u=\tilde{T}}}\right]\\
&\quad-\mathcal{P}\left[\frac{K(\frac{W-w}{h})I(\beta^{\mathrm{T}}Z\leqslant c)\Delta I(\tilde{T}\leqslant t)(\mathcal{P}_n-\mathcal{P})\{K(\frac{W-w}{h})I(\beta^{\mathrm{T}}Z\leqslant c)Y(u)\}|_{u=\tilde{T}}}{\mathcal{P}\{K(\frac{W-w}{h})I(\beta^{\mathrm{T}}Z\leqslant c)Y(u)\}\mathcal{P}_n\{K(\frac{W-w}{h})I(\beta^{\mathrm{T}}Z\leqslant c)Y(u)\}|_{u=\tilde{T}}}\right]\\
&\quad+\left(\mathcal{P}\left[\frac{K(\frac{W-w}{h})I(\beta^{\mathrm{T}}Z\leqslant c)\Delta I(\tilde{T}\leqslant t)}{\mathcal{P}\{K(\frac{W-w}{h})I(\beta^{\mathrm{T}}Z\leqslant c)Y(u)\}|_{u=\tilde{T}}}\right]-\mathcal{P}\left[\left.\frac{I(\beta^{\mathrm{T}}Z\leqslant c)I(\tilde{T}\leqslant t)}{\mathcal{P}\{I(\beta^{\mathrm{T}}Z\leqslant c)Y^*(u)\mid W\}|_{u=\tilde{T}}}\right| W\right]\right). \qquad \text{(A7)}
\end{aligned}$$

The third term on the right-hand side of (A7) is $O(h^l)$ by a simple transformation and Taylor expansion, together with the fact that $\int_0^t E\{\mathrm{d}N(u)I(\beta^{\mathrm{T}}Z\leqslant c)\mid W\}/E\{I(\beta^{\mathrm{T}}Z\leqslant c)Y(u)\mid W\}=\int_0^t E\{\mathrm{d}N^*(u)|Y^*(u)=1,\beta^{\mathrm{T}}Z\leqslant c,W\}$. Thus, by the Duhamel equation and the condition that $nh^{2l}=o(1)$, the second term in (A6) is asymptotically equivalent to $n^{1/2}(\mathcal{P}_n-\mathcal{P})\xi_n(N,Y,Z,W;t,c_v,\beta^*)$, where

$$\xi_n(N,Y,Z,W;t,c,\beta)=-I(\beta^{\mathrm{T}}Z\leqslant c)\left(\mathcal{P}_{Z,W}\left[I(\beta^{\mathrm{T}}Z\leqslant c)S(t;c,\beta,W)h^{-q}K\{(w-W)/h\}\right.\right.$$
$$\left.\left.\times M_n(N,Y;t,c,\beta,W)\right]\right)|_{w=W},$$

$$M_n(N,Y;t,c,\beta,w)=\int_0^t\frac{\mathrm{d}N(u)-Y(u)\,\mathrm{d}\Lambda(u;c,\beta,w)}{\mathcal{P}_n\,[h^{-q}K\{(W-w)/h\}I(\beta^{\mathrm{T}}Z\leqslant c)Y(u)]},$$

and $\mathcal{P}_{Z,W}$ denotes the expectation with respect to Z and W. By the condition that $nh^q \to \infty$, it can be shown that, uniformly in t, c and β, $\xi_n(t; c, \beta)$ converges in probability to

$$\xi(t; c, \beta) = -E\{I(\beta^{\mathrm{T}} Z \leqslant c)|W\} S(t; c, \beta, W) I(\beta^{\mathrm{T}} Z \leqslant c) \int_0^t \frac{\mathrm{d}N(u) - Y(u)\,\mathrm{d}\Lambda(u; c, \beta, W)}{E\{I(\beta^{\mathrm{T}} Z \leqslant c) Y(u)|W\}}.$$

This result, together with the fact that the class of functions $\xi_n(N, Y, Z, w; t, c, \beta)$ indexed by w, t, c and β is $\mathcal{P}$-Donsker, implies that, conditional on $W_1, \ldots, W_n$, the process $n^{1/2}(\mathcal{P}_n - \mathcal{P})\xi_n(t; c, \beta)$ is asymptotically equivalent to $n^{1/2}(\mathcal{P}_n - \mathcal{P})\xi(t; c, \beta)$ by Theorem 2.11.1 of van der Vaart & Wellner (1996), which further implies the unconditional asymptotic equivalence. Therefore, the second and third terms in (A6) are asymptotically equivalent to $n^{1/2}(\mathcal{P}_n - \mathcal{P})\xi(t; c_v, \beta^*)$ and

$$\begin{aligned} n^{1/2}&\mathcal{P}[S(t; \hat{c}_v, \hat{\beta}, W) I(\hat{\beta}^{\mathrm{T}} Z \leqslant \hat{c}_v) - S(t; c_v, \beta^*, W) I(\beta^{*\mathrm{T}} Z \leqslant c_v)] \\ &+ n^{1/2}(\mathcal{P}_n - \mathcal{P})\{\xi(t; \hat{c}_v, \hat{\beta}) - \xi(t; c_v, \beta^*)\}, \end{aligned} \tag{A8}$$

respectively. By the arguments of Appendix 2, (A8) is asymptotically equivalent to

$$v \times \left\{ \left.\frac{\partial G(t; c, \beta^*)}{\partial c}\right|_{c=c_v} n^{1/2}(\hat{c}_v - c_v) + \left.\frac{\partial G(t; c_v, \beta)}{\partial \beta}\right|_{\beta=\beta^*} n^{1/2}(\hat{\beta} - \beta^*) \right\}.$$

As in the case of $n^{1/2}\mathcal{P}[\{\hat{S}(t; \hat{c}_v, \hat{\beta}, W) - S(t; \hat{c}_v, \hat{\beta}, W)\} I(\hat{\beta}^{\mathrm{T}} Z \leqslant \hat{c}_v)]$, we can show that $n^{1/2}\mathcal{P}_n[\{\hat{S}(t; \hat{c}_v, \hat{\beta}, W) - S(t; \hat{c}_v, \hat{\beta}, W)\} I(\hat{\beta}^{\mathrm{T}} Z \leqslant \hat{c}_v)]$ is asymptotically equivalent to $n^{1/2}(\mathcal{P}_n - \mathcal{P})\xi(t; \hat{c}_v, \hat{\beta})$, which implies that the fourth term in (A6) converges uniformly to zero in probability. The fifth term in (A6) converges uniformly to zero in probability by Lemma 19.24 of van der Vaart (1998).

Combining the above results, we obtain the weak convergence of $n^{1/2}\{\tilde{G}(t, v) - G_{Z,\beta^*}(t, v)\}$.

Appendix 4

Validity of the bootstrap method

In this section, we prove the validity of the bootstrap method for $\hat{G}(t, v)$ for survival data. The proofs for other cases are similar and thus omitted. Define

$$\hat{G}^{(b)}(t; c, \beta) = \prod_{0<u\leqslant t} \left\{ 1 - \frac{\sum_{i=1}^n M_{ni} I(\beta^{\mathrm{T}} Z_i \leqslant c)\,\mathrm{d}N_i(u)}{\sum_{i=1}^n M_{ni} I(\beta^{\mathrm{T}} Z_i \leqslant c) Y_i(u)} \right\},$$

where M_{ni} is the number of times that $\{N_i(t), Y_i(t), X_i : t \in [0, \tau]\}$ is redrawn from the original sample. Let $\hat{c}_v^{(b)}$ and $\hat{\beta}^{(b)}$ be the bootstrap counterparts of $\hat{c}_v$ and $\hat{\beta}$. Our goal is to show that the distribution of $n^{1/2}\{\hat{G}(t, v) - G_{Z,\beta^*}(t, v)\}$ can be approximated by the conditional distribution of $n^{1/2}\{\hat{G}^{(b)}(t; \hat{c}_v^{(b)}, \hat{\beta}^{(b)}) - \hat{G}(t, v)\}$ given the data.

By the arguments of Appendix 2, $n^{1/2}\{\hat{G}^{(b)}(t; \hat{c}_v^{(b)}, \hat{\beta}^{(b)}) - \hat{G}(t, v)\}$ is equal to

$$n^{1/2}\{\hat{G}^{(b)}(t; \hat{c}_v, \hat{\beta}) - \hat{G}(t; \hat{c}_v, \hat{\beta})\} + n^{1/2}\{\hat{G}^{(b)}(t; \hat{c}_v^{(b)}, \hat{\beta}^{(b)} - \hat{G}^{(b)}(t; \hat{c}_v, \hat{\beta})\}. \tag{A9}$$

Lemma S1 of the Supplementary Material implies that, given the data, the first term in (A9) has the same asymptotic distribution as $n^{1/2}(\mathcal{P}_n - \mathcal{P})\xi(t; c_v, \beta^*)$, where $\xi(t; c, \beta)$ is defined in Appendix 2. It can also be shown that, given the data, $n^{1/2}(\hat{\beta}^{(b)} - \hat{\beta}) = n^{1/2}(\hat{\mathcal{P}}_n - \mathcal{P}_n)\psi + o_p(1)$ under mild regularity conditions, where $\hat{\mathcal{P}}_n$ is the bootstrap empirical distribution (van der Vaart & Wellner, 1996, p. 345). Combining the arguments of Appendix 2 with those of Lemma S1, we can show that the limiting distribution of $n^{1/2}\{\hat{G}^{(b)}(t; \hat{c}_v^{(b)}, \hat{\beta}^{(b)}) - \hat{G}(t, v)\}$ given the data is the same as the limiting distribution of $n^{1/2}(\mathcal{P}_n - \mathcal{P})\{\xi_1(t, v) + \xi_2(t, v)\}$, where $\xi_1(t, v)$ and $\xi_2(t, v)$ are defined in Appendix 2.

REFERENCES

DABROWSKA, D. M. & DOKSUM, K. A. (1988). Partial likelihood in transformation models with censored data. *Scand. J. Statist.* **15**, 1–23.

FAN, C., OH, D. S., WESSELS, L., WEIGELT, B., NUYTEN, D. S. A., NOBEL, A. B., VAN'T VEER, L. J. & PEROU, C. M. (2006). Concordance among gene-expression-based predictors for breast cancer. *New Engl. J. Med.* **355**, 560–9.

GRAF, E., SCHMOOR, C., SAUERBREI, W. & SCHUMACHER, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statist. Med.* **18**, 2529–45.

HEAGERTY, P. J. & ZHENG, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics* **61**, 92–105.

HENDERSON, R., DIGGLE, P. & DOBSON, A. (2002). Identification and efficacy of longitudinal markers for survival. *Biostatistics* **3**, 33–50.

KORN, E. L. & SIMON, R. (1990). Measures of explained variation for survival data. *Statist. Med.* **9**, 487–503.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.

LIN, D. Y., WEI, L. J. & YING, Z. (2001). Semiparametric transformation models for point processes. *J. Am. Statist. Assoc.* **96**, 620–8.

MOSKOWITZ, C. S. & PEPE, M. S. (2004). Quantifying and comparing the predictive accuracy of continuous prognostic factors for binary outcomes. *Biostatistics* **5**, 113–27.

SCHEMPER, M. (1990). The explained variation in proportional hazards regression. *Biometrika* **77**, 216–8. (Correction: (1994). *Biometrika* **81**, 631.)

SCHEMPER, M. & HENDERSON, R. (2000). Predictive accuracy and explained variation in Cox regression. *Biometrics* **56**, 249–55.

SCHOOP, R., GRAF, E. & SCHUMACHER, M. (2008). Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. *Biometrics* **64**, 603–10.

THERNEAU, T. M. & HAMILTON, S. A. (1997). rhDNase as an example of recurrent event analysis. *Statist. Med.* **16**, 2029–47.

UNO, H., CAI, T., TIAN, L. & WEI, L. J. (2007). Evaluating prediction rules for t-year survivors with censored regression models. *J. Am. Statist. Assoc.* **102**, 527–37.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. New York: Cambridge University Press.

VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

VAN DE VIJVER, M. J., HE, Y. D., VAN'T VEER, L. J., DAI, H., HART, A. A. M., VOSKUIL, D. W., SCHREIBER, G. J., PETERSE, J. L., ROBERTS, C., MARTON, M. J. ET AL. (2002). A gene-expression signature as a predictor of survival in breast cancer. *New Engl. J. Med.* **347**, 1999–2009.

ZENG, D. & LIN, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93**, 627–40.

ZHENG, Y., CAI, T., PEPE, M. S. & LEVY, W. C. (2008). Time-dependent predictive values of prognostic biomarkers with failure time outcome. *J. Am. Statist. Assoc.* **103**, 362–8.

ZHENG, Y. & HEAGERTY, P. J. (2007). Prospective accuracy for longitudinal markers. *Biometrics* **63**, 332–41.

[*Received September* 2010. *Revised January* 2012]