COMMUNICATIONS IN STATISTICS
Theory and Methods
Vol. 33, No. 6, pp. 1385–1401, 2004

Estimation of the Asymptotic Variance of Semiparametric Maximum Likelihood Estimators in the Cox Model with a Missing Time-Dependent Covariate

J.-F. Dupuy[1,*] and M. Mesbah[2]

[1]CEREMADE, Université Paris IX Dauphine, Paris, France
[2]Université de Bretagne Sud, Vannes, France

ABSTRACT

The relationship between a time-to-event and a time-dependent or time-independent covariate is usually assessed using the Cox model. A frequently encountered problem however is occurrence of missing covariate values. A recent approach for estimating the Cox model with a missing covariate jointly models the time-to-event and covariate. In the case of a time-dependent covariate, Dupuy and Mesbah [Dupuy, J.-F., Mesbah, M. (2002). Joint modeling of event time and nonignorable missing longitudinal data. *Lifetime D. Anal.* 8:99–115]

*Correspondence: J.-F. Dupuy, Laboratoire de Statistique et Probabilités – UMR C5583, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex 4, France; Fax: +33-05-61-55-60-89; E-mail: dupuy@math.ups-tlse.fr.

DOI: 10.1081/STA-120030156 0361-0926 (Print); 1532-415X (Online)

have proposed a joint model and have obtained a semiparametric maximum likelihood estimator of the regression parameter of the Cox model that is consistent and asymptotically normal. Furthermore, an explicit expression was obtained for the asymptotic variance of this estimator. In this paper, we examine the problem of estimating this variance. We propose a computationally simple estimator and we show its consistency. We illustrate the approach by applications to real data sets.

Key Words: Time-dependent Cox model; Missing covariate; Semiparametric maximum likelihood estimation; Asymptotic covariance; Consistency.

Mathematics Subject Classification: 62N01 (62N02, 62P10).

1. INTRODUCTION

Many survival studies collect information on each study participant both on a survival time and a time-dependent covariate. A frequent objective of these studies is to evaluate the relationship between survival and the covariate. The Cox regression model (Cox, 1972) is one of the most widely used tool for that purpose. To implement the Cox model with a time-dependent covariate, complete knowledge of the covariate history for each subject is required. A frequently encountered problem in practice, however, is occurrence of missing covariate data. Dupuy and Mesbah (2002) have proposed a joint modeling approach for estimating the Cox model with a nonignorable missing time-dependent covariate. Joint analysis of repeated measurements and survival times has also been extensively used in the context of measurement error of the covariate. References include Henderson et al. (2000), Li and Lin (2000), Song et al. (2002a,b), Tsiatis and Davidian (2001) and Wulfsohn and Tsiatis (1997). Bayesian approaches to estimation in these joint models are studied by Faucett and Thomas (1995), Ibrahim et al. (2001).

Asymptotic properties of estimators obtained from joint models are still an open problem. We note nevertheless that some promising simulations have been performed by Li and Lin (2000), and that Dupuy et al. (2002, 2003) have established consistency and asymptotic normality of estimators obtained in Dupuy and Mesbah (2002) via a joint model. Since estimation of the asymptotic variance of these estimators is of practical interest for constructing tests of hypothesis, we propose in this paper a simple estimator of this asymptotic variance, and we show its consistency.

The paper is organized as follows. The joint model proposed by Dupuy and Mesbah (2002) is briefly presented in Sec. 2. In Sec. 3, a result of asymptotic normality of semiparametric maximum likelihood estimators in this joint model is briefly recalled. We then discuss practical issues arising in the estimation of the asymptotic variance of these estimators. An estimator of this variance is proposed and its consistency is proved. In Sec. 4, we illustrate our results on two real data sets.

2. A JOINT MODEL

Dupuy and Mesbah (2002) propose a joint modeling approach for estimating the parameters in the Cox model with missing values of a time-dependent covariate. The problem and the statistical model can be described as follows.

Consider a sample of n subjects, indexed by i. Observations of a time-dependent covariate are to be measured at fixed times $t_0 = 0 < t_1 < \cdots$ on each of these subjects. Each subject i leaves the study at a random failure time T_i which is subject to right censoring, that is we observe $X_i = \min(T_i, C_i)$, where C_i is a potential censoring time, and the censoring indicator $\Delta_i = 1_{\{T_i \leq C_i\}}$. Assume that the study ends at a fixed time $\tau(\tau < \infty)$ at which any individual still alive is censored.

The covariate process $Z_i(\cdot)\,(i = 1, \ldots, n)$ is assumed to have the form of a step function $Z_i(t) = Z_{ij}$, $t \in (t_{j-1}, t_j]$ [we refer to Bagdonavičius and Nikulin (2001) for analysis of time-to-event models under this kind of covariate process]. Z_{ij} is measured at t_j if failure or censoring do not occur in the interval $[0, t_j]$. Hence Z_{ij} is not observed if failure occur before or at t_j. We assume that Z_{ij} have uniformly bounded total variations, that is $\int_0^\infty |dZ_{ij}(t)| + |Z_{ij}(0)| \leq c$ for some $c > 0$ (this ensures that Z has not too much fluctuation, and this is a reasonable hypothesis in many applications).

The model for the longitudinal covariate is: $Z_{ij} = \alpha Z_{i,j-1} + \epsilon_{ij}$, $j = 1, 2, \ldots,$ where $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ are independent deviations reflecting measurement error. Let us denote $(\alpha, \sigma_e^2)'$ by ξ. The hazard function for the Cox model is given by $\lambda(t)e^{\beta Z_i(t)}$, where $\lambda(\cdot)$ is an unknown baseline hazard function and β is an unknown regression parameter. For ease of presentation, we assume that the current value of the covariate is the appropriate component of the covariate history to use in this model. However, the arguments below would be no more difficult, had the hazard been a more complex functional of the history of the covariate.

Recall that from the experimental design, the realization of $Z_i(X_i)(i = 1, \ldots, n)$ at failure or censoring is not observed. Let $a_{X_i} = \max\{k : t_k < X_i\}$ denote the indice of the last observed value of

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

$Z_i(\cdot)$ before X_i. With these notations, the observed data on subject i are $(X_i, \Delta_i, Z_{i0}, \ldots, Z_{ia_{X_i}})$, independent across i.

Assume that the censoring time C_i is independent of the failure time T_i given the covariate, and that the distribution of C_i does not depend on the parameters $\xi, \beta, \Lambda(\cdot)$, nor on the unobserved value of $Z_i(\cdot)$ at X_i. Under our assumptions, the observed data likelihood L_n is

$$\prod_{i=1}^{n} \left\{ \int \left\{ \lambda(X_i) e^{\beta Z_i(X_i)} \right\}^{\Delta_i} \cdot \exp\left[- \int_0^{X_i} \lambda(u) e^{\beta Z_i(u)} \, du \right] \times f_\xi(Z_{i0}, \ldots, Z_{ia_{X_i}}, Z_i(X_i)) dZ_i(X_i) \right\},$$

where $f_\xi(Z_{i0}, \ldots, Z_{il})$ is the density of the random vector $(Z_{i0}, \ldots, Z_{il})(l \geq 1)$.

Let $\theta_0 = (\Lambda_0, \beta_0, \xi_0)$ denote the true value of the parameter. The problem of finding an estimator for θ_0 is semiparametric since the component Λ_0 belongs to a set of nondecreasing absolutely continuous functions. The maximum in Λ of the likelihood function does not exist. To define an estimator for Λ_0, we proceeds by the method of sieves, which consists in replacing the space Θ of parameters (Λ, β, ξ) by an increasing sequence of approximating spaces Θ_n, so that there exists a maximum likelihood estimator in each Θ_n. For a fixed sample size n, the above likelihood is then maximized over Θ_n. Following the approach of Johansen (1983), we maximize the likelihood L_n over the space $\{\theta = (\Lambda_n, \beta, \xi)\}$, where $\beta \in \mathbb{R}$, $\xi \in \mathbb{R}^2$ and Λ_n is an increasing step function with positive jumps at each uncensored failure time. This leads to so-called semiparametric maximum likelihood (SPML) estimators. We refer to Li and Lin (2000), Murphy (1995), Parner (1998), Scharfstein et al. (1998) and Wulfsohn and Tsiatis (1997) for use of SPML in various situations. Another example of use of a sieve procedure is given by Murphy and Sen (1991) for estimating time-dependent regression coefficients in the Cox model.

We used the expectation-maximization (EM) algorithm to calculate the SPML estimates, along with Gauss–Hermite quadrature for evaluation of integrals. Formulas are given in Dupuy and Mesbah (2002).

3. ESTIMATION OF THE ASYMPTOTIC VARIANCE OF SPML ESTIMATORS IN THE JOINT MODEL

3.1. Preliminaries

In the following, let $BV[0, \tau]$ denote the space of bounded variation real-valued functions defined on $[0, \tau]$. Let $H = \{(h_1, h_2, h_3) : h_1 \in BV[0, \tau],$

$h_2 \in \mathbb{R}, h_3 \in \mathbb{R}^2\}$. With $h \in H$, we define the norm on H to be $\|h\|_H = \|h_1\|_\nu + |h_2| + \|h_3\|_2$, where $\|\cdot\|_2$ denotes the Euclidean norm and $\|h_1\|_\nu$ is the absolute value of $h_1(0)$ plus the total variation of h_1 on the interval $[0, \tau]$. We will consider the parameter θ as a functional on H given by

$$\theta(h) = \int_0^\tau h_1(u) d\Lambda(u) + h_2\beta + h_3'\xi, \tag{1}$$

and the parameter space Θ as a subset of $l^\infty(H)$, the space of bounded real-valued functions on H, equipped with the supremum norm $\|\theta\| = \sup_{h\in H}|\theta(h)|$.

Note that specific choices for h in (1) correspond to the quantities of interest. For example, if we let $h_1(u) = 0$ for all $u \in [0, \tau], h_2 = 1$ and $h_3 = 0$, then (1) reduces to β. If we let $h_1(u) = 1_{\{u\le t\}}(t \in [0, \tau])$, $h_2 = 0$ and $h_3 = 0$, then (1) reduces to $\Lambda(t)$.

In this paper, we are interested in estimating the asymptotic variance of the SPML estimator of β_0 and Λ_0, so we restrict attention to the limiting distribution of

$$\left(\sqrt{n}(\hat{\Lambda}_n - \Lambda_0), \sqrt{n}(\hat{\beta}_n - \beta_0)\right),$$

which is given by Dupuy et al. (2003). Dupuy et al. (Submitted) gives the more general result of the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0) = (\sqrt{n}(\hat{\Lambda}_n - \Lambda_0), \sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\xi}_n - \xi_0))$. This may be relevant in applications where one is also interested in estimating and testing the parameters of the distribution of the time-dependent covariate.

Dupuy et al. (2003) give in the following theorem an asymptotic normality result for $(\sqrt{n}(\hat{\Lambda}_n - \Lambda_0), \sqrt{n}(\hat{\beta}_n - \beta_0))$, considered as a functional on the space $H_{12} = \{h \in H : h_3 = 0, \|h\|_H < \infty\}$. Again, if we let $h_1(u) = 0$ for all $u \in [0, \tau]$ and $h_2 = 1$, this reduces to $\sqrt{n}(\hat{\beta}_n - \beta_0)$. Similarly, letting $h_1(u) = 1_{\{u\le t\}}(t \in [0, \tau])$ and $h_2 = 0$ yields to $\sqrt{n}(\hat{\Lambda}_n(t) - \Lambda_0(t))$.

In the following, we denote expectation with respect to the empirical distribution of the data by $\mathbb{E}_n$, and expectation with respect to the true underlying distribution by E_{θ_0}. Moreover, we denote by $E_{i\theta}[h(Z(X))]$ the conditional expectation of a function $h(Z(X))$ given the observations on the i-th subject and the parameter value θ.

Theorem 1. *The sequence* $(\sqrt{n}(\hat{\Lambda}_n - \Lambda_0), \sqrt{n}(\hat{\beta}_n - \beta_0))$ *converges in distribution in the space* $l^\infty(H_{12})$ *to a zero mean Gaussian process G with covariance process*

$$cov(G(g), G(g^*)) = \int_0^\tau g_1(u)\sigma_{1,\theta_0}^{-1}(g^*)(u) d\Lambda_0(u) + \sigma_{2,\theta_0}^{-1}(g^*)g_2,$$

where

$$\sigma_{1,\theta_0}(h)(u) = E_{\theta_0}\left[(h_1(u) + h_2 Z(u))e^{\beta_0 Z(u)} 1_{\{u \le X\}}\right],$$
$$\sigma_{2,\theta_0}(h) = E_{\theta_0}\left[\int_0^X Z(u)e^{\beta_0 Z(u)}(h_1(u) + h_2 Z(u))d\Lambda_0(u)\right].$$

Estimation of the asymptotic variance

$$\int_0^\tau h_1(u)\sigma_{1,\theta_0}^{-1}(h)(u)d\Lambda_0(u) + \sigma_{2,\theta_0}^{-1}(h)h_2, \tag{2}$$

of $(\sqrt{n}(\hat{\Lambda}_n - \Lambda_0), \sqrt{n}(\hat{\beta}_n - \beta_0))(h)$ is crucial for constructing tests of hypotheses and calculating confidence intervals. In the following, we discuss practical and theoretical aspects of the problem of estimation of (2).

3.2. A Convergent Estimator of the Asymptotic Variance

A natural estimator of (2) may be obtained by replacing θ_0 by $\hat{\theta}_n$ in $\sigma_{\theta_0} = (\sigma_{1,\theta_0}, \sigma_{2,\theta_0})$. This, however, raises problems. We first illustrate these problems on the particular example of estimation of the asymptotic variance of $\sqrt{n}(\hat{\beta}_n - \beta_0)$. We then propose an alternative convergent estimator of (2).

Estimating the asymptotic variance $\sigma_{2,\theta_0}^{-1}(h)$ of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ by replacing θ_0 by $\hat{\theta}_n$ in σ_{2,θ_0} leads to the following estimator:

$$\sigma_{2,\hat{\theta}_n}^{-1}(h) = E_{\hat{\theta}_n}\left[\sum_{k=1}^{p} Z(s_k)^2 e^{\hat{\beta}_n Z(s_k)} \Delta\hat{\Lambda}_n(s_k) 1_{\{s_k \le X\}}\right]^{-1}. \tag{3}$$

Calculation of such expectations are however cumbersome, due to the dimension of the integrals involved. One may then approximate (3) by using empirical means, such as

$$\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{p} Z_i(s_k)^2 e^{\hat{\beta}_n Z_i(s_k)} \Delta\hat{\Lambda}_n(s_k) 1_{\{s_k \le X_i\}}\right]^{-1}.$$

This, however, would require observation of the missing $Z_i(X_i)$.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Our proposal is then to first estimate $\sigma_{\theta_0} = (\sigma_{1,\theta_0}, \sigma_{2,\theta_0})$ by $\hat{\sigma}_{\hat{\theta}_n} = (\hat{\sigma}_{1,\hat{\theta}_n}, \hat{\sigma}_{2,\hat{\theta}_n})$, where

$$\hat{\sigma}_{1,\hat{\theta}_n}(h)(u) = \frac{1}{n}\sum_{i=1}^{n} E_{i\hat{\theta}_n}\left[(h_1(u) + h_2 Z(u))e^{\hat{\beta}_n Z(u)} 1_{\{u \le X\}}\right],$$

$$\hat{\sigma}_{2,\hat{\theta}_n}(h) = \frac{1}{n}\sum_{i=1}^{n} E_{i\hat{\theta}_n}\left[\int_0^X Z(u)e^{\hat{\beta}_n Z(u)}(h_1(u) + h_2 Z(u))d\hat{\Lambda}_n(u)\right],$$

and then to estimate the asymptotic variance (2) by

$$\int_0^\tau h_1(u)\hat{\sigma}^{-1}_{1,\hat{\theta}_n}(h)(u)d\hat{\Lambda}_n(u) + \hat{\sigma}^{-1}_{2,\hat{\theta}_n}(h)h_2. \tag{4}$$

For example, following this proposition, the estimator of the asymptotic variance of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ has the form

$$\hat{\sigma}^{-1}_{2,\hat{\theta}_n}(h) = \left[\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{p} E_{i\hat{\theta}_n}[Z(s_k)^2 e^{\hat{\beta}_n Z(s_k)}]\Delta\hat{\Lambda}_n(s_k)1_{\{s_k \le X_i\}}\right]^{-1}. \tag{5}$$

Next theorem shows that (4) is a consistent estimator of the asymptotic variance (2).

Theorem 2. *For $(g_1, g_2) \in H_{12}$, the solution $h = \hat{\sigma}^{-1}_{\hat{\theta}_n}(g)$ to $g_1 = \hat{\sigma}_{1,\hat{\theta}_n}(h)$, $g_2 = \hat{\sigma}_{2,\hat{\theta}_n}(h)$ exists with probability going to one as n tends to infinity, and*

$$\int_0^\tau h_1(u)\hat{\sigma}^{-1}_{1,\hat{\theta}_n}(h)(u)\,d\hat{\Lambda}_n(u) + \hat{\sigma}^{-1}_{2,\hat{\theta}_n}(h)h_2$$

converges in probability to $\int_0^\tau h_1(u)\sigma^{-1}_{1,\theta_0}(h)(u)d\Lambda_0(u) + \sigma^{-1}_{2,\theta_0}(h)h_2$.

We shall note, however, that we can not compute the exact value of $\hat{\sigma}^{-1}_{1,\hat{\theta}_n}$ and $\hat{\sigma}^{-1}_{2,\hat{\theta}_n}$ since the EM algorithm provides us only an approximation of $\hat{\theta}_n$. Moreover, θ contains a functional part and convergence properties of the EM algorithm in this case are not well established. We shall note that simulations by Chen and Little (1999) and Li and Lin (2000) in the contexts of missing and mismeasured time-independent covariates respectively seem to point to the validity of such an approach.

Proof. We give an outline of our proof, which follows the method proposed by Murphy (1995) for proving consistency of an estimator of

the asymptotic variance of estimators in the frailty model. This theorem can be proved by the following two steps. □

Lemma 1. $\sup_{h\in H_{12}}\|\hat{\sigma}_{\hat{\theta}_n}(h)-\sigma_{\theta_0}(h)\|_H$ *converges in probability to* 0.

Proof. We first show that the functions under the sign $\sum$ in $\hat{\sigma}_{1,\hat{\theta}_n}$ and $\hat{\sigma}_{2,\hat{\theta}_n}$ form Donsker classes (for $h \in H_{12}$). As an illustration of the method, we prove that

$$\left\{\int_0^\tau h_1(u)Z(u)e^{\beta Z(u)}1_{\{u\le X\}}\,d\Lambda(u) : h_1 \in BV[0,\tau]\right\}$$

form a Donsker class.

From Lemma 2 of Parner (1998), the classes $f_{1,u}: 1_{\{[\cdot,\infty)\}} \longmapsto 1_{\{[u,\infty)\}}$ and $f_{2,u}: Z(\cdot)\longmapsto Z(u)\,(u\in[0,\tau])$ are Donsker. Then $f_{3,u}: e^{\beta Z(\cdot)}\longmapsto e^{\beta Z(u)}$ is Donsker since the exponential function is Lipschitz on compact sets of real line. Finally, $f_{4,u}: Z(\cdot)e^{\beta Z(\cdot)}1_{\{[\cdot,\infty)\}}\longmapsto Z(u)e^{\beta Z(u)}1_{\{[u,\infty)\}}(u\in[0,\tau])$ is Donsker since $Z(\cdot)$ is uniformly bounded.

Let $\psi: u\longmapsto\psi(u)$ be some uniformly bounded real-valued function defined on $[0,\tau]$. Define the map $\varphi:\psi\longmapsto\varphi(\psi)$ from $l^\infty([0,\tau])$ to $l^\infty(BV[0,\tau])$ by

$$\varphi(\psi)\,(h)=\int_0^\tau h(u)\psi(u)df(u),$$

for $f\in BV[0,\tau]$. Assume that the class $j_u:\psi(\cdot)\longmapsto\psi(u)(u\in[0,\tau])$ form a Donsker class. Then $\{\mathbb{G}_n j_u : u\in[0,\tau]\}$ (where $\mathbb{G}_n j_u=\sqrt{n}(\mathbb{E}_n j_u - E_{\theta_0}j_u)$) converges to some tight limit process. The function φ is continuous, and from the continuous mapping theorem, it follows that $\varphi(\mathbb{G}_n j_u)$ converges in distribution. Now,

$$\begin{aligned}\varphi(\mathbb{G}_n j_u)\,(h)&=\int_0^\tau h(u)\mathbb{G}_n j_u\,df(u)=\int_0^\tau h(u)\sqrt{n}(\mathbb{E}_n j_u-E_{\theta_0}j_u)df(u)\\&=\sqrt{n}\left(\mathbb{E}_n\int_0^\tau h(u)\psi(u)df(u)-E_{\theta_0}\int_0^\tau h(u)\psi(u)df(u)\right)\\&=\sqrt{n}(\mathbb{G}_n(\varphi(\psi)\,(h))).\end{aligned}$$

Hence $\psi(\cdot)\longmapsto\int_0^\tau h(u)\psi(u)\,df(u)(h\in BV[0,\tau])$ is Donsker. Applying this with $\psi(\cdot)=Z(\cdot)e^{\beta Z(\cdot)}1_{\{[\cdot,\infty)\}}$ gives that the class $Z(\cdot)e^{\beta Z(\cdot)}1_{\{[\cdot,\infty)\}}\longmapsto\int_0^\tau h_1(u)Z(u)e^{\beta Z(u)}1_{\{u\le X\}}\,d\Lambda(u)$ is Donsker $(h_1\in BV[0,\tau])$.

The same arguments for the remaining parts of $\hat{\sigma}_{1,\hat{\theta}_n}$ and $\hat{\sigma}_{2,\hat{\theta}_n}$ leads to the desired conclusion. Details can be found in Dupuy et al. (Submitted).

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

A Donsker class is also a Glivenko–Cantelli class. Hence, having proved that functions under $\sum$ in $\hat{\sigma}_{1,\hat{\theta}_n}$ and $\hat{\sigma}_{2,\hat{\theta}_n}$ are Donsker, it follows that $\sup_{h\in H_{12}}\|\hat{\sigma}_{\hat{\theta}_n}(h)-\sigma_{\theta_0}(h)\|_H$ converges in probability to 0. □

Lemma 2. $\sup_{h\in H_{12}}\|\hat{\sigma}^{-1}_{\hat{\theta}_n}(h)-\sigma^{-1}_{\theta_0}(h)\|_H$ *converges in probability to* 0.

Proof. From Lemma 4 of Dupuy et al. (Submitted), σ_{θ_0} is continuously invertible, then $\hat{\sigma}_{\hat{\theta}_n}$ is continuously invertible with a probability going to 1 as n tends to infinity. For $0<q<\infty$, let $H^q_{12}=\{h\in H_{12}:\|h\|_H\le q\}$. Then, with a probability going to 1 as n tends to infinity, for every $H^p_{12}\subset H_{12}$ there exists $H^q_{12}\subset H_{12}$ such that $\hat{\sigma}^{-1}_{\hat{\theta}_n}(H^q_{12})\subset H^p_{12}$, and for every $g\in H^q_{12}$, there exists $h\in H^p_{12}$ such that $\hat{\sigma}^{-1}_{\hat{\theta}_n}(g)=h$. Therefore

$$\begin{aligned}\|\hat{\sigma}^{-1}_{\hat{\theta}_n}(g)-\sigma^{-1}_{\theta_0}(g)\|_H &= \|\sigma^{-1}_{\theta_0}(\sigma_{\theta_0}(h))-\sigma^{-1}_{\theta_0}(\hat{\sigma}_{\hat{\theta}_n}(h))\|_H\\ &\le \sup_{h\in H^q_{12}}\frac{\|\sigma^{-1}_{\theta_0}(h)\|_H}{\|h\|_H}\sup_{h\in H^p_{12}}\|\sigma_{\theta_0}(h)-\hat{\sigma}_{\hat{\theta}_n}(h)\|_H.\end{aligned}$$

From Lemma 1, it follows that $\sup_{g\in H_{12}}\|\hat{\sigma}^{-1}_{\hat{\theta}_n}(g)-\sigma^{-1}_{\theta_0}(g)\|_H$ converges to 0. □

From Lemma 2, it follows that $\int_0^\tau h_1(u)\hat{\sigma}^{-1}_{1,\hat{\theta}_n}(h)\,(u)d\hat{\Lambda}_n(u)+\hat{\sigma}^{-1}_{2,\hat{\theta}_n}(h)h_2$ converges in probability to $\int_0^\tau h_1(u)\sigma^{-1}_{1,\theta_0}(h)\,(u)\,d\Lambda_0(u)+\sigma^{-1}_{2,\theta_0}(h)h_2$.

4. APPLICATIONS: TWO REAL EXAMPLES

Example 1. We illustrate the proposed method by first analyzing quality-of-life (QoL) data from a cancer clinical trial. Two groups of patients with metastatic colorectal cancer (referred to as groups A and B in the following) and treated with two different therapies were followed for their QoL until dropout. Collection of QoL data was done by administering a questionnaire to study participants at a number of prespecified time points. The aim of our analysis is to evaluate the relationship between dropout and the longitudinal QoL in each group, and to test for a difference between groups.

The Cox model is a standard tool for that purpose. However, Cox regression analysis is complicated here by missingness of QoL scores at dropout. A commonly used approach to implement the Cox model in such a situation substitutes in the Cox partial likelihood, for each subject at his failure time, the last covariate value prior to that time. This method, termed "last value carried forward" (LCF)

(Altman and De Stavola, 1994) can induce considerable bias into the parameter and variance estimators. One other strategy is a two-stage or "regression calibration" approach (Wang et al., 1997), where a model is fitted to the time-dependent covariate and predictors of the missing values are used to impute missing values in the partial likelihood. This method however produces an inconsistent estimator of the regression parameter of the Cox model (Wang et al., 1997). Some alternative approaches have recently been proposed: the approximate partial likelihood method of Lin and Ying (1993) consists in replacing any missing term at time t in the Cox partial likelihood score function by an estimate, calculated on subjects who are at risk at t and for whom observed covariate measurements at t are available. The estimated partial likelihood of Zhou and Pepe (1995) is an imputation method, which requires a validation sample of subjects with no missing covariate measurements. Paik and Tsai (1997) elaborate on this approach and propose an imputation method that does not require that some subjects are fully observed. However, for these methods to be valid, it is assumed that the covariate is missing completely at random or at random, that is, missingness does not depend on the unobserved covariate value. However, it has now been recognized that QoL is generally not missing at random (see Mesbah et al., 2002). In particular, following Mesbah et al. (Submitted) who provide a detailed description of the study and data, "it may be assumed that the patients with missing data had a worse QoL score." The authors note that "the relationship between data missingness and the possibly low values of the missing QoL scores suggested that the missingness mechanism was not at random."

Hence we propose to use the joint model proposed by Dupuy and Mesbah (2002) to take account of the missing QoL scores in estimating the Cox model. We then use our asymptotic results, and particularly the convergence in probability of our variance estimator, to test for association between dropout and the longitudinal QoL in each group, and to test for a difference between groups A and B. Results are compared with an at-random-analysis which assumes the following hazard function for dropout: $\lambda(t)\exp(\beta Z(t))$, where $Z(t) = Z_{a_t}$ is the last observed QoL score prior to dropout. Partial likelihood (PL) estimation is then used to estimate β.

Groups A and B had, respectively, 115 and 120 patients. The number of QoL scores per patient in group A (respectively, group B) ranges between 1 and 10 (respectively, 1 and 13). Scores of QoL range between 0 and 7, with a higher score indicating a better overall QoL.

Table 1 displays the estimates from both at-random-analysis and the joint model, together with estimated standard errors, which are obtained using formula (5) for the joint model.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Table 1. PL and SPML estimates and standard errors for at random and nonignorable analysis of time to dropout and longitudinal QoL.

	Group	Estimate of β	Estimated standard error (S.E.)	Est./S.E.	p-value
PL for at-random-analysis	A	−0.1636	0.0781	−2.0947	0.037
	B	−0.1672	0.0779	−2.1463	0.032
SPML for joint model	A	−0.3620	0.0863	−4.1947	<0.001
	B	−0.3161	0.0870	−3.6333	<0.001

The negative values for the regression parameter β obtained from both estimation methods imply that patients with low levels of QoL are more likely to dropout. However, we note that estimates from at-random-analysis are less significant (p-values are, respectively, 0.037 and 0.032 in groups A and B). Failure of the at-random-analysis to take account of the worsening of QoL just before dropout may explain that estimates of the association of dropout and QoL are biased towards the null. The at-random-analysis uses for estimation the last QoL score prior to dropout instead of the actual (unobserved) QoL value at dropout, which is certainly lower. This in turn attenuates the impact of a decrease in QoL on dropout and this may explain the observed bias.

We note that the estimated standard error obtained from the joint model is greater compared with the at-random-approach. This is because the estimate of β may be affected by the uncertainty in the estimate of the other parameters. More variability is incorporated.

It is also of practical interest to note that the estimated relative risks between 2 patients in group A with a difference in QoL of x units are, respectively, $e^{-0.1636x}$ (PL method for at-random-analysis) and $e^{-0.3620x}$ (SPML method for the joint model). These are represented as functions of x on Fig. 1, together with similar curves for group B.

It appears that when dropout is nonignorable, the at-random-analysis strongly underestimates the risk of dropout for negative x (in both groups A and B), that is, strongly underestimates risk of dropout for patients with a decreasing QoL. Our method proves useful to reduce bias in estimation of this relative risk, which may be of practical interest for practitioners and clinical care of the patients.

One may also be interested in testing for a difference in β between groups A and B. From the result proved in Sec. 3, $\sqrt{n}(\hat{\beta}_n - \beta_0)/\sqrt{\hat{\sigma}^{-1}_{2,\hat{\theta}_n}(h_\beta)}$ converges in distribution to $\mathcal{N}(0,1)$, where $h_\beta = (h_1, 1)$ with

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

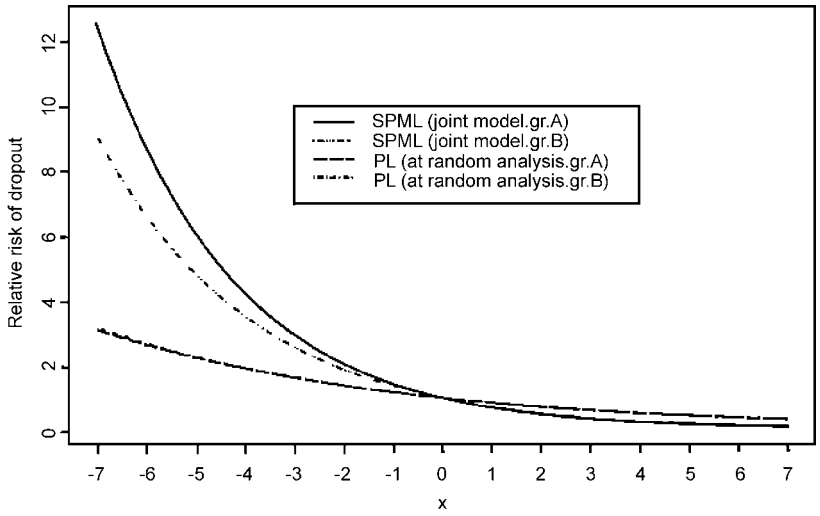

Figure 1. Relative risk of dropout vs. x estimated by the SPML and PL methods.

$h_1(u) = 0$, $u \in [0, \tau]$. Let denote this statistic by

$$\sqrt{n_A}\frac{\hat{\beta}^A_{n_A} - \beta^A_0}{\sqrt{\hat{\sigma}^{-1}_{2,\hat{\theta}^A_{n_A}}}} \quad \text{and} \quad \sqrt{n_B}\frac{\hat{\beta}^B_{n_B} - \beta^B_0}{\sqrt{\hat{\sigma}^{-1}_{2,\hat{\theta}^B_{n_B}}}}$$

for group A and B respectively. Since $n_A = 115$ and $n_B = 120$, we approximate the law of

$$\frac{\hat{\beta}^A_{n_A} - \hat{\beta}^B_{n_B} - (\beta^A_0 - \beta^B_0)}{\sqrt{\frac{\hat{\sigma}^{-1}_{2,\hat{\theta}^A_{n_A}}}{n_A} + \frac{\hat{\sigma}^{-1}_{2,\hat{\theta}^B_{n_B}}}{n_B}}}$$

by $\mathcal{N}(0,1)$.

Testing $H_0 : \beta^A_0 = \beta^B_0$ against $H_1 : \beta^A_0 \neq \beta^B_0$, we accept H_0 at the 5% level using both at-random-analysis ($p = 0.96$) and the joint modeling approach ($p = 0.70$). However, the at-random-analysis seems again to yield results biased towards the null, which suggests that is some situations, the at-random-analysis would falsely conclude to absence of difference. Again, in case of nonignorable dropout, our joint model and its asymptotic properties appear as an appealing alternative to reduce this bias.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

As a by-product of the joint modeling approach, we obtain estimates for the parameters of the longitudinal covariate. This would allow us to test also for differences in these parameters between groups.

Example 2. Our results are now used to analyze data consisting of repeated measures of CD4 lymphocyte counts on 101 patients with advanced disease due to the human immunodeficiency virus (HIV) (source: `http://www.maths.lancs.ac.uk/diggle/lda/Datasets/`). CD4 lymphocyte count is known to be associated with clinical outcome and the degree of this association is of interest from a prognostic viewpoint. A standard analysis is to use the Cox regression model to evaluate the relationship between CD4 count (the covariate) and survival.

Each subject is followed until death (there is no censoring in our data). The current value of CD4 count on the time interval where death occurs can not be observed and is then missing for inference in the Cox model. The LCF and two-stage approaches are the most widely used methods in practice, to resolve such an issue in estimating the Cox model. However, the LCF method can induce considerable bias into the estimate of the true association of interest. Simulations of Dafni and Tsiatis (1998) in the context of covariate measurement error in the Cox model indicate that the two-stage approach yields parameter estimates that are still biased towards the null, even though the majority of the bias using more naive modeling approaches is eliminated. This method may, however, still give erroneous results.

As is customary, CD4 counts were log-transformed to ensure that the assumption of normality is better satisfied, that is Z represents here log CD4. We then estimated the regression parameter in the Cox model using both LCF and two-stage methods, and our joint model. Results are presented in Table 2.

It is anticipated that imputing the missing value by LCF will yield a seriously biased estimate of β since CD4 count may vary dramatically before death. Failure of the LCF approach to capture this variation

Table 2. Cox regression analysis of time to death for patients with AIDS. No censoring.

Estimation method	Estimate of β	Estimated standard error (S.E.)	Est./S.E.	p-value
LCF	−0.2634	0.1872	−1.4071	0.158
Two-stage	−0.5190	0.1829	−2.8376	0.005
Joint model	−0.6498	0.1909	−3.4039	<0.001

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

indeed leads to acceptation of the null hypothesis of no association between CD4 and death ($p = 0.158$). In comparison to this approach, the two-stage method results in a much less biased estimate of the regression parameter β ($\hat{\beta} = -0.5190$) and concludes to a significant association between log CD4 and death ($p < 0.004$). The parameter we obtain from the joint model is further from the null hypothesis ($\hat{\beta} = -0.6498$ and $p < 0.001$), indicating that we have probably reduced the bias even further.

The estimated standard errors for the LCF, two stage and joint modeling approaches are, respectively, 0.1872, 0.1829, 0.1909. The standard error obtained from the joint model is again greater, as in Example 1.

In comparing the estimates obtained from the three approaches, we see that our estimate of the association of death and log CD4 is less biased towards the null. We may then be interested in assessing the effect of censoring of the terminal event on relative performance of the three methods. As a preliminary attempt to evaluate this, we carried out the following "simulation": we randomly censored 10%, 25% and 50% of the study subjects, and we fitted the Cox model using the three methods. Results are presented in Table 3.

From these results, we see that presence of censoring increases the estimated standard error (as in the no censoring case, the standard error is greater for the joint model than for the LCF and two-stage approaches). Censoring then results in increasing the p-value for testing

Table 3. Cox regression analysis of time to death for patients with AIDS. Censoring.

% Censoring	Estimation method	Estimate of β	Estimated standard error (S.E.)	Est./ S.E.	p-value
10	LCF	−0.2629	0.1970	−1.3345	0.183
	Two-stage	−0.5211	0.1923	−2.7098	0.007
	Joint model	−0.6569	0.1913	−3.4339	<0.001
25	LCF	−0.2931	0.2157	−1.3588	0.174
	Two-stage	−0.5433	0.2112	−2.5724	0.010
	Joint model	−0.6727	0.2205	−3.0508	0.002
50	LCF	−0.2359	0.2647	−0.8912	0.373
	Two-stage	−0.4943	0.2597	−1.9034	0.057
	Joint model	−0.6314	0.2672	−2.3630	0.018

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

the null hypothesis of no association between death and log CD4 count. For 50% censoring, the two-stage approach even leads to the conclusion that at the 5% level, there is no association between death and the longitudinal marker. We note that the test obtained from the joint model is still significant at the 5% level.

5. DISCUSSION

We have proposed an estimator for the asymptotic variance of SPML estimators in a joint model for survival and longitudinal data which was developed by Dupuy and Mesbah (2002) to estimate the Cox model with a missing time-dependent covariate. This variance estimator is consistent and easy to compute. Tests of hypothesis based on more "naive" methods of estimation in the Cox model with a missing time-dependent covariate may yield erroneous conclusions. This is particularly evident when the censoring is heavy or when missingness of the covariate is falsely assumed to be at random instead of nonignorable. Our method provides a mean of correcting this. Simulations for comparing our joint model to some alternative methods are subject for future research.

The validity of our approach, asymptotic results, and of the statistical analysis based on them (e.g., tests of hypothesis) however depends on the parametric distributional assumptions for the longitudinal covariate. Future research is then needed to investigate sensitivity to violation of these assumptions.

REFERENCES

Altman, D. G., De Stavola, B. L. (1994). Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates. *Statist. Med.* 13:301–341.

Bagdonavičius, V., Nikulin, M. S. (2001). *Accelerated Life Models: Modeling and Statistical Analysis*. Monographs on Statistics and Applied Probability, Chapman & Hall.

Chen, H. Y., Little, R. J. A. (1999). Proportional hazards regression with missing covariates. *J. Am. Statist. Assoc.* 94:896–908.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc., Ser. B* 34:187–220.

Dafni, U. G., Tsiatis, A. A. (1998). Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics* 54:1445–1462.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Dupuy, J.-F., Mesbah, M. (2002). Joint modeling of event time and nonignorable missing longitudinal data. *Lifetime D. Anal.* 8: 99–115.

Dupuy, J.-F., Grama, I., Mesbah, M. Asymptotic theory for the Cox model with missing time-dependent covariate. Submitted, 2002.

Dupuy, J.-F., Grama, I., Mesbah, M. (2003). Normalité asymptotique des estimateurs semiparamétriques dans le modèle de Cox avec covariable manquante nonignorable. *C. R. de l'Acad. des Sciences de Paris, Sér. I* 336:81–84.

Faucett, C. L., Thomas, D. (1995). Simultaneously modeling censored survival data and repeatedly measured covariates using Gibbs sampling. *Statist. Med.* 15:1663–1685.

Henderson, R., Diggle, P., Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* 1:465–480.

Ibrahim, J. A., Chen, M.-H., Sinha, D. (2001). *Bayesian Survival Analysis*. Springer Series in Statistics, New York: Springer-Verlag.

Johansen, S. (1983). An extension of Cox's regression model. *Internat. Statist. Rev.* 51:258–262.

Lin, D., Ying, Z. (1993). Cox regression with incomplete covariate measurements. *J. Am. Statist. Assoc.* 88:1341–1349.

Li, Y., Lin, X. (2000). Covariate measurement errors in frailty models for clustered survival data. *Biometrika* 87:849–866.

Mesbah, M., Cole, B. F., Lee, M. L. T. (2002). *Statistical Methods for Quality of Life Studies: Design, Measurement and Analysis*. Boston: Kluwer.

Mesbah, M., Dupuy, J.-F., Heutte, N., Awad, L. Joint analysis of longitudinal quality of life and survival processes. In: Balakrishnan, N., Rao, C., eds. *Advances in Survival Analysis, Handbook of Statistics*. Vol. 22. Amsterdam: North Holland, 2003. To appear.

Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* 23:182–198.

Murphy, S. A., Sen, P. K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stoch. Proc. Appl.* 39:153–180.

Paik, M. C., Tsai, W.-Y. (1997). On using the Cox proportional hazards model with missing covariates. *Biometrika* 84:579–593.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* 26:183–214.

Scharfstein, D. O., Tsiatis, A. A., Gilbert, P. B. (1998). Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime D. Anal.* 4:355–391.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Song, X., Davidian, D., Tsiatis, A. A. (2002a). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* 58:742–753.

Song, X., Davidian, D., Tsiatis, A. A. (2002b). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics* 3:511–528.

Tsiatis, A. A., Davidian, D. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* 88:447–458.

Wang, C. Y., Hsu, L., Feng, Z. D., Prentice, R. L. (1997). Regression calibration in failure time regression. *Biometrics* 53:131–144.

Wulfsohn, M. S., Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* 53:330–339.

Zhou, H., Pepe, M. S. (1995). Auxiliary covariate data in failure time regression. *Biometrika* 82:139–149.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016