STATISTICS IN MEDICINE
Statist. Med. **17**, 2061–2077 (1998)

A COMPARISON OF SMOOTHING TECHNIQUES FOR CD4 DATA MEASURED WITH ERROR IN A TIME-DEPENDENT COX PROPORTIONAL HAZARDS MODEL

PAUL BYCOTT[1] AND JEREMY TAYLOR[2]*

[1]*Department of Biometrics, Parke-Davis Pharmaceutical Research, 2800 Plymouth Rd., Ann Arbor, MI 48105, U.S.A.*
[2]*Department of Biostatistics, UCLA School of Public Health, Los Angeles, CA 90095-1772, U.S.A.*

SUMMARY

The use of CD4 + T-lymphocyte counts as a covariate presents some unique challenges in survivorship analyses due to the variability of this marker. If one does not account for the measurement error component of this variability in some manner, the estimate of the relative risk parameter in a time-dependent Cox model is biased towards zero, and coverage levels of confidence intervals may be seriously incorrect. We use a two-stage approach to reduce the variability in the observed CD4 counts in order to obtain a more accurate estimate of the relative risk parameter and more valid summary statistics. In the first stage, population based smoothing methods derived from a random-effects model plus a stochastic process or individual based smoothing methods are used to replace the observed longitudinal CD4 counts with less variable imputes at each failure time. In the second stage, we use the imputes in a time-dependent Cox model to estimate the risk parameter and its associated summary statistics. We compare the smoothing methods in simulation studies and find that the use of these smoothing methods results in a substantial reduction in bias for the true risk parameter estimate, better efficiency, and more accurate coverage rates in confidence intervals. We apply our two-stage smoothing methods to the marker CD4 in the ACTG-019 clinical trial part B. © 1998 John Wiley & Sons, Ltd.

1. INTRODUCTION

In the analysis of biomedical data, a question often of interest is what factors are predictive of time to an event such as death, recurrence of cancer, or AIDS? Failure time data are distinguished from other types of data by the presence of censoring, which implies the availability of only partial information on time. Often, the analysis of these kind of data involves a Cox proportional hazards model .[1] If a covariate changes with time, as opposed to being fixed over time, then one can incorporate this information into a time-dependent Cox model. The hazard function for the Cox model is

$$\lambda(t; Z_i(t)) = \lambda_0(t)\mathrm{e}^{\beta Z_i(t)}$$

* Correspondence to: Jeremy Taylor, Department of Biostatistics, UCLA School of Public Health, Los Angeles, CA 90095-1772, U.S.A.

Contract/grant sponsor: National Institutes of Health
Contract/grant number: AI07370, CA64235, AI29196

CCC 0277–6715/98/182061–17$17.50

Received March 1997
Revised November 1997

where $Z_i(t)$ is the time-dependent covariate, $\lambda_0(t)$ is an unspecified baseline hazard function, and β is the relative risk parameter. The risk parameter β in the time-dependent Cox model represents the effect on the hazard of a unit difference in the covariate at time zero or at any time after entry under the assumption that the effect of the covariate is time invariant. The partial likelihood function[2] for the time-dependent Cox model assuming k unique failure times $(t_1, \ldots, t_k)$ is

$$L(\beta) = \prod_{i=1}^{k} \frac{e^{\beta Z_i(t_i)}}{\sum_{j \in R_i} e^{\beta Z_j(t_i)}}$$

where R_i represents the set of subjects at risk just prior to the ith failure, referred to as the risk set. As we see from the partial likelihood function, this model requires knowledge of the covariate for all subjects in the risk set at the time of each failure.

In a clinical trial, subjects fail on a continuous basis while the marker, however, is usually measured at only discrete time points, and thus no measurements for the covariate exist for members in the risk set when a failure occurs between scheduled follow-up visits. There are a number of approaches,[3] but what is usually done is to pull forward the nearest preceding value of the marker and treat it as if it were the current value of the marker at the failure time. An adaption of this approach discussed by Gail[4] is to use the preceding marker value as an estimate of the current value if it was observed within δ time units prior to the failure, or else to exclude the subject from the risk set for that event. One problem with both of these approaches is that neither accounts for measurement error in the covariate value which, as Raboud *et al.*[5] and Prentice[6] argued, causes the estimated relative risk parameter in the time-dependent Cox model to be biased toward the null, and the extent of this bias is directly proportional to the amount of measurement error in the observed covariate. Also, using the nearest preceding marker value may not be biologically plausible if the marker is tremendously variable, such as are CD4 counts for HIV patients. Additionally, by using the proximate method suggested by Gail, subjects potentially fall in and out of the risk set which reduces power of tests by increasing the estimate of the standard error for the risk parameter. Further, the definition of a proximate value is somewhat arbitrary.

Our aim is to investigate the use of population based smoothing techniques involving mixed models which we can use to estimate missing covariate values in the risk set of a survivorship analysis, and to reduce the impact of measurement error on the estimate of β. We apply all of the methods we discuss in a two-stage approach.[7] In the first stage, we use the smoothing methods to impute missing covariate values for each subject in the risk set, and we repeat this process across all risk sets. In the second stage, we take these imputes and treat them as the true values of CD4 at the time of each failure for purposes of fitting a time-dependent Cox proportional hazards model.

The motivation for looking at smoothing methods to obtain more valid inferences from a time-dependent Cox model arose from the ACTG-019 clinical trial data. The ACTG-019 clinical trial part B was a randomized double-blind trial in asymptomatic HIV infected adults who had CD4 counts of fewer than 500 per cubic millimetre at study entry.[8] The subjects were randomly assigned to one of three treatment arms: placebo; AZT 500 mg per day, or AZT 1500 mg per day. The trial was conducted to determine the safety of AZT and its efficacy in prolonging survival and delaying the onset of AIDS and advanced AIDS-related complex. In this trial, the observed longitudinal CD4 counts had considerable within-subject variability, due both to biological fluctuations in the marker over time and measurement error.

In the two-stage approach as described by Tsiatis *et al.*[7] they use longitudinal modelling to help predict the true CD4 value for each member of the risk set at the time of failure. They assumed the true log CD4 value followed a random-effects model with simple error structure and then, conditional on the marker history, they used the conditional mean of the multivariate normal to impute the missing covariate values, and showed why this conditional mean is appropriate to use in a Cox model. To estimate the parameters of the random-effects models, they fit the model to the prior CD4 measurements of subjects in the risk set, and repeat this procedure at each failure time. This method only conditions on past data and only considers a specific simple error structure.

In an article focused on methods for longitudinal data, Taylor *et al.*[9] considered what stochastic models are appropriate for the analysis of longitudinal CD4 + cell data. In particular, they not only looked at random-effects models with simple error structure, but they also looked at alternative models that included a stochastic process such as Brownian motion or an integrated Ornstein–Uhlenbeck process plus measurement error. They found Brownian motion gave the best fit to the CD4 + cell data in terms of a larger likelihood with fewer parameters to estimate. When fitting these models, they took a one-fourth root transformation of the CD4 counts to better meet the underlying normal assumption in these approaches and to achieve homogeneity of within-subject variance.

LaValley and DeGruttola[10] used a two-stage procedure in the context of an AIDS clinical trial to fit a time-dependent Cox model. They modelled longitudinal log CD4 data using a random-effects model plus a stochastic term, and generated imputed values at each event time using the conditional expectation of the multivariate normal distribution conditional on the observed CD4 counts prior to the event time. They used these imputed values in a time-dependent Cox model for time to a new AIDS-defining infection or death.

Raboud *et al.*[5] took a different approach to modelling CD4 counts as an interim step in fitting a time-dependent Cox model. These authors looked at a variety of individual smoothing techniques including fitting a straight line to an individual's marker values, running means, running medians, and running lines. Somewhat surprisingly, the authors found that simply fitting a straight line to an individual's sequence of markers is the most effective method of reducing bias in the estimate of β in the time-dependent Cox model.

In this paper we evaluate and compare the statistical properties of the estimates from the time-dependent Cox model, when a variety of different smoothing techniques are used in the first stage of the two-stage procedure. We compare smoothing techniques based on different longitudinal models as well as one technique based on individual smoothing.

In Section 2, we discuss the methodology for the smoothing techniques. In Section 3, we describe the design of a simulation study and then discuss the results from the study. In Section 4, we perform some diagnostics on the simulation runs that enhance understanding of some of the results of the study. In Section 5, we apply our methods to the ACTG-019 trial. Finally, in Section 6, we offer some general discussion of the smoothing methods and we make recommendations for their use.

2. METHODS

2.1. Longitudinal Models

Consider the longitudinal data situation where we have repeated measures of a marker on $i = 1, \ldots, N$ subjects at time points $j = 1, \ldots, n_i$. These observations are not necessarily recorded

at the same point in time for each subject and each individual may have a different number of observations. We assume that the model for the observed value of the covariate, $\mathbf{Y}_i(\mathbf{t})$ is

$$\mathbf{Y}_i(\mathbf{t}) = \mathbf{Y}_i^*(\mathbf{t}) + \varepsilon_\mathbf{i}, \quad i = 1, \ldots, N$$
$$\varepsilon_\mathbf{i} \sim \mathrm{N}_{n_i}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_\mathbf{i})$$

where $\mathbf{Y}_i^*(\mathbf{t})$ is the true unobserved covariate value and $\varepsilon_\mathbf{i}$ is a measurement error term, with variance parameter σ_ε^2, and $\mathbf{I}_\mathbf{i}$ is the identity matrix. We further assume that $\mathbf{Y}_i^*(\mathbf{t})$ can be written as

$$\mathbf{Y}_i^*(\mathbf{t}) = \mathbf{X}_\mathbf{i}(\mathbf{t})\alpha + \mathbf{V}_\mathbf{i}(\mathbf{t})\mathbf{b}_\mathbf{i} + \mathbf{S}_\mathbf{i}(\mathbf{t}), \quad i = 1, \ldots, N$$
$$\mathbf{b}_\mathbf{i} \sim \mathrm{N}(\mathbf{0}, \mathbf{G}).$$

If, for simplicity, we assume a linear function of time, then $\mathbf{X}_\mathbf{i}(\mathbf{t})$ is a known matrix that consists of a column of ones and a column of time points when the observed CD4 values are recorded, and α is a vector of fixed-effects terms consisting of a population intercept and slope. The matrix $\mathbf{V}_\mathbf{i}(\mathbf{t})$ consists of the identical columns as $\mathbf{X}_\mathbf{i}(\mathbf{t})$, or a subset of the columns of $\mathbf{X}_\mathbf{i}(\mathbf{t})$, $\mathbf{b}_\mathbf{i}$ are random effects independent of $\varepsilon_\mathbf{i}$, and $\mathbf{G}$ is an unstructured covariance matrix. In the examples we consider $\mathbf{b}_\mathbf{i}$ consists of a set of random intercepts or a set of random intercepts and slopes. The term $\mathbf{S}_\mathbf{i}(\mathbf{t})$ represents a Gaussian stochastic process, independent of $\varepsilon_\mathbf{i}$ and $\mathbf{b}_\mathbf{i}$. In the model, $\mathbf{S}_\mathbf{i}(\mathbf{t})$ is designed to capture the biological variation and heterogeneity in the pattern of CD4 trajectories over time seen in individuals. The remaining variability due to other factors such as measurement error, diurnal variation or irrelevant short term variability is reflected in the term $\varepsilon_\mathbf{i}$.

We consider two examples of these kinds of longitudinal models, in one the rate of decline for an individual's CD4 counts remains constant indefinitely, in the other the rate of decline is random and changing constantly within an individual over time. These two special cases are:

Model 1: *Pure random-effects model*

$$\mathbf{Y}_\mathbf{i}^*(\mathbf{t}) = \mathbf{X}_\mathbf{i}(\mathbf{t})\alpha + \mathbf{V}_\mathbf{i}(\mathbf{t})\mathbf{b}_\mathbf{i}, \qquad i = 1, \ldots, N$$
$$\mathbf{b}_\mathbf{i} \sim \mathrm{N}(\mathbf{0}, \mathbf{G})$$

and there is no stochastic process term. This model is the same as that used by Tsiatis *et al.*[7] Here $\mathbf{b}_\mathbf{i}$ is a vector that consists of an individual's intercept and slope. Therefore

$$E(\mathbf{Y}_\mathbf{i}(\mathbf{t})) = \mathbf{X}_\mathbf{i}(\mathbf{t})\alpha.$$
$$\mathbf{\Sigma}_\mathbf{i} = \mathbf{V}_\mathbf{i}(\mathbf{t})\mathbf{G}\mathbf{V}_\mathbf{i}(\mathbf{t})' + \sigma_\varepsilon^2 \mathbf{I}_\mathbf{i}$$

where $\mathbf{\Sigma}_\mathbf{i}$ is the covariance matrix for $\mathbf{Y}_\mathbf{i}(\mathbf{t})$.

Model 2: *Random intercept plus Brownian motion*

$$\mathbf{Y}_\mathbf{i}^*(\mathbf{t}) = \mathbf{X}_\mathbf{i}(\mathbf{t})\alpha + \mathbf{V}_\mathbf{i}(\mathbf{t})\mathbf{b}_\mathbf{i} + \mathbf{BM}_\mathbf{i}(\mathbf{t}), \qquad i = 1, \ldots, N$$
$$\mathbf{b}_\mathbf{i} \sim \mathrm{N}(\mathbf{0}, \mathbf{G})$$
$$\mathbf{BM}_\mathbf{i}(\mathbf{t}) \sim \mathrm{N}_{n_i}(\mathbf{0}, c^2\mathbf{T}_\mathbf{i})$$

where $\mathbf{T}_\mathbf{i}$ is a matrix of visit times and $\mathbf{BM}_\mathbf{i}(\mathbf{t})$ is independent of $\mathbf{b}_\mathbf{i}$ and $\varepsilon_\mathbf{i}$, and $\mathbf{G} = \sigma_{\mathrm{int}}^2$. $\mathbf{T}_\mathbf{i} = (t_{lm})$ where $(t_{lm}) = \min(t_l, t_m)$ where $l = 1, \ldots, n_i$ and $m = 1, \ldots, n_i$ and c^2 is a positive variance

parameter of the process. $\mathbf{X_i(t)}$ consists of both a column of ones and the time points at which the observed CD4 values are recorded, and $\mathbf{V_i(t)}$ merely consists of a column of ones. Thus for this model the pattern for an individual's measurements are described by random intercepts, a linear population drift and scaled Brownian motion to capture the potentially erratic nature of the trajectory. The slope is treated as a fixed-effects term for reasons discussed in Taylor *et al.*[9] These authors found that when they fit a random-effects model with a Brownian motion term, the variance component for the random slope was forced to zero. The Brownian motion term in effect makes each individual's slope a random process over time. Hence, inclusion of a random slope term as well is redundant and the data forces that term to a fixed effect. Since the Brownian motion process is not stationary, there must be a definitive time zero that we take to be the date of randomization.

Therefore, for this model we have

$$E(\mathbf{Y_i(t)}) = \mathbf{X_i(t)}\alpha.$$

$$\mathbf{\Sigma_i} = \sigma^2_{\text{int}}\mathbf{1_i}\mathbf{1_i'} + c^2\mathbf{T_i} + \sigma^2_\varepsilon\mathbf{I_i}$$

where 1_i is a vector of ones. Except for differences in the mean structure and slight differences in the specification of the random effects, model 2 is the same model as used by LaValley and DeGruttola[10] in their two-stage procedure.

2.2. Imputation Using the Multivariate Normal Distribution

We assume for individual i, the vector of true $\text{CD4}^{1/4}$ (fourth root transformation of CD4) values is

$$\mathbf{Y_i^*(t)} = \mathbf{X_i(t)}\alpha + \mathbf{V_i(t)b_i} + \mathbf{S_i(t)}$$

where

$$\mathbf{b_i} \sim \text{N}(\mathbf{0}, \mathbf{G}),$$

$$\mathbf{S_i(t)} \sim \text{N}_{n_i}(\mathbf{0}, \mathbf{\Gamma_i(t)}).$$

Then the observed vector $\mathbf{Y_i(t)}$ is distributed

$$\text{N}_{n_i}(\mathbf{X_i(t)}\alpha, \mathbf{V_i(t)GV_i(t)'} + \mathbf{\Gamma_i(t)} + \sigma^2_\varepsilon\mathbf{I_i}).$$

For individual i, who is at risk at time t, $(Y_i(t_1), \ldots, Y_i(t_j), Y_i^*(t))$ follows a multivariate normal distribution, and we wish to impute a CD4 value at time t. It can be shown that the mean of the true covariate, $Y_i^*(t)$, given the observed covariates, $\mathbf{Y}_i$, is

$$\mu(t) + \mathbf{\Sigma_{21}\Sigma_{11}^{-1}}(\mathbf{Y_i} - \mu_\mathbf{Y}) \tag{1}$$

where $\mu(t) = E(Y_i^*(t))$. If we let $C_{t,t_m} = \text{cov}(Y_i^*(t), Y_i(t_m))$ for $m = 1, \ldots, j$, then

$$\mathbf{\Sigma_{21}} = [C_{t,t_1}, C_{t,t_2}, \ldots, C_{t,t_j}]$$

and

$$\boldsymbol{\Sigma}_{\mathbf{11}}^{-\mathbf{1}} = \begin{bmatrix} C_{t_1,t_1} + \sigma_\varepsilon^2 & C_{t_1,t_2} & \cdots & C_{t_1,t_j} \\ C_{t_2,t_1} & C_{t_2,t_2} + \sigma_\varepsilon^2 & \cdots & C_{t_2,t_j} \\ \vdots & \vdots & \ddots & \vdots \\ C_{t_j,t_1} & C_{t_j,t_2} & \cdots & C_{t_j,t_j} + \sigma_\varepsilon^2 \end{bmatrix}^{-1}$$

where C_{t_l,t_m} equals $\text{cov}(Y_i^*(t_l), Y_i^*(t_m))$ for $l, m = 1, \ldots, j$. The means are given by

$$\mu_{\mathbf{Y}} = [\alpha_0 + \alpha_1 t_1 \quad \alpha_0 + \alpha_1 t_2 \cdots \quad \alpha_0 + \alpha_1 t_j]'$$

and

$$\mu(t) = \alpha_0 + \alpha_1 t$$

where α_0 and α_1 are the fixed-effects estimates of the population mean intercept and slope, respectively.

For model 1 (pure random-effects model), C_{t_l,t_m} would include a variance term for the slope as well as the covariance term between the slope and intercept while the covariance term for the stochastic process would be zero.

For model 2, for individual i, C_{t_l,t_m} is given by

$$\text{cov}(b_i + s_i(t_l), b_i + s_i(t_m)) = \text{var}(b_i) + \text{cov}(s_i(t_l), s_i(t_m)).$$

Using the above concepts, we can now impute a value for each member in the risk set using equation (1). However, we must first estimate the parameters of the above model. As suggested by Tsiatis *et al.*,[7] we can condition on past marker values at each failure time for those individuals still in the risk set (that is, t_j in our above notation would be the last observed time prior to t, the target time for imputation), and estimate the above parameters. We can extend this by conditioning on any window of data we choose, since in an applied setting one often has such information. We hypothesize that by doing this, we can better fit the data with our stochastic models and obtain a better prediction of the CD4 count at the time of failure. Further, since we are assuming that the stochastic processes follow multivariate normal distributions for each individual, then, conditional on any subset of these marker values, the distribution still follows a multivariate normal with its appropriate mean and variance structure.

In our application to clinical trials we fit the mixed models used to impute CD4 values at each failure time separately for the treatment and placebo arms. We calculate a covariate value at time t for individual i from equation (1), and t_j can be larger than t. Equation (1) generalizes in the obvious way if $\mathbf{X_i(t)}\alpha$ is a more complex function than linear in time. We use the Fisher scoring algorithm, as defined by Jennrich and Schluchter[11] to estimate the parameters in the above models.

2.3. Future versus Historical

When we use the words 'historical' and 'future', we are referring to the time window of observed CD4 counts used in estimating the parameters in the longitudinal models and the time window of data used in generating the imputations of current predicted true CD4 counts for each subject in the risk set. Two numbers ω_1 and ω_2 define these two time windows. Let t be the target time at which we wish to impute CD4 values. For estimation of the longitudinal model parameters, we

use all observed CD4 values at times $\leqslant t + \omega_1$ where $0 \leqslant \omega_1 < \infty$ and ω_1 is the size of the window beyond time t used to estimate the parameters. When we generate our imputes for each subject in the risk set, we condition on that subject's observed CD4 values recorded at all times $\leqslant t + \omega_2$ where $0 \leqslant \omega_2 < \infty$. For the 'historical' approach we use $(\omega_1, \omega_2) = (0, 0)$, which was the procedure adopted by Tsiatis *et al.*,[7] and for the 'future' approach we use $(\omega_1, \omega_2) = (16$ weeks, 16 weeks) for the ACTG-019 trial and (12 weeks, 12 weeks) for the simulation study. The choices of ω_1 and ω_2 in the future models effectively allow for use of one additional observed CD4 value in estimating the longitudinal model parameters and for generating the imputations for all subjects in the risk set who have an observed visit within the time frame. We note that ω_1 and ω_2 do not have to be equal; we will present later an analysis of the ACTG 019 data in which $\omega_1 = \infty$ and $\omega_2 = 0$. The motivation for using future data in the estimation stage ($\omega_1 > 0$) is to give more precise estimates of the longitudinal model parameters, since they are based on more data, and in the imputation stage is to give less variable estimates of an individual's CD4 count since the procedure uses extra information on that individual from a proximate time. The drawback of using future data is the greater potential for bias which needs to be balanced against the reduced variability.

3. SIMULATIONS

3.1. Design of the Study

The aim in this simulation study is to evaluate and compare the statistical properties of the estimate of the relative risk parameter, β, in the Cox model, when the various smoothing techniques are used. We study the impact of varying amounts of measurement error on the different smoothing methods. We investigate what cost we pay in terms of lost efficiency and increased bias in our risk parameter estimate when we mismodel the true underlying CD4 process in stage 1 of our approach. Also, we determine if use of a window of future data improves our predictions of the true CD4 values, and leads to a better estimate of the risk parameter. In the simulation design, we have both a treatment and a placebo group to mimic the situation in ACTG-019. Although our primary focus is on β, we also comment on the relative hazard between the two treatment arms.

In the simulations, we evaluate each of the smoothing techniques in terms of MSE, actual coverage in a nominal 95 per cent confidence interval, average bias in the relative risk parameter estimate, and average of the Wald statistic estimate. The Wald statistic gives a measure of the power to test the significance of β.

We simulate data for $CD4^{1/4}$ under two different longitudinal models: the pure random-effects and Brownian motion models of Section 2, and two values of $\sigma_\varepsilon = (0{\cdot}1, 0{\cdot}3)$. The value of 0·3 arises from the ACTG-019 trial, and represents a fairly noisy data set. We chose the value 0·1 because it represents an ideal situation where measurement error is kept to a minimum. We choose the following model, also used by Tsiatis *et al.*,[7] with a change point at eight weeks for the CD4 process over time in the treatment arms:

$$Y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \alpha_2 (t_{ij} - 0{\cdot}15332)^+ + b_{0i} + b_{1i} t_{ij} + s_{ij} + \varepsilon_{ij}$$

where Y_{ij} is the observed $CD4^{1/4}$ on individual i at time j, and t_{ij} is measured in years. In this model α_0, α_1 and α_2 are the intercept, slope prior to 8 weeks and the change in slope, respectively. The terms b_{0i} and b_{1i} are random intercepts and slopes and s_{ij} is a stochastic process. In all of the

analyses presented, we use $CD4^{1/4}$ because it has roughly a Gaussian distribution, which is the underlying assumption of the mixed model, and it stabilizes within-subject variance.[12] The term $(t_{ij} - 0{\cdot}15332)^+$ is defined to be the maximum of $(t_{ij} - 0{\cdot}15332)$ and 0. The model for the placebo arm is

$$Y_{ij} = \alpha_0 + \alpha_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + s_{ij} + \varepsilon_{ij}.$$

When simulating data under the random-effects model, α is $(4{\cdot}2638, \ -0{\cdot}0226)^{\mathrm{T}}$ for the placebo arm and α is $(4{\cdot}2143, \ 1{\cdot}1632, \ -1{\cdot}2293)^{\mathrm{T}}$ for the treatment arm. The components of $\mathbf{G}$ are $\mathrm{var}(b_{0i}) = 0{\cdot}1320$, $\mathrm{cov}(b_{0i}, b_{1i}) = 0{\cdot}0287$ and $\mathrm{var}(b_{1i}) = 0{\cdot}1098$ for the placebo arm and $\mathrm{var}(b_{0i}) = 0{\cdot}1122$, $\mathrm{cov}(b_{0i}, b_{1i}) = 0{\cdot}0491$ and $\mathrm{var}(b_{1i}) = 0{\cdot}1333$ for the treatment arm. The fixed-effects vector for the Brownian motion model is $(4{\cdot}2520, \ -0{\cdot}0084)^{\mathrm{T}}$ for the placebo arm and $(4{\cdot}2143, 1{\cdot}1476, \ -1{\cdot}2097)^{\mathrm{T}}$ for the treatment arm. The covariance parameters to simulate the data for the Brownian motion model are $\sigma^2_{\mathrm{int}} = 0{\cdot}1347$ for the placebo arm and $\sigma^2_{\mathrm{int}} = 0{\cdot}1224$ for the treatment arm and $c^2 = 0{\cdot}15$.

We simulated failure times with AIDS under a proportional hazards model

$$\lambda_0(t) \mathrm{e}^{\beta Y_i^*(t) + \gamma A_i}$$

where $\lambda_0(t) = 2{\cdot}5$ and the risk parameter for $CD4^{1/4}$, $\beta = -2$, and for treatment (A_i), $\gamma = -1$. We fit each imputation model separately for each treatment arm to each simulated longitudinal data set and we impute the missing covariates in each risk set. We then analyse each of the imputed data sets using a Cox proportional hazards model, $\lambda_0(t) \mathrm{e}^{\beta \hat{Y}_i^*(t) + \gamma A_i}$, where $\hat{Y}_i^*(t)$ is the imputed value and determine the effect that the imputation schemes and assumed models have on the estimate of β and γ. We run 200 Monte Carlo simulations, and for each simulation, we consider 300 subjects randomized to two treatment arms with between 2 and 10 visits spaced at 3 month intervals with the first visit at $t = 0$.

The simulated CD4 data are loosely based on the structure of the ACTG-019 trial. Hence, parameter values for the mixed models used to generate the CD4 data are those obtained from the appropriate fits to the ACTG-019 data. Also, the parameters used for the hazard of progression to AIDS are those derived from the ACTG-019 data. The rationale for mimicking the ACTG-019 trial in our simulations is to make our study more comparable to the real data, and to help enhance our insight into observations of a statistical nature made from these trial data. The computations are carried out using SAS IML[13] and SAS PROC PHREG.[14]

3.2. Results

Table I gives the results when the model generating the CD4 counts is the pure random-effects model (model 1 in Section 2), and Table II gives the results when we use the Brownian motion model (model 2 in Section 2) to generate the CD4 counts.

3.2.1. Bias

In Tables I and II, we see for $\sigma_\varepsilon = 0{\cdot}3$ that all risk parameter estimates are biased downward on average compared to the TRUE (true value of $CD4^{1/4}$ measured without error) except for the HBM (historical Brownian motion) model. Conversely, the FBM (future Brownian motion) approach gives a worse fit, in general, compared to the other smoothing methods. When using the nearest preceding CD4 value (OBS), the average relative risk parameter estimate we obtain is greatly biased towards zero. In Table I, both the HRE (historical random effects) and FRE (future

Table I. Simulation results for data generated from the pure random-effects model. Summary statistics for β (true value: $-2{\cdot}0$)

Model	Estimate	Standard deviation	MSE	Coverage	Wald
TRUE	$-2{\cdot}04$	0·25	0·067	0·95	63·8
$\sigma_\varepsilon = 0{\cdot}3$					
OBS	$-1{\cdot}58$	0·20	0·219	0·545	49·9
IRL	$-1{\cdot}83$	0·24	0·087	0·865	58·0
HRE	$-1{\cdot}93$	0·25	0·069	0·935	56·2
FRE	$-1{\cdot}93$	0·25	0·068	0·945	54·9
HBM	$-2{\cdot}05$	0·27	0·072	0·945	55·8
FBM	$-1{\cdot}85$	0·24	0·081	0·905	49·5
$\sigma_\varepsilon = 0{\cdot}1$					
OBS	$-2{\cdot}03$	0·25	0·064	0·965	60·8
IRL	$-1{\cdot}96$	0·25	0·065	0·96	60·0
HRE	$-2{\cdot}02$	0·25	0·066	0·965	62·5
FRE	$-2{\cdot}01$	0·25	0·064	0·97	62·0
HBM	$-2{\cdot}10$	0·27	0·080	0·95	62·0
FBM	$-1{\cdot}89$	0·24	0·067	0·94	56·0

Table II. Simulation results for data generated from the Brownian motion model. Summary statistics for β (true value: $-2{\cdot}0$)

Model	Estimate	Standard deviation	MSE	Coverage	Wald
TRUE	$-2{\cdot}01$	0·21	0·043	0·965	82·2
$\sigma_\varepsilon = 0{\cdot}3$					
OBS	$-1{\cdot}61$	0·18	0·182	0·50	64·9
IRL	$-1{\cdot}90$	0·23	0·060	0·89	79·9
HRE	$-1{\cdot}85$	0·24	0·076	0·895	68·1
FRE	$-1{\cdot}89$	0·23	0·062	0·925	66·7
HBM	$-2{\cdot}00$	0·24	0·059	0·95	68·9
FBM	$-1{\cdot}79$	0·20	0·086	0·86	60·6
$\sigma_\varepsilon = 0{\cdot}1$					
OBS	$-1{\cdot}98$	0·21	0·044	0·955	75·4
IRL	$-2{\cdot}01$	0·23	0·049	0·935	80·8
HRE	$-1{\cdot}93$	0·23	0·056	0·915	74·6
FRE	$-1{\cdot}98$	0·23	0·049	0·935	75·5
HBM	$-2{\cdot}06$	0·23	0·054	0·96	76·3
FBM	$-1{\cdot}86$	0·20	0·061	0·89	69·2

random effects) models perform fairly well in terms of a small bias in the risk parameter estimate since they correctly specify the true underlying CD4 process. Neither method removed all of the bias, however. As Dafni[15] pointed out, most of the residual bias remaining is probably due to the fact that subjects with more negative slopes have a tendency to fail earlier and more often. Therefore, the distribution of the random effects becomes more skewed with time and the assumption of normality is most likely violated. There is some residual bias due to measurement error not accounted for by the smoothing process and that is reflected in the downward bias as

well. In Table II, both the FRE and HRE models perform reasonably well even though they misspecify the underlying stochastic process. Both models underestimate the true risk parameter, but offer a substantial improvement over use of the nearest preceding neighbour.

The IRL (individual regression lines) approach fits a straight line (intercept and slope) through each subject's observed $CD4^{1/4}$ counts. The IRL model also does a reasonable job in terms of reducing the bias in the risk parameter estimate, and has the added advantage that it is much easier to implement than the methods based on random-effects modelling. However, this simulation study is an overly optimistic setting for the IRL approach since there are no missed visits. As the amount of missed visits increases, the IRL method performs increasingly poorly. Also, this method is much more susceptible to outliers and unrealistic predictions when extrapolating outside the range of the data used to estimate the parameters of the model. A main reason that some residual bias remains in the IRL approach is that the model ignores the change in slope at eight weeks for subjects in the treatment arm. When $\sigma_\varepsilon = 0{\cdot}1$, the difference in terms of bias in the risk parameter estimate is much more negligible among the various methods, including use of just the nearest preceding CD4 value.

Across all of these particular simulation scenarios, measurement error has no appreciable effect on the estimate of γ (results not shown). In all cases, $\hat{\gamma}$ is close to -1 regardless of the method used to estimate the current CD4 value, and actual coverage rates are around the nominal 0·95 level. In other work,[16] we found evidence that measurement error in the time-dependent covariate can induce bias in the estimate of γ.

3.2.2. *MSE and Standard Deviation*

For $\sigma_\varepsilon = 0{\cdot}3$ in Tables I and II, the average MSE is the smallest for the TRUE model. However, with the exception of the OBS model, all of the MSEs are reasonably close to the TRUE in magnitude. When $\sigma_\varepsilon = 0{\cdot}1$, we see that all methods give small MSE estimates. It is clear that the bias in the risk parameter estimate is the dominating factor in the MSE. Also in both tables the OBS and future smoothing approaches result in somewhat smaller standard deviations than do the historical approaches.

3.2.3. *Coverage Rates for Confidence Intervals*

In Table I, when $\sigma_\varepsilon = 0{\cdot}3$, all the smoothing techniques have less than the 95 per cent nominal coverage level. The HRE and FRE methods actually are very close to correct since they represent a correctly specified model for the underlying CD4 process. The HBM approach also does well in terms of correct coverage. For Table II, only the FRE and HBM approaches have coverage rates near the nominal level of 95 per cent. The major reason for the lower coverage rates for the other methods is the bias remaining in the parameter estimates. It is worth noting that use of the nearest preceding CD4 value gives very poor coverage of the true risk parameter in a nominal 95 per cent confidence interval of 54·5 per cent in Table I and only 50 per cent in Table II. When $\sigma_\varepsilon = 0{\cdot}1$, all of the methods have an actual coverage level fairly close to the nominal level of 95 per cent. Overall, it appears that only the HBM and FRE approaches consistently give adequate coverage rates across the four simulation scenarios.

3.2.4. *Wald Statistics*

The Wald test statistic for all the smoothing methods in both tables is smaller on average than the value obtained from the TRUE data, especially when $\sigma_\varepsilon = 0{\cdot}3$. This implies that we are less likely

Table III. Mean residual fit and estimated variance for pure random-effects model with $\sigma_\varepsilon = 0{\cdot}3$

Model	Mean (r)	SE (r)	Mean (s^2)	SE (s^2)
OBS	−0·0023	0·0005	0·0919	0·0002
IRL	0·0057	0·0004	0·0172	0·0001
HRE	0·0002	0·0007	0·0408	0·0002
FRE	−0·0007	0·0005	0·0257	0·0001
HBM	−0·0005	0·0006	0·0436	0·0002
FBM	−0·0013	0·0005	0·0284	0·0001

to reject the null hypothesis $\beta = 0$ when in fact the alternative is true. The smoothing methods, with the exception of the FBM approach, give larger average Wald statistics than does use of simply the nearest preceding CD4 value. For $\sigma_\varepsilon = 0{\cdot}1$, the smoothing methods on average give Wald test statistics much closer in size to that obtained from the TRUE approach.

4. DIAGNOSTICS

To evaluate the fit of our longitudinal models to the true CD4 counts and to better understand the results of Tables I and II, we compare the predicted CD4 counts with the true (unobserved) values and we consider both the bias of the predictions and their variability. We examine the case where we generated the true CD4 from a pure random-effects model with $\sigma_\varepsilon = 0{\cdot}3$.

We consider the residual fit for each observation across all subjects

$$r = \sum_{i=1}^{N} \sum_{j=1}^{m_i} \frac{(Y_{ij}^* - \hat{Y}_{ij}^*)}{\sum_{i=1}^{N} m_i}.$$

As before, Y_{ij}^* denotes the true $\mathrm{CD4}^{1/4}$ value for individual i at time j, and $\hat{Y}_{ij}^*$ is the predicted true $\mathrm{CD4}^{1/4}$ value for individual i at time j, where m_i is the number of risk sets for subject i.

We also consider the statistic

$$s^2 = \sum_{i=1}^{N} \sum_{j=1}^{m_i} \frac{(Y_{ij}^* - \hat{Y}_{ij}^*)^2}{\sum_{i=1}^{N} m_i}.$$

If $\hat{Y}_{ij}^*$ is just observed $\mathrm{CD4}^{1/4}$ count, s^2 is an estimate of σ_ε^2. Table III presents the average of these statistics and their standard errors across all 200 simulated data sets.

In Table III, the OBS and FBM approaches tend to predict values of CD4 that are higher than the true value at the time of failure while the IRL approach tends to predict low on average. For the OBS, IRL and FBM approaches, zero is not within 2 standard errors of the mean of r. A reason for the overestimation of the true CD4 count by the OBS approach is that a pulled forward preceding observation tends to overestimate the true value because subjects on average follow a decreasing slope with time.

The underestimation of the true CD4 counts on average by the IRL approach is due to the lack of accounting for the change point at eight weeks in the treatment arm. When we look at the average of r across the 200 simulations by treatment arm, the underestimation vanishes for the placebo arm, but remains for the treatment arm.

The fact that the FBM model is predicting high on average, as seen in Table III, causes the risk parameter to be biased downwards towards the null. This is why the FBM approach does not perform as well as the other smoothing techniques in terms of bias. What we believe is causing the FBM approach to predict high is the informative right censoring and the fact that the estimates of the fixed-effects parameters for each risk set are biased because they are being estimated from a sample of subjects who tend to be healthier because they do not drop out and have a flatter rate of decline in their CD4 counts. This is causing the FBM approach to predict CD4 values that are higher on average than the unobserved true CD4 value at the time of failure. This overprediction is largest for those subjects who have the event at the risk set time.

For the OBS approach in Table III, we see that the mean of s^2 is roughly equal to 0·09 which we expect since we simulated the data with $\sigma_\varepsilon^2 = 0{\cdot}09$. All of the smoothing techniques give substantially smaller values of s^2, which is the objective of applying these methods. The FRE, FBM and IRL approaches give smaller average estimates of s^2 than do the historical approaches. This is due to the fact that the future approaches make use of more data in generating predictions which reduces the variability of these predictions. However, these methods have higher bias than the historical approaches. The IRL approach has the smallest estimate of s^2 on average because it uses the entire set of data for each subject, but also entertains the largest bias in r.

An interesting observation in Table III, is that bias in predicting the true CD4 value at the time of failure appears to have a bigger effect on how well the smoothing methods perform in Table I. The FBM and IRL methods have large means for r, but small values for s^2, and they perform poorly in Table I compared to the other approaches. In contrast, the HRE and HBM approaches, which have larger values of s^2 but substantially smaller mean values of r in Table III, do much better in Table I in terms of smaller bias, smaller MSE estimates, and better actual coverage rates.

5. ACTG-019 TRIAL

In the ACTG-019 trial, the protocol indicated observation of patients' CD4 values at baseline, 8, 16, 32, 48, 64 and 80 weeks. There were 428 individuals in the placebo arm, 453 in the low-dose arm, and 456 in the high-dose arm. For the purposes of all analyses, we combined the low-dose and high-dose arms, since there was no significant difference in the survival curves for time to AIDS between the two arms. The average follow-up overall was 50 weeks with a maximum of about 108 weeks. The number of individuals who progressed to AIDS in the placebo arm was 33, while 25 progressed in the treatment arm (p-value = 0·0021 by the logrank test). Figure 1 shows the Kaplan–Meier curves for failure with AIDS.

Figure 2 shows the median CD4 count by treatment arm over time. From this figure we see the effect of AZT is initially to increase sharply the median CD4 count in the treatment arm, and then the effect dissipates. However, the treatment arm consistently maintains a higher level of CD4 than does the placebo arm over time, as seen in the figure. This pattern has been consistent in many trials of AZT versus placebo. The apparent increasing trend of CD4 over time is most likely due to the attrition of less healthy individuals from the trial which causes an inflated estimate of the average CD4 count as time progresses.

5.1. Results

In the two-stage procedure for fitting the Cox models, we fit all mixed models separately to each treatment arm. Of the 58 failures that occurred in the ACTG-019 clinical trial, seven occurred

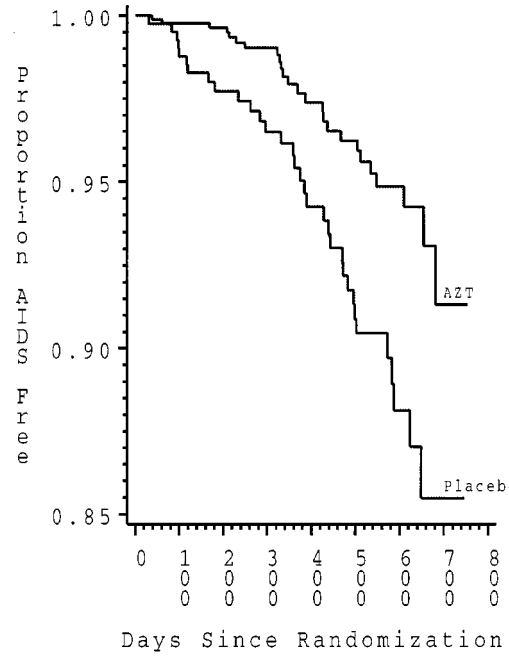

Figure 1. Product-limit estimate of AIDS free survival measured from date of randomization

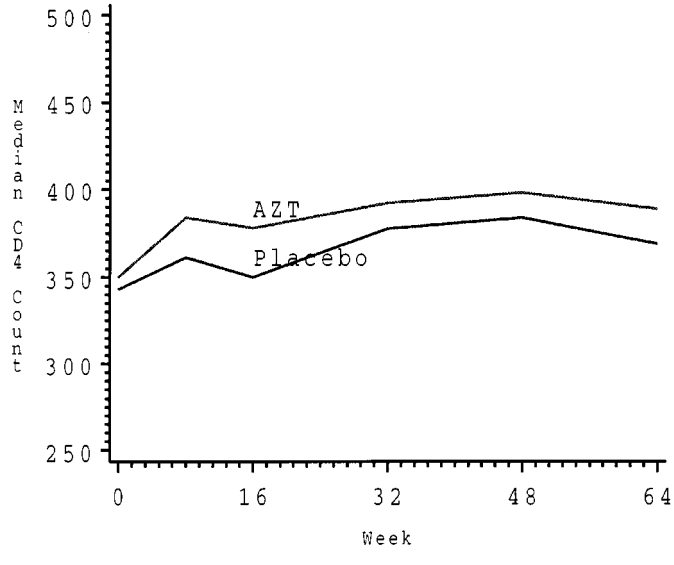

Figure 2. Change in median CD4 count by treatment arm

Table IV. Estimated relative risk parameters and standard errors for $CD4^{1/4}$ (β) and treatment (γ) under various smoothing approaches

Model	Parameter	Standard errors	Coeff./SE
OBS	−2·03	0·18	−11·28
	−0·81	0·27	−3·00
IRL	−0·77	0·09	−8·56
	−0·61	0·27	−2·26
HRE	−2·13	0·20	−10·65
	−1·00	0·27	−3·70
FRE	−2·27	0·21	−10·81
	−0·97	0·27	−3·59
HBM	−2·39	0·20	−11·95
	−0·90	0·27	−3·33
FBM	−2·43	0·22	−11·05
	−0·88	0·27	−3·26

prior to the 16 week follow-up visit. Therefore, estimation of the mixed model parameters for the first seven risk sets is impossible for the HRE and HBM models. Hence, for the first seven failures, we use the baseline covariate measurement for all individuals in a given risk set for these two methods. For the IRL method, if an individual has only one measurement, then we pull that forward for all eligible risk sets, since estimation of the parameters of the regression model is not possible. Table IV displays the fit of the models $\lambda_0(t)e^{\beta CD4^{1/4}(t)+\gamma A_i}$ to the ACTG-019 data. For each approach, the first row is the risk parameter estimate and summary statistics for $CD4^{1/4}$, and the second row is for the treatment parameter.

In Table IV, excluding IRL, we see the estimate for the relative risk parameter is the smallest when we use the most recently observed $CD4^{1/4}$ value as the prediction for a subject's true value at each failure time. The FRE, FBM, HRE and HBM models lead to larger estimates of the relative risk term in absolute value than does the observed model. This agrees with the results of Raboud[5] and the fact that the parameter estimate of the relative risk term for the observed data is biased towards the null when we measure the CD4 counts with error. The different results of the IRL model are most likely due to the fact that several individuals have only one or two post baseline CD4 measurements. Thus the regression coefficient estimates are not very reliable, and often extrapolation well outside the range of the data can lead to very unreasonable predictions. The population based smoothing techniques can circumvent this problem by borrowing strength across the population when imputing CD4 values. We see for all models that the $CD4^{1/4}$ count is very predictive of the time until AIDS. In Table IV, we see that treatment is a strong predictor of progression to AIDS, with $\hat{\gamma}$ ranging between $-0{\cdot}6$ and $-1{\cdot}0$ depending on the smoothing method.

For comparison purposes, we also looked at estimation of the mixed model parameters one time from the whole data set, with imputations then generated by using this same set of parameter estimates for each risk set. The reason for doing this alternative approach is to deal with the fact that the first seven failures with AIDS occurred prior to the third visit, thus making estimation of the random intercept, slope, error and Brownian motion variance parameters within each of these risk sets impossible. Table V presents the results.

Table V. Estimated relative risk parameters and standard errors for $CD4^{1/4}$ and treatment after using global estimation of mixed model parameters

Model	Parameter	Standard error	Coeff./SE
HRE	−2·25	0·21	−10·71
	−0·96	0·27	−3·56
HBM	−2·55	0·23	−11·09
	−0·90	0·27	−3·33

Both models in Table V give larger estimates of the relative risk parameter than the one obtained from the OBS model. This again agrees with the results of Raboud.[5] This approach also has the advantage that it is much less computer intensive since we estimate the mixed model parameters only once as opposed to estimation at each event time.

6. DISCUSSION

With a covariate measured repeatedly over time with error, one can apply smoothing techniques to reduce the variability of the data. We use these smoothing techniques in the first stage of our two-stage procedure. We considered two basic kinds of smoothing methods: population based and individual. Population based smoothing techniques use random-effects models plus possibly a stochastic term to generate imputes for each subject in the risk set at the time of failure. Individual techniques involve fitting regression lines or performing some other type of averaging of each unique individual's data to obtain imputes at each failure time. Once we generate our imputes at each event time, we then use them in the second stage in a time-dependent Cox model, which we saw led to better inferences for the risk parameter, β.

When all subjects have most of their data observed over time, individual methods of smoothing do a very good job of reducing the variability found in the original measurements, and are less time consuming and computer intensive than are population based methods. Conversely, when here is a lot of missing data, we recommend use of population based methods because they can better handle subjects with sparse data by smoothing their predictions towards the population mean. One way to reduce the computational burden of re-estimating random-effects models at each failure time is to estimate the mixed model parameters once by fitting to the entire data set, and to use these estimates repeatedly across risk sets. We found that this resulted in a substantial reduction in computational time and provided a good fit in the Cox model when we applied it to the ACTG-019 data.

We also looked at use of a window of future data to help estimate our mixed model parameters, and to generate our imputes at each failure time. Overall, we found that use of future data did not, in general, result in substantial improvements for inferences based on β in a time-dependent Cox model compared to use of only a historical window, and furthermore, they could induce some bias. We recommend use of a historical window except when longitudinal data are sparse, such as for early failures. Then a window of future data may allow for a more stable estimate of the mixed model parameters. Further, we found that if we misspecify the covariance structure of the true CD4 process in our imputation models, this does not greatly affect our results.

A primary objective of applying the smoothing techniques is to reduce the bias in the risk parameter estimate. We achieve this reduction by performing single imputation of the conditional expectation of the covariate at each failure time for all subjects in the risk set. A further refinement, which recognizes the uncertainty in the conditional expectations, is multiple imputation[17] with the associated methods for combining estimates. We might expect such a procedure to more accurately estimate the variance of the risk parameter.

There is the potential for the missingness pattern of the covariate due to informative dropout to bias the estimation of the parameters in the longitudinal model. This might be due to the more seriously ill patients dying or to healthy subjects dropping out. This may lead to bias in the imputes which, as we have seen in this article, leads to bias in the estimation of the relative risk parameter. One possible way to handle this, although computationally burdensome, is joint modelling of the longitudinal data, the failure time data and the dropout mechanism. This is a topic of current research interest.[18,19]

We found that reducing the bias in our predicted true CD4 values appeared to have greater importance in terms of obtaining an accurate estimate of the risk parameter in the Cox model than did reducing the amount of variability in our predictions. In the simulation study, we did not investigate the impact of using smoothing methods which mismodel the mean structure of the longitudinal marker process. We would expect such mismodelling to lead to biased estimates of the risk parameter, and inappropriate confidence intervals. In summary, the research presented in this paper shows that application of smoothing approaches in the first stage of a two-stage procedure for fitting a time-dependent Cox model leads to improved statistical properties of the risk parameter estimate.

ACKNOWLEDGEMENTS

This work was partially supported by grants AI07370, CA64235 and AI29196 from the National Institutes of Health.

REFERENCES

1. Cox, D. R. 'Regression models and life tables (with discussion)', *Journal of the Royal Statistical Society, Series B*, **34**, 187–220 (1972).
2. Cox, D. R. 'Partial likelihood', *Biometrika*, **62**, 269–276 (1975).
3. Altman, D. G. and De Stavola, B. L. 'Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates', *Statistics in Medicine*, **13**, 301–341 (1994).
4. Gail, M. 'Evaluating serial cancer marker studies in patients at risk of recurrent disease', *Biometrics*, **37**, 67–78 (1981).
5. Raboud, J., Reid, N., Coates, R. A. and Farewell, V. T. 'Estimating risks of progressing to AIDS when covariates are measured with error', *Journal of the Royal Statistical Society, Series A*, **156**, 393–406 (1993).
6. Prentice, R. L. 'Covariate measurement errors and parameter estimation in a failure time regression model', *Biometrika*, **69**, 331–342 (1982).
7. Tsiatis, A. A., DeGruttola, V. and Wulfsohn, M. S. 'Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with Aids', *Journal of the American Statistical Association*, **429**, 27–37 (1995).
8. Volberding, P. A., Lagakos, S. W., Koch, M. A. *et al.*, 'Zidovidine in asymptomatic human immunodeficiency virus infection', *New England Journal of Medicine*, **322**, 941–949 (1990).
9. Taylor, J. M. G., Cumberland, W. G. and Sy, J. P. 'A stochastic model for analysis of longitudinal AIDS data', *Journal of the American Statistical Association*, **89**, 727–736 (1994).
10. LaValley, M. P. and DeGruttola, V. 'Models for empirical Bayes estimators of longitudinal CD4 counts', *Statistics in Medicine*, **15**, 2289–2305 (1996).

11. Jennrich, R. I. and Schluchter, M. 'Unbalanced repeated-measures with structured covariance matrices', *Biometrics*, **42**, 805–820 (1986).
12. Taylor, J. M. G., Fahey, J. L., Detels, R. and Giorgi, J. V. 'CD4 percentage, CD4 number, and CD4:CD8 ratio in HIV infection: which to choose and how to use', *Journal of Acquired Immune Deficiency Syndromes*, **2**, 114–124 (1989).
13. SAS. *SAS/IML Software*, Version 6, First Edition, SAS Institute Inc., Cary NC, 1990.
14. SAS. 'The PHREG Procedure', in SAS/STAT *Software*: *Changes and Enhancements through Release 6.12*, SAS Institute Inc., Cary NC, 1997, Chapter 24.
15. Dafni, U. 'Evaluating surrogate markers of clinical outcomes when measured with error', Doctoral Dissertation, Harvard School of Public Health, 1993.
16. Bycott, P. W. 'A comparison of smoothing techniques for a covariate measured with error in a time-dependent Cox proportional hazards model', Doctoral Thesis, Department of Biostatistics, UCLA, 1997.
17. Rubin, D. B. *Multiple Imputation for Nonresponse in Surveys*, Wiley, 1987.
18. Faucett, C. L. and Thomas, D. C. 'Simultaneously modeling censored survival data and repeatedly measured covariates: a Gibbs sampling approach', *Statistics in Medicine*, **15**, 1663–1685 (1996).
19. Wulfsohn, M. S. and Tsiatis, A. A. 'A joint model for survival and longitudinal data measured with error', *Biometrics*, **53**, 330–339 (1997).

