A Semiparametric Estimator for the Proportional Hazards Model with Longitudinal Covariates Measured with Error

Author(s): Anastasios A. Tsiatis and Marie Davidian

Source: *Biometrika*, Vol. 88, No. 2 (Jun., 2001), pp. 447-458
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2673492
Accessed: 24-05-2018 06:14 UTC

Biometrika (2001), **88**, 2, *pp.* 447–458

Printed in Great Britain

A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error

BY ANASTASIOS A. TSIATIS AND MARIE DAVIDIAN
Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203, U.S.A.
tsiatis@stat.ncsu.edu davidian@stat.ncsu.edu

SUMMARY

A common objective in longitudinal studies is to characterise the relationship between a failure time process and time-independent and time-dependent covariates. Time-dependent covariates are generally available as longitudinal data collected periodically during the course of the study. We assume that these data follow a linear mixed effects model with normal measurement error and that the hazard of failure depends both on the underlying random effects describing the covariate process and other time-independent covariates through a proportional hazards relationship. A routine assumption is that the random effects are normally distributed; however, this need not hold in practice. Within this framework, we develop a simple method for estimating the proportional hazards model parameters that requires no assumption on the distribution of the random effects. Large-sample properties are discussed, and finite-sample performance is assessed and compared to competing methods via simulation.

Some key words: Conditional score; Measurement error; Mixed effects model; Regression calibration; Semiparametric; Survival analysis.

1. INTRODUCTION

Many longitudinal studies collect information on each participant both on a time-to-event, henceforth 'survival' or 'failure', and covariates, some of which vary with time. A frequent objective is to characterise the relationship between survival and covariates; the proportional hazards model (Cox, 1972) is a standard framework. To implement the Cox model with time-dependent covariates, complete knowledge of the true covariate history for each subject is required; however, time-dependent covariates are generally measured intermittently, often at different times for each subject and with error. A naive approach is to substitute for each subject at each failure time in the Cox partial likelihood (Cox, 1975) the closest observed covariate value prior to that time, often termed 'last value carried forward'. It is well known (Prentice, 1982) that substituting mismeasured values for true covariates in the Cox model leads to biased estimation.

Recent interest has focused on joint models for longitudinal covariate data and a survival endpoint. A popular approach assumes that the longitudinal data follow a linear mixed effects model (Laird & Ware, 1982) and that survival depends on the covariate through a proportional hazards relationship with the underlying random effects. A common strat-

egy for estimation of the proportional hazard regression parameters is a two-stage approach, where

(i) the mixed effects model is fitted to data at each risk set assuming normality both of random effects and intra-subject error from which estimated empirical Bayes, i.e. best linear unbiased predictors of the random effects are obtained, and

(ii) predictors for the covariate for each subject at each failure time based on the relevant fit are substituted for the true covariate values in the Cox partial likelihood (Pawitan & Self, 1993; Tsiatis et al., 1995; Dafni & Tsiatis, 1998).

This approximate method uses regression calibration (Carroll et al., 1995, Ch. 3) to reduce bias of the naive approach but may still yield biased estimators for large measurement error. Alternatively, the joint likelihood of the survival and longitudinal data may form the basis for inference. DeGruttola & Tu (1994) assumed the covariate process and survival times to be multivariate normal and fitted the model via parametric maximum likelihood; Wulfsohn & Tsiatis (1997) adopted the less rigid proportional hazards relationship and used nonparametric maximum likelihood, but continued to assume normal random effects. Henderson et al. (2000) used normal random effects in Gaussian covariate processes. Faucett & Thomas (1996) assumed normality and took a Bayesian approach.

These strategies rely heavily on the assumption of normality of random effects characterising the true covariate process; however, this assumption may be over-restrictive and the consequences if it is violated are unknown, so it is natural to be concerned that inference on parameters describing the relationship between survival and covariates may be compromised. In this paper, we assume that survival is related to the covariate through a proportional hazards relationship with the underlying random effects; the model is formulated in § 2. In § 3, we develop a simple method for inference that does not put any restrictions on the distribution of the random effects by exploiting the conditional score approach of Stefanski & Carroll (1987). Large-sample properties are discussed in § 4, and performance in finite samples of both the new and competing estimators is demonstrated in § 5.

2. MODEL AND ASSUMPTIONS

For each subject i ($i = 1, \ldots, n$) let T_i and C_i denote times to failure and censoring, respectively, where time on study $V_i = \min(T_i, C_i)$ and failure indicator $\Delta_i = I(T_i \leqslant C_i)$ are observed; all variables are independent across i. Let Z_i denote time-independent covariates and $X_i(u)$ denote time-dependent covariates at time u for subject i; for simplicity, we assume $X_i(u)$ scalar, but generalisation to vector-valued $X_i(u)$ is straightforward. Assume that $X_i(u)$ follows a subject-specific linear model; we take $X_i(u) = \alpha_{0i} + \alpha_{1i}u$, where $\alpha_i = (\alpha_{0i}, \alpha_{1i})^{\mathrm{T}}$ are the intercept and slope for i, which may be generalised to more complex polynomial or regression spline growth curves. The covariate process $X_i(u)$ is not directly observed; rather, longitudinal measurements $W_i(t_{ij})$ are obtained at ordered times $t_i = (t_{i1}, \ldots, t_{im_i})^{\mathrm{T}}$, for $t_{im_i} \leqslant V_i$, where $W_i(t_{ij}) = X_i t_{ij}) + e_{ij}$, with $e_i = (e_{i1}, \ldots, e_{im_i})^{\mathrm{T}}$. The errors e_{ij} reflect uncertainty in measuring $X_i(u)$ at t_{ij} and are assumed identically normally distributed and independent with mean zero and variance σ^2, independent of $(T_i, C_i, \alpha_i, Z_i, t_i, m_i)$. More precisely,

$$(e_i | T_i, C_i, \alpha_i, Z_i, t_i, m_i) \sim N_{m_i}(0, \sigma^2 I_{m_i}), \tag{1}$$

where N_q denotes the q-variate normal distribution and I_q the q-dimensional identity matrix. Although normality may be a reasonable assumption for within-subject error in

continuous covariates, the standard assumption that the α_i are multivariate normal may be too restrictive or unrealistic to represent the nature of the true covariate trajectories $X_i(u)$ in the population.

The survival model assumes that the hazard of failure is related to $X_i(u)$ and Z_i through a proportional hazards regression model; that is,

$$\begin{aligned}\lambda_i(u) &= \lim_{du \to 0} du^{-1} \operatorname{pr}\{u \leqslant T_i < u + du \mid T_i \geqslant u, \alpha_i, Z_i, C_i, e_i(u), t_i(u)\} \\ &= \lim_{du \to 0} du^{-1} \operatorname{pr}\{u \leqslant T_i < u + du \mid T_i \geqslant u, \alpha_i, Z_i\} \\ &= \lambda_0(u) \exp\{\gamma X_i(u) + \eta^{\mathrm{T}} Z_i\}, \end{aligned} \tag{2}$$

where $\lambda_0(u)$ denotes an unspecified baseline hazard function, the collection of times of longitudinal measurements up to and including u is denoted by $t_i(u) = (t_{ij} \leqslant u)$, $e_i(u) = (e_{ij} : t_{ij} \leqslant u)$, and η is $(q \times 1)$. Equation (2) makes explicit the nature of our assumption that timing of measurements and censoring are noninformative. In particular, timing of measurements is noninformative in the sense that it does not impart additional prognostic effect beyond that given by the covariates and the α_i. A natural question involves the restriction of conditioning to the set $t_i(u)$ in (2). Conditioning on all of t_i would be informative and invalidate the hazard relationship in (2), which is critical to the developments of the next section.

Interest focuses on estimation of the parameters γ and ζ. We seek an estimator that requires no distributional assumption on the random effects.

3. Conditional score estimator

Let $\hat{X}_i(u)$ be the ordinary least squares estimator of $X_i(u)$ using all the longitudinal data up to and including time u, that is based on $t_i(u)$. Note that this requires at least two longitudinal measurements on i up to and including u, for $t_{i2} \leqslant u$. Define the counting process increment

$$dN_i(u) = I(u \leqslant V_i < u + du,\ \Delta_i = 1,\ t_{i2} \leqslant u)$$

and the 'at risk' process

$$Y_i(u) = I(V_i \geqslant u,\ t_{i2} \leqslant u);$$

that is $dN_i(u)$ puts point mass at time u corresponding to the observed death time for the ith subject as long as this occurs after the second longitudinal measurement, and $Y_i(u)$ is the indicator that subject i is at risk with at least two longitudinal measurements at time u. Then the estimator $\hat{X}_i(u)$, conditional on $\{\alpha_i, t_i(u), Y_i(u) = 1, Z_i\}$, is normally distributed with mean $X_i(u) = \alpha_{0i} + \alpha_{1i} u$ and variance $\sigma^2 \theta_i(u)$, the usual variance of the estimated mean $\hat{X}_i(u)$ at u using data up to and including u, which depends on timing of measurements for i up to and including u. For $X_i(u) = \alpha_{0i} + \alpha_{1i} u$,

$$\theta_i(u) = 1/m_{i,u} + (u - \bar{t}_{i,u})^2 / \mathrm{ss}_{i,u},$$

where $t_i(u)$ contains $m_{i,u}$ time-points t_{ij} with mean $\bar{t}_{i,u}$, and $\mathrm{ss}_{i,u} = \sum_{j=1}^{m_{i,u}} (t_{ij} - \bar{t}_{i,u})^2$. For now, we will assume that σ^2 is known; this will be relaxed subsequently.

Our approach is motivated by the conditional score method of Stefanski & Carroll (1987) and the following heuristic argument. At any time u, the conditional density for $\{dN_i(u) = r, \hat{X}_i(u) = x\}$, given i is at risk at time u so that $Y_i(u) = 1$, random effects α_i,

longitudinal measurements taken up to and including time u at times $t_i(u)$, and time-independent covariates Z_i, is

$$\text{pr}\{dN_i(u) = r \,|\, Y_i(u) = 1,\ \hat{X}_i(u) = x,\ \alpha_i,\ Z_i,\ t_i(u)\} \times \text{pr}\{\hat{X}_i(u) = x \,|\, Y_i(u) = 1,\ \alpha_i,\ Z_i,\ t_i(u)\},$$

which equals

$$\frac{[\lambda_0(u)\, du \exp\{\gamma X_i(u) + \eta^{\mathrm{T}} Z_i\}]^r [1 - \lambda_0(u)\, du \exp\{\gamma X_i(u) + \eta^{\mathrm{T}} Z_i\}]^{1-r}}{\{2\pi\sigma^2\theta_i(u)\}^{\frac{1}{2}}} \exp\left[- \frac{\{x - X_i(u)\}^2}{2\sigma^2\theta_i(u)} \right];$$

thus, the conditional likelihood of $\{dN_i(u), \hat{X}_i(u)\}$ given $\{Y_i(u) = 1, \alpha_i, Z_i, t_i(u)\}$, up to order du, is

$$\begin{aligned}
&[\lambda_0(u)\, du \exp\{\gamma X_i(u) + \eta^{\mathrm{T}} Z_i\}]^{dN_i(u)} \frac{\exp[-\{\hat{X}_i(u) - X_i(u)\}^2 / \{2\sigma^2\theta_i(u)\}]}{\{2\pi\sigma^2\theta_i(u)\}^{\frac{1}{2}}} \\
&= \exp\left[X_i(u) \left\{ \gamma dN_i(u) + \frac{\hat{X}_i(u)}{\sigma^2\theta_i(u)} \right\} \right] \frac{\{\lambda_0(u) \exp(\eta^{\mathrm{T}} Z_i)\, du\}^{dN_i(u)}}{\{2\pi\sigma^2\theta_i(u)\}^{\frac{1}{2}}} \exp\left\{ - \frac{\hat{X}_i^2(u) + X_i^2(u)}{2\sigma^2\theta_i(u)} \right\}.
\end{aligned} \tag{3}$$

This representation implies that, conditional on $Y_i(u) = 1$,

$$S_i(u, \gamma, \sigma^2) = \gamma\sigma^2\theta_i(u)\, dN_i(u) + \hat{X}_i(u)$$

is a complete sufficient statistic for α_i, suggesting that, at each time u, conditioning on $S_i(u, \gamma, \sigma^2)$ would remove the dependence of the conditional distribution on the random effects α_i. In the Appendix, we show that the conditional intensity process defined as

$$\lim_{du \to 0} du^{-1} \text{pr}\{dN_i(u) = 1 \,|\, S_i(u, \gamma, \sigma^2), Z_i, t_i, (u), Y_i(u)\}$$

is equal to

$$\lambda_0(u) \exp\{\gamma S_i(u, \gamma, \sigma^2) - \gamma^2\sigma^2\theta_i(u)/2 + \eta^{\mathrm{T}} Z_i\} Y_i(u). \tag{4}$$

We now outline the reasoning underlying the conditional score estimator, which follows by analogy with that for estimators for the proportional hazards model with no measurement error. The conditional intensity of $dN(u) = \sum_{j=1}^n dN_j(u)$ given

$$\{S_i(u, \gamma, \sigma^2), Z_i, t_i(u), Y_i(u), i = 1, \ldots, n\}$$

is $\lambda_0(u) E_0(u, \gamma, \eta, \sigma^2)$, where

$$E_0(u, \gamma, \eta, \sigma^2) = \sum_{j=1}^n E_{0j}(u, \gamma, \eta, \sigma^2),$$

$$E_{0j}(u, \gamma, \eta, \sigma^2) = \exp\{\gamma S_j(u, \gamma, \sigma^2) - \gamma^2\sigma^2\theta_j(u)/2 + \eta^{\mathrm{T}} Z_j\} Y_j(u).$$

This suggests that a reasonable estimator for $\lambda_0(u)\, du$ is given by

$$\hat{\lambda}_0(u)\, du = dN(u) / E_0(u, \gamma, \eta, \sigma^2).$$

By analogy with the usual score equations derived from the partial likelihood in a proportional hazards model, we suggest estimating (γ, η) by solving the $(q+1) \times 1$ set of estimating equations

$$\sum_{i=1}^n \int \{S_i(u, \gamma, \sigma^2), Z_i^{\mathrm{T}}\}^{\mathrm{T}} \{dN_i(u) - E_{0i}(u, \gamma, \eta, \sigma^2) \hat{\lambda}_0(u)\, du\} = 0,$$

which, upon substitution of $\hat{\lambda}_0(u)$ for $\lambda_0(u)$, may be written as

$$\sum_{i=1}^{n} \int \{S_i(u, \gamma, \sigma^2), Z_i^{\mathrm{T}}\}^{\mathrm{T}} \left\{ dN_i(u) - \frac{dN(u) E_{0i}(u, \gamma, \eta, \sigma^2)}{E_0(u, \gamma, \eta, \sigma^2)} \right\} = 0. \tag{5}$$

Defining

$$E_{1j}(u, \gamma, \eta, \sigma^2) = \{S_j(u, \gamma, \sigma^2), Z_j^{\mathrm{T}}\}^{\mathrm{T}} \exp\{\gamma S_j(u, \gamma, \sigma^2) - \gamma^2 \sigma^2 \theta_j(u)/2 + \eta^{\mathrm{T}} Z_j\} Y_j(u),$$

$$E_1(u, \gamma, \eta, \sigma^2) = \sum_{j=1}^{n} E_{1j}(u, \gamma, \eta, \sigma^2),$$

and interchanging the sums in (5), we may express the estimating equations as

$$\sum_{i=1}^{n} \int \left[\{S_i(u, \gamma, \sigma^2), Z_i^{\mathrm{T}}\}^{\mathrm{T}} - \frac{E_1(u, \gamma, \eta, \sigma^2)}{E_0(u, \gamma, \eta, \sigma^2)} \right] dN_i(u) = 0. \tag{6}$$

With no measurement error, $\sigma^2 = 0$, (6) is identical to the score equations for the maximum partial likelihood estimator of Cox (1975). Moreover, with $X_i(u)$ time-independent and σ^2 known, it is straightforward to show that the equations are asymptotically equivalent to those proposed by Nakamura (1992). Nakamura claimed that his estimator is approximate, with the asymptotic theory only holding for σ^2 decreasing to zero with sample size; however, because of the equivalence, the results of § 4 suggest that the Nakamura estimator is consistent and asymptotically normal even for fixed $\sigma^2 > 0$. An alternative semiparametric estimator with time-independent covariates is given by Buzas (1998).

4. Large-sample properties

We give a heuristic sketch of steps involved in showing that solving (6) with σ^2 known should yield consistent, asymptotically normal estimators for (γ, η). We demonstrate these properties via simulation in § 5.

Defining $\bar{S}(u, \gamma, \eta, \sigma^2) = E_1(u, \gamma, \eta, \sigma^2)/E_0(u, \gamma, \eta, \sigma^2)$ to be the weighted average of vectors $\{S_i(u, \gamma, \sigma^2), Z_i^{\mathrm{T}}\}^{\mathrm{T}}$ among individuals i at risk at time u, and letting $\mu(u, \gamma, \eta, \sigma^2)$ denote the probabilistic limit of $\bar{S}(u, \gamma, \eta, \sigma^2)$, by adding and subtracting common terms, the equivalent version of (6), (5), may be written

$$\sum_{i=1}^{n} \int [\{S_i(u, \gamma, \sigma^2), Z_i^{\mathrm{T}}\}^{\mathrm{T}} - \bar{S}(u, \gamma, \eta, \sigma^2)] \{dN_i(u) - E_{0i}(u, \gamma, \eta, \sigma^2) \lambda_0(u)\, du\}, \tag{7}$$

which may be reexpressed as

$$\sum_{i=1}^{n} \int [\{S_i(u, \gamma, \sigma^2), Z_i^{\mathrm{T}}\}^{\mathrm{T}} - \mu(u, \gamma, \eta, \sigma^2)] \{dN_i(u) - E_{0i}(u, \gamma, \eta, \sigma^2) \lambda_0(u)\, du\} \tag{8a}$$

$$+ \sum_{i=1}^{n} \int \{\mu(u, \gamma, \eta, \sigma^2) - \bar{S}(u, \gamma, \eta, \sigma^2)\} \{dN_i(u) - E_{0i}(u, \gamma, \eta, \sigma^2) \lambda_0(u)\, du\}. \tag{8b}$$

We first outline steps needed to conclude the existence of consistent solutions to (6) by exploiting its representation as the sum of (8a) and (8b). If set equal to zero with the function $\mu(u, \gamma, \eta, \sigma^2)$ known, (8a) is an unbiased estimating equation for (γ, η), which follows as, at the true values (γ_0, η_0), (8a) is a sum of independent and identically distributed

zero-mean random vectors. This may be shown by taking the expectation inside the integral for the ith summand and conditioning on $[\{S_i(u, \gamma_0, \sigma^2), Z_i^{\mathrm{T}}\}^{\mathrm{T}}, Z_i, t_i(u), Y_i(u)]$, which yields

$$\int E\{[\{S_i(u, \gamma_0, \sigma^2), Z_i^{\mathrm{T}}\}^{\mathrm{T}} - \mu(u, \gamma_0, \eta_0, \sigma^2)] \times (E[dN_i(u)|\{S_i(u, \gamma_0, \sigma^2), Z_i^{\mathrm{T}}\}^{\mathrm{T}}, Z_i, t_i(u), Y_i(u)] - E_{0i}(u, \gamma_0, \eta_0, \sigma^2)\lambda_0(u)\, du)\}; \quad (9)$$

as

$$E[dN_i(u)|\{S_i(u, \gamma_0, \sigma^2), Z_i^{\mathrm{T}}\}^{\mathrm{T}}, Z_i, t_i(u), Y_i(u)] = E_{0i}(u, \gamma_0, \eta_0, \sigma^2)\lambda_0(u)\, du$$

is the conditional intensity in (4), the inner expectation is equal to zero, so that (9) is zero, demonstrating the unbiasedness. That n^{-1} times (8b) converges in probability to zero uniformly in a neighbourhood $\mathcal{N}(\gamma_0, \eta_0)$ of (γ_0, η_0) follows from the inequality

$$\sup_{\mathcal{N}(\gamma_0, \eta_0)} \left| \int \{\mu(u, \gamma, \eta, \sigma^2) - \bar{S}(u, \gamma, \eta, \sigma^2)\} n^{-1} \sum_{i=1}^{n} \{dN_i(u) - E_{0i}(u, \gamma, \eta, \sigma^2)\lambda_0(u)\, du\} \right|$$

$$\leqslant \sup_{\mathcal{N}(\gamma_0, \eta_0)} \left[\sup_u \{|\mu(u, \gamma, \eta, \sigma^2) - \bar{S}(u, \gamma, \eta, \sigma^2)|\} \right] \quad (10)$$

$$\times \left[n^{-1} \sum_{i=1}^{n} \int dN_i(u) + n^{-1} \sum_{i=1}^{n} \sup_{\mathcal{N}(\gamma_0, \eta_0)} \left\{ \int E_{0i}(u, \gamma, \eta, \sigma^2)\lambda_0(u)\, du \right\} \right]. \quad (11)$$

The first term in (11) is bounded by 1, and the second converges to

$$E\left\{ \sup_{\mathcal{N}(\gamma_0, \eta_0)} \int E_{0i}(u, \gamma, \eta, \sigma^2)\lambda_0(u)\, du \right\}$$

in probability. Uniform convergence of $n^{-1}E_0(u, \gamma, \eta, \sigma^2)$ and $n^{-1}E(u, \gamma, \eta, \sigma^2)$, and hence of $\bar{S}(u, \gamma, \eta, \sigma^2)$, both in u and (γ, η) in $\mathcal{N}(\gamma_0, \eta_0)$, could be established by a modification of the Glivenko–Cantelli lemma, thus showing convergence in probability to zero of (10). Collecting these developments demonstrates the result.

Combining these arguments, we deduce that the behaviour of the estimators solving (6) will be dictated by (8a). Since (8a) set to zero is an unbiased estimating equation, under regularity conditions, a consistent sequence of solutions to it exists, indicating the existence of consistent solutions to (6). By analogy with Stefanski & Carroll (1987), (6) need not have a unique solution, and inconsistent solutions may exist. Our experience shows, however, that this problem may not be a significant drawback in practice; we discuss this issue further in § 5.

We now argue that, at (γ_0, η_0), $n^{-\frac{1}{2}}$ times (6), equal to the sum of $n^{-\frac{1}{2}}$ times the sum of (8a) and (8b), is asymptotically normal with mean zero, so that the consistent solution $(\hat{\gamma}, \hat{\eta})$ should be asymptotically normal. In contrast to the case $\sigma^2 = 0$, standard martingale theory arguments do not apply. However, evaluated at (γ_0, η_0), $n^{-\frac{1}{2}}$ times (8a) is a normalised sum of independent and identically distributed zero-mean random vectors from above, so is asymptotically normal. Expression (8b) times $n^{-\frac{1}{2}}$ may be written

$$\int n^{\frac{1}{2}}\{\mu(u, \gamma_0, \eta_0, \sigma^2) - \bar{S}(u, \gamma_0, \eta_0, \sigma^2)\} n^{-1} \sum_{i=1}^{n} \{dN_i(u) - E_{0i}(u, \gamma_0, \eta_0, \sigma^2)\lambda_0(u)\, du\}. \quad (12)$$

By arguments similar to those in Breslow & Crowley (1974, Theorem 4) or Tsiatis (1981, Theorem 5.1), we expect that, under regularity conditions, (12) converges in probability to zero. If we combine these results, $n^{-\frac{1}{2}}$ times (6), evaluated at the truth, is asymptotically equivalent to $n^{-\frac{1}{2}}$ times (8a), and thus converges to a normal random vector with mean zero and covariance matrix equal to that of a single element in the sum (8a).

The derivation of § 3 and the above arguments assume σ^2 is known; in practice, σ^2 will almost always be unknown. A natural strategy is to replace σ^2 in (6) by the pooled estimator

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} I(m_i > 2) R_i}{\sum_{i=1}^{n} I(m_i > 2)(m_i - 2)}, \tag{13}$$

where R_i is the residual sum of squares for the least squares fit to all m_i observations for subject i, and solve for (γ, η). That $\hat{\sigma}^2$ is a consistent estimator for σ^2 under our assumptions is shown in the Appendix. It is straightforward to show that

$$n^{\frac{1}{2}}(\hat{\sigma}^2 - \sigma^2) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \{E(m_i) - 2\}^{-1} \{R_i - (m_i - 2)\sigma^2\} + o_p(1).$$

Thus, because (6) is asymptotically equivalent to the sum (8a) of independent and identically distributed terms, standard M-estimator arguments as in Carroll et al. (1995, §§ A.3.3, A.3.6) apply, from which it may be concluded that (6) with σ^2 replaced by $\hat{\sigma}^2$ again converges to a zero-mean multivariate normal random vector.

To obtain approximate standard errors for the estimators $(\hat{\gamma}, \hat{\eta})$, these developments suggest that the usual 'sandwich' technique may be used; for example, when $\hat{\sigma}^2$ is substituted, the asymptotic variance of $(\hat{\gamma}, \hat{\eta})$ may be deduced by appealing to the results of Carroll et al. (1995, § A.3.6). In the simulations of § 5, we demonstrate that this strategy yields reliable estimates of uncertainty.

5. Simulation studies

We carried out a number of simulation studies under the following scenario, a modification of that in Dafni & Tsiatis (1998). For simplicity, we focus on the situation of a single, time-dependent covariate $X_i(u)$ and no time-independent covariate, so that $\eta = 0$, and estimation of γ only is of interest. In each case, $E(\alpha_i) = (4{\cdot}173, -0{\cdot}0103)^{\mathrm{T}}$, $\gamma = -1{\cdot}0$, and the censoring distribution was exponential with mean 110 weeks, with additional censoring at the end of the study, 80 weeks. Nominal times of observation for $X_i(u)$ were at (0, 2, 4, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80) weeks, with a 10% chance of a missing observation at any time except baseline, and the hazard was given by (2) for $u \geqslant 16$ and 0 otherwise to represent the common clinical trial situation where early observation is frequent and events are not seen immediately following enrolment. Three true, underlying random effects distributions were considered: normal, with $\mathrm{cov}(\alpha_i) = D$, where D has distinct elements $(D_{11}, D_{12}, D_{22}) = (1{\cdot}24, -0{\cdot}0114, 0{\cdot}003)$; a bimodal mixture of normals with mixing proportion 0·5 generated as described in Davidian & Gallant (1993, § 5) with their sep = 4 and R chosen to yield $\mathrm{cov}(\alpha_i)$ with the same diagonal elements in the normal case and $D_{12} = 0{\cdot}039$; and a bivariate skew-normal distribution (Azzalini & Dalla Valle, 1996) chosen so that $\mathrm{cov}(\alpha_i)$ was the same as the normal case and the components α_{0i} and α_{1i} had coefficients of skewness of $-0{\cdot}07$ and 0·85, respectively, representing moderately skewed α_{1i}.

In each scenario, 500 Monte Carlo datasets were generated, and, for each, γ was estimated in five ways:

(i) using the 'ideal' estimator that could be obtained by fitting by partial likelihood with time-dependent covariates if the true values $X_i(u)$ were available for all subjects at each failure time;
(ii) via the conditional score method with σ^2 estimated as in (13);
(iii) using a version of regression calibration as in Dafni & Tsiatis (1998), described below, where empirical Bayes predictors were substituted in the usual partial likelihood;
(iv) using 'naive regression', where predicted values were imputed at each failure time from a single least squares fit to all the available longitudinal data for each i with $m_i \geqslant 2$ and substituted in the usual partial likelihood; and
(v) via 'last value carried forward' as described in § 1.

Ideally, implementation of regression calibration (iii), as discussed by Tsiatis et al. (1995) and Dafni & Tsiatis (1998), would require that a linear mixed model be fitted to the longitudinal data from all individuals at risk at each failure time and that predicted values be imputed for each subject still at risk at each failure time based on the relevant fit. This is computationally prohibitive in a simulation, as the number of mixed model fits equals the number of failures. As an alternative, we implemented this method by fitting the mixed model four times, using available data up to the 25th, 50th, 75th and 100th percentiles of the ordered failure times, respectively, and imputed values for each i still at risk at each failure time within each quartile. Preliminary simulations demonstrated that differences in performance between full regression calibration and this simpler modification are negligible. To calculate approximate standard errors, the expression for the usual Cox partial likelihood was used in (i) and (iii)–(v), with no adjustment for imputation in (iii)–(v). This coincides with suggestions in the literature and standard practice; (iv) and (v) are typically used in this way, and, for (iii), several authors (Tsiatis et al., 1995; Dafni & Tsiatis, 1998) have remarked that adjustment of standard errors to account for repeated fitting of the mixed model seems unnecessary in Monte Carlo studies, and, moreover, calculation of such adjustment would be difficult. For (ii), standard errors were calculated using the 'sandwich' method outlined in § 4; the expression for the asymptotic variance of the conditional score estimator of γ is given in the Appendix. For all methods, 95% Wald confidence intervals for γ based on the standard normal critical value 1·96 were constructed. Although previous studies of (iii) have investigated validity of standard errors, performance of confidence intervals has not been evaluated.

For a range of n and σ^2 values and over the three α_i distributions, we investigated the severity of the problem of multiple roots for (6), discussed in § 4; as pointed out by L. A. Stefanski in an unpublished paper presented at the Third International Workshop on Statistical Modeling, this issue is expected to persist for all n. In the context of generalised linear models, Stefanski conjectured that the consistent root of the conditional score equation will be closest to the naive estimator obtained by solving the usual estimating equation with mismeasured covariates treated as the true values. Accordingly, Stefanski suggested that using the naive estimator as the starting value would be a practical strategy for locating the consistent root. We have observed that a similar phenomenon holds true for our estimator. Figure 1 shows (6) as a function of γ for one of the simulated datasets with $n = 200$, $\sigma^2 = 0{\cdot}30$ and α_i following the mixture of normals, and is representative of the pattern for all datasets inspected across choices of n, σ^2 and α_i distribution. Figure 1(a) shows that the solution to (6) near $\gamma = -1{\cdot}0$ is well determined and close to those for the

'naive' last value carried forward and naive regression estimators. Figure 1(b) indicates further that, if Stefanski's strategy were employed, other potential solutions to (6) would not be considered.

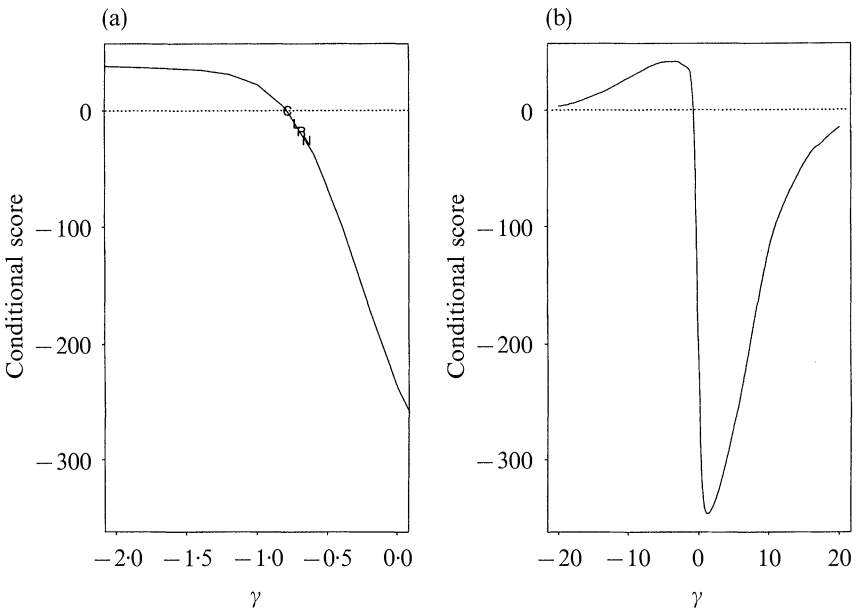

Fig. 1. Equation (6) as a function of γ for a representative dataset, $n = 200$, $\sigma^2 = 0{\cdot}30$, α_i from the mixture scenario. (a) shows a plot over the range of γ near the true value; C denotes the conditional score estimate, R denotes the regression calibration estimate, N denotes the naive regression estimate, and L denotes the last value carried forward estimate. (b) shows a plot over a wide range of γ values.

Results for $n = 200$ and $\sigma^2 = 0{\cdot}30$ are summarised in Table 1. The conditional score estimator shows negligible bias coinciding with that of the unachievable 'ideal' estimator across all distributions, and estimated standard errors track the Monte Carlo standard deviation well. In contrast, the regression calibration, naive regression and last value carried forward methods all exhibit nonnegligible biases; this is most dramatic for the latter two. The most striking results are those for Wald coverage probabilities. The conditional score method provides intervals that achieve the nominal level, as does the 'ideal'; the remaining feasible methods achieve coverages well below the nominal level. We investigated other choices for n and σ^2 with similar results. The conditional score estimator

Table 1. *Simulation results for three underlying random effect distributions*

	Normal				Mixture				Skewed			
Method	Mean	SD	SE	Cov	Mean	SD	SE	Cov	Mean	SD	SE	Cov
I	−1·01	0·08	0·09	0·96	−1·02	0·10	0·12	0·95	−1·01	0·09	0·08	0·97
CS	−1·01	0·11	0·12	0·95	−1·03	0·24	0·25	0·95	−1·01	0·12	0·12	0·95
RC	−0·93	0·08	0·09	0·87	−0·88	0·07	0·09	0·75	−0·92	0·07	0·09	0·85
NR	−0·88	0·07	0·08	0·65	−0·83	0·06	0·08	0·44	−0·88	0·07	0·08	0·68
LV	−0·87	0·07	0·08	0·67	−0·87	0·06	0·08	0·65	−0·86	0·07	0·08	0·61

SD, Monte Carlo standard deviation; SE, average of estimated standard errors; Cov, Wald coverage probabilities. Methods: I, 'ideal'; CS, conditional score; RC, regression calibration; NR, naive regression; LV, last value carried forward.

always showed the same, negligible degree of bias as the ideal and achieved nominal coverage, and the bias and optimism of confidence intervals for methods (iii)–(v) worsened with increasing σ^2.

6. DISCUSSION

We have proposed an estimator for a popular joint model for survival and longitudinal data that is semiparametric in not requiring a distributional assumption on random effects characterising the longitudinal covariate process. The estimator is easily computed; S-Plus functions implementing all approaches studied in § 5 are available from the authors. Although technically the proposed estimating equation may have multiple roots, using the strategy outlined in § 5 we have found that the 'correct' consistent solution may be identified reliably in practice. Rigorous proofs of the large-sample properties of the estimator would be highly technical and are an open problem.

When the hazard is thought to depend on α_i through $X_i(u)$, for data where the proportion of individuals with few longitudinal measurements that do not span the entire range of time on study is large, this method may yield unstable results. However, under such conditions, attempting to deduce reliably the relationship of hazard to $X_i(u)$ may be a fruitless enterprise regardless of estimation method. Although we have presented the estimator for this case, the conditional score approach we propose could be applied when the hazard depends on any linear combination of the elements of α_i. Moreover, the method is applicable to the case of time-independent covariates at baseline; if replicate such measurements were available, σ^2 could be estimated by analogy with the method given in § 4.

Since we focus on the situation where no assumption is made on the distribution of the random effects, our approach requires that, at any risk time u, estimation of $X_i(u)$ does not involve longitudinal covariate data beyond time u. This may result in a loss of efficiency relative to models in which a parametric specification for the random effects is made. A parametric specification provides the necessary structure for making use of these data; see, for example, the EM algorithm of Wulfsohn & Tsiatis (1997), where the random effects are assumed normally distributed.

If one wishes to relax parametric assumptions, an alternative strategy to the one proposed here would be a likelihood-based approach such as an extension of the semiparametric method proposed by Hu et al. (1998, § 3.2) in the case of time-independent X_i, where the distribution of true X_i is restricted only in the sense that it is required to have a 'smooth' density. However, in the presence of additional covariates Z_i, the likelihood approach requires modelling of the joint distribution of X_i and Z_i and can entail a significant computational challenge.

Our estimator is semiparametric in the sense that the random effects distribution is left unspecified. The semiparametric efficiency bound and construction of a semiparametric efficient estimator are open problems.

ACKNOWLEDGEMENT

The research of both authors are supported by grants from the National Cancer Institute and the National Institute of Allergy and Infectious Diseases.

Appendix

Technical details

Derivation of (4). After some straightforward algebra, (3) can be expressed as

$$\{\lambda_0(u)\,du\exp(\eta^{\mathrm{T}}Z_i)\}^{dN_i(u)}\exp\left\{-\frac{S_i^2(u,\gamma,\sigma^2)}{2\sigma^2\theta_i(u)}+\gamma S_i(u,\gamma,\sigma^2)\,dN_i(u)-\frac{\gamma^2\sigma^2\theta_i(u)\,dN_i(u)}{2}\right\}$$
$$\times K\{u, X_i(u), S_i(u,\gamma,\sigma^2)\}, \tag{A1}$$

where

$$K\{u, X_i(u), S_i(u,\gamma,\sigma^2)\}=\{2\pi\sigma^2\theta_i(u)\}^{-\frac{1}{2}}\exp\left\{\frac{2X_i(u)S_i(u,\gamma,\sigma^2)-X_i^2(u)}{2\sigma^2\theta_i(u)}\right\}.$$

The conditional probability $\mathrm{pr}\{dN_i(u)=1\,|\,S_i(u,\gamma,\sigma^2)=s, Z_i, t_i(u), Y_i(u)=1\}$ is

$$\frac{\int \mathrm{pr}\{dN_i(u)=1,\, S_i(u,\gamma,\sigma^2)=s\,|\,\alpha_i, Z_i, Y_i(u)=1\}p\{\alpha_i\,|\,Z_i, t_i(u), Y_i(u)=1\}\,d\alpha_i}{\mathrm{num}+\int \mathrm{pr}\{dN_i(u)=0,\, S_i(u,\gamma,\sigma^2)=s\,|\,\alpha_i, Z_i, Y_i(u)=1\}p\{\alpha_i\,|\,Z_i, t_i(u), Y_i(u)=1\}\,d\alpha_i}, \tag{A2}$$

where 'num' denotes the numerator of (A2), and $p\{\alpha_i\,|\,Z_i, t_i(u), Y_i(u)=1\}$ is the density of α_i conditional on Z_i, $t_i(u)$ and $Y_i(u)$. By (A1), the numerator of (A2) up to order du is

$$\lambda_0(u)\,du\exp(\eta^{\mathrm{T}}Z_i)\exp\left\{-\frac{s^2}{2\sigma^2\theta_i(u)}+\gamma s-\frac{\gamma^2\sigma^2\theta_i(u)}{2}\right\}\int K\{u, X_i(u), s\}p\{\alpha_i\,|\,Z_i, t_i(u), Y_i(u)=1\}\,d\alpha_i$$

and the denominator up to order 1 is given by

$$\exp\left\{-\frac{s^2}{2\sigma^2\theta_i(u)}\right\}\int K\{u, X_i(u), s\}p\{\alpha_i\,|\,Z_i, t_i(u), Y_i(u)=1\}\,d\alpha_i.$$

Thus, (A2) is equal to $\lambda_0(u)\,du\exp(\eta^{\mathrm{T}}Z_i)\exp\{\gamma s-\gamma^2\sigma^2\theta_i(u)/2\}+o_p(du)$, which implies (4).

Consistency of $\hat{\sigma}^2$ *in* (13). If we define $W_i=A_i\alpha_i+e_i$, where A_i is the $(m_i\times 2)$ matrix with first column all ones and second column t_i, $\hat{\sigma}^2$ is the solution to the estimating equation

$$\sum_{i=1}^{n} I(m_i>2)\{(W_i-A_i\hat{\alpha}_i)^{\mathrm{T}}(W_i-A_i\hat{\alpha}_i)-(m_i-2)\sigma^2\}=0,$$

where

$$\hat{\alpha}_i=(A_i^{\mathrm{T}}A_i)^{-1}A_i^{\mathrm{T}}W_i=\alpha_i+(A_i^{\mathrm{T}}A_i)^{-1}A_i^{\mathrm{T}}e_i.$$

As

$$(W_i-A_i\hat{\alpha}_i)^{\mathrm{T}}(W_i-A_i\hat{\alpha}_i)=e_i^{\mathrm{T}}\{I_{m_i}-A_i(A_i^{\mathrm{T}}A_i)^{-1}A_i^{\mathrm{T}}\}e_i,$$

that this is an unbiased estimating equation for σ^2 follows by noting that

$$E(I(m_i>2)[e_i^{\mathrm{T}}\{I_{m_i}-A_i(A_i^{\mathrm{T}}A_i)^{-1}A_i^{\mathrm{T}}\}e_i-(m_i-2)\sigma^2])$$
$$=E\{E(I(m_i>2)[e_i^{\mathrm{T}}\{I_{m_i}-A_i(A_i^{\mathrm{T}}A_i)^{-1}A_i^{\mathrm{T}}\}e_i-(m_i-2)\sigma^2]\,|\,T_i, C_i, \alpha_i, Z_i, t_i, m_i)\}. \tag{A3}$$

By the assumption on e_i in (1) and standard results on quadratic forms,

$$E[I(m_i>2)e_i^{\mathrm{T}}\{I_{m_i}-A_i(A_i^{\mathrm{T}}A_i)^{-1}A_i^{\mathrm{T}}\}e_i\,|\,T_i, C_i, \alpha_i, Z_i, t_i, m_i]=I(m_i>2)(m_i-2)\sigma^2,$$

showing that the inner conditional expectation in (A3) is zero, as required.

Expression for the 'sandwich' variance of the conditional score estimator for the model of § 5. Here there is no time-independent covariate, so $\eta = 0$. The derivatives of (5),

$$\Lambda_\gamma(\gamma, \sigma^2) = \sum_{i=1}^{n} \int \partial/\partial\gamma\{S_i(u, \gamma, \sigma^2) - \bar{S}(u, \gamma, 0, \sigma^2)\}\, dN_i(u),$$

$$\Lambda_{\sigma^2}(\gamma, \sigma^2) = \sum_{i=1}^{n} \int \partial/\partial\sigma^2\{S_i(u, \gamma, \sigma^2) - \bar{S}(u, \gamma, 0, \sigma^2)\}\, dN_i(u)$$

are straightforward to compute. If we let $(\hat{\gamma}, \hat{\sigma}^2)$ denote the estimators, the variance estimator is given by

$$\Lambda_\gamma^{-2}(\hat{\gamma}, \hat{\sigma}^2) \sum_{i=1}^{n} \left(\left[\int \{S_i(u, \hat{\gamma}, \hat{\sigma}^2) - \bar{S}(u, \hat{\gamma}, 0, \hat{\sigma}^2)\} \left\{ dN_i(u) - dN(u) \frac{E_{0i}(u, \hat{\gamma}, 0, \hat{\sigma}^2)}{E_0(u, \hat{\gamma}, 0, \hat{\sigma}^2)} \right\} \right] + \frac{\Lambda_{\sigma^2}(\hat{\gamma}, \hat{\sigma}^2)\{R_i - (m_i - 2)\hat{\sigma}^2\} I(m_i > 2)}{\sum_{i=1}^{n} I(m_i > 2)(m_i - 2)} \right)^2.$$

REFERENCES

AZZALINI, A. & DALLA VALLE, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715–26.

BRESLOW, N. E. & CROWLEY, J. J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2**, 437–53.

BUZAS, J. S. (1998). Unbiased scores in proportional hazards regression with covariate measurement error. *J. Statist. Plan. Infer.* **67**, 247–57.

CARROLL, R. J., RUPPERT, D. & STEFANSKI, L. A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman and Hall.

COX, D. R. (1972). Regression models and life tables (with Discussion). *J. R. Statist. Soc.* B **34**, 187–220.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–76.

DAFNI, U. G. & TSIATIS, A. A. (1998). Evaluating surrogate markers of clinical outcome measured with error. *Biometrics* **54**, 1445–62.

DAVIDIAN, M. & GALLANT, A. R. (1993). The nonlinear mixed effects model with a smooth random effects density. *Biometrika* **80**, 475–88.

DEGRUTTOLA, V. & TU, X. M. (1994). Modeling progression of CD-4 lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003–14.

HENDERSON, R., DIGGLE, P. & DOBSON, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–80.

FAUCETT, C. J. & THOMAS, D. C. (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statist Med.* **15**, 1663–85.

HU, P., TSIATIS, A. A. & DAVIDIAN, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* **54**, 1407–19.

LAIRD, N. M. & WARE, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963–74.

NAKAMURA, T. (1992). Proportional hazards models with covariates subject to measurement error. *Biometrics* **48**, 829–38.

PAWITAN, Y. & SELF, S. (1993). Modeling disease marker processes in AIDS. *J. Am. Statist. Assoc.* **83**, 719–26.

PRENTICE, R. (1982). Covariate measurement errors and parameter estimates in a failure time regression model. *Biometrika* **69**, 331–42.

STEFANSKI, L. A. & CARROLL, R. J. (1987). Conditional scores and optimal scores in generalized linear measurement error models. *Biometrika* **74**, 703–16.

TSIATIS, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93–108.

TSIATIS, A. A., DEGRUTTOLA, V. & WULFSOHN, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and CD4 counts in patients with AIDS. *J. Am. Statist. Assoc.* **90**, 27–37.

WULFSOHN, M. S. & TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–9.

[*Received November* 1999. *Revised November* 2000]