A particular diffusion model for incomplete longitudinal data: application to the multicenter AIDS cohort study

CYNTHA A. STRUTHERS*, DONALD L. MCLEISH

Department of Statistics and Actuarial Science, University of Waterloo,
Waterloo, Ontario N2L 3G1, Canada
castruth@uwaterloo.ca

SUMMARY

Longitudinal studies, in which individuals are measured repeatedly in time, are often incomplete. We model continuous-time longitudinal data from the Multicenter AIDS Cohort Study using a diffusion model in which the diffusion parameters are functions of the covariates. These data are jointly modeled with the process of time-to-death due to AIDS. We show that, even for large data sets with a large number of missing variables, a Bayesian analysis is feasible using Gibbs sampling and compare a complete case analysis with a Bayesian treatment of missing values.

Keywords: Diffusion model; Gibbs sampling; Longitudinal models; Missing data; Multicenter AIDS Cohort Study; Ornstein–Uhlenbeck process; Time-to-event data.

1. INTRODUCTION

In longitudinal studies, individuals are measured repeatedly, often at irregular discrete time intervals (see Diggle and others, 2005, for a general discussion). An example that motivated our work is the analysis of the Multicenter AIDS Cohort Study (MACS) covering the years 1984–1990. The data include 2376 repeated measurements on CD4 count (an immune cell that fights infectious agents) and a number of time-varying covariates on 369 HIV-infected individuals. The objectives of the analysis of Diggle and others (2005) were to (1) estimate the average time of CD4 cell depletion, (2) estimate the time course for individuals taking account of the measurement error, and (3) characterize the degree of heterogeneity across men in the rate of progression. Typically, CD4 count is roughly constant until seroconversion occurs, and then decreases, quickly at first. Diggle and others (2005) fit a regression model of the form

\[ Y(t) = \mu(t) + \beta x + \varepsilon, \quad (1.1) \]

where \( Y(t) \) is the square root of the CD4 count at time \( t \), \( \mu(t) \) is assumed constant prior to seroconversion (at \( t = 0 \)) and a quadratic function thereafter, \( \beta \) is a vector of unknown parameters, and \( x \) is a vector of covariates. The zero-mean random variable \( \varepsilon \) is assumed to have variance/covariance structure with 3 components corresponding to random effects (specific to a given subject), serial correlation between

*To whom correspondence should be addressed.
the measurements on a given individual at different times, and measurement error (specific to a given individual time point).

Taylor and others (1994) proposed a model in which \( Y(t) \), the fourth root of the CD4 count at time \( t \), satisfies \( Y(t) = a + bt + \beta X(t) + W(t) + e(t) \), where \( e(t) \) is the measurement error and \( W(t) \) is an “integrated” Ornstein–Uhlenbeck process. They applied this model to data including only CD4 counts of seroconverters, with the date of seroconversion approximated by the midpoint of a seroconversion interval known to be of length less than 15 months and before drug therapies were begun. Their covariate \( X(t) \) was time independent and only influenced the initial level of the process not its rate of change over time. The intercept \( a \) was assumed normally distributed. They concluded that the best fit was obtained when \( W(t) \) was a Brownian motion, that is, \( dY(t) = b \, dt + dW(t) \), where \( b \) is the rate of change of \( Y(t) \).

We assume that measurements on a given individual are obtained by measuring the values of a continuous-time stochastic process, subject to measurement error, with observations only at specific times. There is a number of underlying time-dependent covariates. The use of diffusion models is not uncommon in longitudinal studies (see Taylor and others, 1994; Sy and others, 1997; Boscardin and others, 1998; Zhang and others, 1998) but there is a material and a philosophical difference between our approach and the current literature modeling the response using regression. We suggest that a more natural model is a stochastic process for which the covariates are embedded in the time-dependent parameter \( \theta(t) \) of the stochastic process. We fit a special case of a diffusion in which a power of the CD4 counts follows a generalized Ornstein–Uhlenbeck model with parameter \( \theta(t) \). In this case, a covariate change can induce a shock that alters the drift of the process and dissipates slowly over time. Because the parameter \( \theta(t) \) is time dependent, our model is not equivalent to one of the regression models in the literature. In addition to the objectives (1)–(3) of Diggle and others (2005), we add a further objective: (4) to characterize the relationship between CD4 count and the progression to AIDS or death. For an overview on joint models for longitudinal data and time-to-event data see Tsiatis and Davidian, 2004.

We also provide a more complete treatment of the substantial amount of missing or incomplete data in the MACS public data set covering the years 1984–2002. Previous analyses have been on complete cases and a much smaller subset of the data. We use Gibbs sampling or “data augmentation,” commonly employed in regression models but which can be easily adapted to the diffusion model used here. As in Faucett and others (2002), we treat missing data by multiply imputing the missing data and analyzing the augmented data. Our objectives are to determine the feasibility of missing data methodology, such as Gibbs sampling for a large data set, and to determine to what extent the incomplete data may affect the analysis. A Bayesian analysis, although slow, is feasible in this case, and the results are similar to those obtained from either a Bayesian or a maximum likelihood analysis of the complete cases, probably because the data set is large and mechanisms that control the missingness are sufficiently close to completely at random as to impart little or no bias to the complete case estimators. There is no general guarantee that a complete case analysis is devoid of bias, and we contend that both analyses, complete case, and the more computationally intense Bayesian analysis investigated here should be done. Since we do not attempt to model the mechanism by which data are incomplete, we implicitly make the standard assumption that the observations are missing at random. There is no evidence for or against informative dropout in this context, nor are we able, with the information provided, to address the missing at random assumption. See Diggle and others (2007) for a discussion of the complex issues and biases that may arise with informative dropout or censoring.

2. Longitudinal Model

Let \( Y(t) \) be the true CD4 count for one individual at time \( t \). We contend that the model for \( Y(t) \) should allow for an equilibrium, at least in the nondiseased state and in the absence of parameter changes. This
limits possible models for $\mathcal{Y}(t)$ to those that are potentially stationary, ruling out, for example, the integrated Ornstein–Uhlenbeck or a Brownian motion process. The easiest example of a stationary dependent process is the well-known Ornstein–Uhlenbeck. There are physiological arguments and precedents for the mean reversion property of the Ornstein–Uhlenbeck in the literature (see Castiglione and others, 2007, p. 3352).

Diggle and others (2005) and McNeil and Gore (1996) use a square-root transformation and Taylor and Law (1998) use a fourth-root transformation of CD4 count so evidently the scaling implicit in the measurement of CD4 count is not uniquely consistent with the Ornstein–Uhlenbeck model. We will model $\mathcal{Y}(t)$ as a power transformation of an Ornstein–Uhlenbeck process with time-varying parameter. In particular, we assume that the diffusion model for the CD4 counts is

$$d\mathcal{Y}(t) = m(\mathcal{Y}(t))dt + \sqrt{2\kappa} \left( \frac{\sigma_\kappa}{\xi} \right) \mathcal{Y}(t)^{1-\xi} dW(t),$$

(2.1)

where

$$m(y) = \frac{\kappa}{\xi} \left[ \theta(t)y^{1-\xi} - y + \sigma_\kappa^2 \left( \frac{1}{\xi} - 1 \right) y^{1-2\xi} \right],$$

(2.2)

where $W(t)$ is a standard Wiener process and $\kappa, \xi$, and $\sigma_\kappa$ are positive parameters which will be estimated from the data. The function $m(y)$, though complex, behaves in a natural way. When the CD4 count is below a target value, the drift in (2.2) is positive and the process has a tendency to increase. When it is above a target value, it has a tendency to decrease. We assume that the initial CD4 count is positive and to avoid negative values, consider this model only up to a stopping time defined to be the first passage time of $\mathcal{Y}(t)$ to 0.

We see below that our assumed model is a power transformation of a model which, when parameters and covariates are constant, is an Ornstein–Uhlenbeck process. More complex diffusion models are also possible in view of the ability to simulate and determine likelihoods exactly (see, e.g. Beskos and others, 2006), but we limit our analysis to (2.1) in view of the large number of parameters and missing values that are involved. The power transformation provides considerable flexibility and includes as special cases many of the models in the literature.

It is easy to express (2.1) as a function of a generalized Ornstein–Uhlenbeck process with time-varying parameter if we let $Y(t) = (\mathcal{Y}(t))^\xi$. Using Ito’s formula, we can show

$$dY(t) = \kappa [\theta(t) - Y(t)]dt + \sqrt{2\kappa} \sigma_\kappa dW(t).$$

(2.3)

If $\theta(t)$ is constant in $t$, then (2.3) is the Ornstein–Uhlenbeck process.

In general, the parameters $\sigma_\kappa$, $\kappa$, and the “target” $\theta(t)$ in (2.3) may depend on time $t$ and the covariates. HIV disease progresses to AIDS if the CD4 count decreases. However, there are many factors that can also influence CD4 counts. For example, illnesses such as pneumonia, influenza, herpes simplex virus infection etc., or chemotherapy treatment can cause CD4 counts to decrease, while antiviral drugs can cause an increase in CD4 counts. We model the effect of the covariates mainly through $\theta(t)$ that varies in $t$ due to time-varying covariates and a seroconversion time that is interval censored. The parameter $\kappa$, which is the rate of reversion toward the target, and $\sigma_\kappa$ are assumed for simplicity the same for all individuals.

For $\kappa > 0$ and constant initial value $y_0$, (2.3) dynamically tends toward the current target $\theta(t)$ and is a Gaussian process. From (2.3),

$$Y(t) = y_0 + \int_{t_0}^t \kappa [\theta(s) - Y(s)]ds + \sqrt{2\kappa} \sigma_\kappa [W(t) - W(t_0)].$$

(2.4)
Let \( \mu(t) = E[Y(t)]|Y(t_0) = y_0 \) and take the expected value of both sides of (2.4), conditional on \( Y(t_0) = y_0 \). Then \( \mu(t) \) must satisfy \( \mu(t) = y_0 + \int_{t_0}^{t} \kappa(\theta(s) - \mu(s))\,ds \) or \( \mu'(t) + \kappa \mu(t) = \kappa \theta(t) \) that gives \( \mu(t) = e^{-\kappa(t-t_0)}y_0 + \int_{t_0}^{t} \theta(s)\kappa e^{\kappa(s-t_0)}\,ds \). If the process began in the distant past so that \( t_0 \to -\infty \), then

\[
\mu(t) = \int_{-\infty}^{t} \theta(s)\kappa e^{-\kappa(t-s)}\,ds, \tag{2.5}
\]

which is the convolution of \( \theta(s) \) and the exponential probability density function with mean 1/\( \kappa \). Thus, \( \mu(t) \) is a weighted average of the previous values of \( \theta(s) \), that is, \( \mu(t) \approx E[\theta(t - \psi)] \), where \( \psi \) has an exponential distribution with mean 1/\( \kappa \). Since \( \theta(s) \) is a function of the covariates, the parameter \( \kappa \) determines the extent of the smoothing of these covariate values. Since \( P(\psi > s) = 0.5 \) when \( s = \kappa^{-1}\ln 2 \), the parameter \( \kappa^{-1}\ln 2 \) is the “half-life” of the covariate memory effect. As \( \kappa \) decreases, the half-life increases.

To achieve a smoothing over time, Diggle and others (2005) assume the regression model (1.1), where \( \mu(t) \) is constant prior to seroconversion and a quadratic function thereafter. In our case, the smoothing is embedded in (2.5) and is a consequence of the diffusion model. If \( \theta(u) = \theta \) is constant for \( u \in [s, t) \), then (2.5) implies \( \mu(t) = e^{-\kappa(t-s)}(\mu(s) + \theta(1 - e^{-\kappa(t-s)}) \), a weighted average of \( \mu(s) \) and \( \theta \) with weights \( e^{-\kappa(t-s)} \) and \( 1 - e^{-\kappa(t-s)} \), respectively. Thus, even in an interval in which \( \theta(t) \) is constant because the covariates do not change, the mean \( \mu(t) \) does change due to the memory effect.

The form of (2.5) suggests transforming to \( Z(t) = Y(t) - \mu(t) \) for which

\[
dZ(t) = dY(t) + \left\{ \kappa^2 e^{-\kappa t} \int_{-\infty}^{t} \theta(s)e^{\kappa s}\,ds - \kappa e^{-\kappa t}\theta(t) e^{\kappa t} \right\} dt \tag{2.6}
\]

\[
= -\kappa Z(t) dt + \sqrt{2\kappa}\sigma_k dW(t).
\]

This implies that the transformed process \( Z(t) \) is a standard Ornstein–Uhlenbeck process. From (2.6), we obtain, for \( t_0 < s < t \),

\[
\text{cov}[Y(s), Y(t)|Y(t_0)] = \text{cov}[Z(s), Z(t)|Z(t_0)] = \sigma_k^2[1 - e^{-2\kappa(s-t_0)}]e^{-\kappa|t-s|}
\]

and the conditional distribution of \( Y(t) \) given \( Y(t_0) = y_0 \) is

\[
N\left[e^{-\kappa(t-t_0)}y_0 + \kappa \int_{t_0}^{t} \theta(s)e^{\kappa(s-t)}\,ds, \sigma_k^2[1 - e^{-2\kappa(t-t_0)}] \right].
\]

Let \( V(t) \) be the observed value of \( Y(t) \) for one individual at time \( t \). We assume that

\[
V(t) = Y(t) + \varepsilon(t), \tag{2.7}
\]

where the measurement errors \( \varepsilon(t) \) are independent \( N(0, \sigma_\varepsilon^2) \) random variables. \( V(t) \) is Gaussian with the same mean as \( Y(t) \) and covariance \( \text{cov}[V(s), V(t)|Y(t_0) = y_0] = \sigma_k^2[1 - e^{-2\kappa(s-t_0)}]e^{-\kappa|t-s|} + \sigma_\varepsilon^2 I(s = t) \). If the process began in the distant past so that \( t_0 \to -\infty \), then \( \text{cov}[V(s), V(t)|Y(t_0) = y_0] \simeq \sigma_k^2 e^{-\kappa|t-s|} + \sigma_\varepsilon^2 I(s = t) \).

### 2.1 Covariates and incomplete observations

The data, taken from the 2007 release of the MACS public data set for the years 1984–2002 (www.statepi.jhsph.edu/macs), consist of a total of 11,743 visits for 595 seroconverters. A large amount of information is either missing or censored to ensure privacy of individuals. Seroconversion times are “all” interval
particular diffusion model for incomplete longitudinal data

Table 1. Frequency of missing observations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number missing (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD4 count</td>
<td>1190 (10.1)</td>
</tr>
<tr>
<td>Cigarette smoking</td>
<td>114 (1)</td>
</tr>
<tr>
<td>Recreational drugs</td>
<td>237 (2)</td>
</tr>
<tr>
<td>Antiviral AIDS drugs</td>
<td>112 (1)</td>
</tr>
<tr>
<td>Non-antiviral AIDS drugs</td>
<td>111 (1)</td>
</tr>
<tr>
<td>Depressive symptoms</td>
<td>666 (5.7)</td>
</tr>
<tr>
<td>No. of male sexual partners</td>
<td>302 (2.6)</td>
</tr>
</tbody>
</table>

censored since seroconversion is only known to have occurred on an interval between consecutive visits. For some individuals, this interval is only 6 months long, while for others, the interval is completely unknown. These data provide an ideal testing ground for the treatment of missing and incomplete data. Since most analyses (including, as far as we can determine, that in Diggle and others, 2005) use earlier releases of these data and delete any observations that have missing components, it is valuable to determine whether a more sophisticated treatment of these missing and incomplete data gives results that differ significantly from an analysis based on complete cases.

A full description of the covariates used is given in Appendix A of the supplementary material available at Biostatistics online. Table 1 summarizes the missingness for the response and the covariates. If any visit for which either the CD4 count or a covariate is missing is discarded, the number of visits is reduced to 9997 or about 85% of the original data. If the visits corresponding to seroconverters whose interval of seroconversion is greater than 6 months are also discarded the data would be further reduced to 8005 visits or 68% of the original data. Since the remaining data are more sparse in time and carry less information about the time dynamics of the process, we prefer to keep all 11743 visits and develop a method for parameter estimation as described in Section 3.1 which handles the missing information on covariates (including seroconversion time) and responses.

2.2 Complete data likelihood and estimation for the longitudinal model

In this section, we construct a likelihood function using the model specified by (2.3), (2.7), and the “complete data,” that is, the subset of the data with no missing responses or covariates and seroconversion times known to within a 6-month interval. We assume that

$$\theta(t) = x(t)\beta + Z,$$

where $\beta$ is a $p \times 1$ vector of unknown regression coefficients, $x(t)$ is a $1 \times p$ vector of time-varying covariates for this individual, and $Z$ is a $N(0, \sigma_i^2)$ random effect.

Let $V(t) = \{V(t_1), \ldots, V(t_d)\}$ be the $d \times 1$ vector of responses for one individual with $t = (t_1, \ldots, t_d)$, $t_1 < \cdots < t_d$, and $V(t)$ is (observed CD4 count at time $t$). Our analysis could be extended to allow for continuous changes in the covariate process, but since the covariates are only observed at discrete times, we assume that $\theta(t)$ only changes at these observation times and at the time of seroconversion ($t = 0$). The covariate values $x(t_j)$ measured at observation time $t_j$ are assumed to apply over the previous time interval $(t_{j-1}, t_j)$ and to remain constant for $t > t_d$. This assumption and (2.8) imply that $\theta(t)$ is a piecewise constant function on the intervals $(t_j, t_{j+1}]$, $j = 1, \ldots, d$. By (2.5) and (2.8), the mean for this individual at time $t$ is equal to $b_{\kappa}(t)\beta$, where

$$b_{\kappa}(t) = \int_{-\infty}^{t} x(s)\kappa e^{\kappa(s-t)}ds.$$
\[
\begin{align*}
\mathbf{x}(t_j) = & \ e^{-\kappa t_j} \sum_{i=1}^{j-1} \mathbf{e}^{\kappa t_i} [\mathbf{x}(t_i) - \mathbf{x}(t_1)] + e^{\kappa t_j} [\mathbf{x}(t_j) - \mathbf{x}(t_{j-1})], & t_{j-1} < t_j, \\
\mathbf{x}(t_1), & t_l \leq t_1, \\
\mathbf{x}(t_d), & t > t_d.
\end{align*}
\] (2.9)

We see that \( \mathbf{b}_\kappa(t) \) is a weighted average of past covariate values for this individual. Note also that as \( \kappa \to \infty \), \( \mathbf{b}_\kappa(t) \to \mathbf{x}(t) \) and as \( \kappa \to 0 \), \( \mathbf{b}_\kappa(t) \) approaches the long-run average of the values \( \mathbf{x}(s), s \leq t \), that is, \( \mathbf{b}_\kappa(t) \) is a weighted average of the observed covariate values with weights proportional to the length of the time interval over which they are assumed to be constant. If the process is assumed to start infinitely far in the past so that \( t_0 \to -\infty \), then this weighted average is dominated by \( \mathbf{x}(t_1) \), the covariates observed at the first observation time \( t_1 \), and thus \( \mathbf{b}_\kappa(t) \to \mathbf{x}(t_1) \) as \( \kappa \to 0 \). If we let \( \mathbf{B}_\kappa(t) \) be the \( d \times p \) matrix whose \( j \)th row is given by \( \mathbf{b}_\kappa(t_j) \), \( j = 1, \ldots, d \), then as \( \kappa \to 0 \), \( \mathbf{B}_\kappa(t) \) approaches a matrix with \( d \) identical rows all equal to \( \mathbf{x}(t_1) \).

For known \( \xi \), we can construct the complete data likelihood for the longitudinal data since, conditional on the initial values, the process is Gaussian. Let \( \mathbf{I}_d \) be a \( d \times 1 \) vector of ones, and let \( \mathbf{I}_d \) be the \( d \times d \) identity matrix. Then for one individual \( \mathbf{V}(t) \) has an MVN distribution and

\[
\text{Σ}_V(t) = \sigma_\kappa^2 \mathbf{A}_\kappa(t) + \sigma_\epsilon^2 \mathbf{I}_d + \sigma_\sigma^2 \mathbf{I}_d \mathbf{I}_d^T
\] (2.10)
and \( \mathbf{A}_\kappa(t) \) is the \( d \times d \) matrix with \( (j, l) \) component equal to \( e^{-\kappa |t_j - t_l|}, j, l = 1, \ldots, d \). Note that \( \sigma_\kappa^2 \) governs the temporal variability in the real underlying process for a given individual, \( \sigma_\epsilon^2 \) measures the variability in the measurement error, and \( \sigma_\sigma^2 \) measures the variability among individuals. Let \( \Phi(\mathbf{v}; \mathbf{u}, \Sigma) \) be the probability density function of an MVN distribution evaluated at \( \mathbf{v} \). Denote the likelihood contribution for the \( i \)th individual using a superscript \((i)\), and let

\[
\mathcal{L}_V^{(i)} = \Phi(\mathbf{V}^{(i)}(t^{(i)}); \mathbf{B}_\kappa^{(i)}(t^{(i)}), \mathbf{B}_\kappa^{(i)}(t^{(i)})), \text{Σ}_V^{(i)}(t^{(i)}))
\]

The likelihood for \( N \) individuals given \( \xi \) is

\[
\mathcal{L}_V(\mathbf{β}, \sigma_\epsilon^2, \kappa, \sigma_\kappa^2, \sigma_\sigma^2) = \prod_{i=1}^{N} \mathcal{L}_V^{(i)}.
\] (2.11)

Since the mean structure is linear in \( \mathbf{β} \), this is a standard linear model.

When \( \xi \) is unknown, the full likelihood given the data are

\[
\mathcal{L}_V(\mathbf{β}, \sigma_\epsilon^2, \kappa, \sigma_\kappa^2, \sigma_\sigma^2) \propto \prod_{i,j} |V^{(i)}(t_j)|^{1-1/\delta} (Hwang, 2004). We estimated \( \xi \) by maximizing the profile likelihood for the complete data over \( \xi \), obtaining, rounded to one decimal \( \xi = 0.4 \). We will fix this value for the remainder of the paper. The remaining parameters were then estimated using the conditional likelihood \( \mathcal{L}_V(\mathbf{β}, \sigma_\epsilon^2, \kappa, \sigma_\kappa^2, \sigma_\sigma^2) \). Estimates of the parameters \( (\mathbf{β}, \sigma_\epsilon^2, \gamma) \) are available in closed form, while the parameters \( (\kappa, \sigma_\kappa^2, \sigma_\sigma^2) \) must be found iteratively. The maximum likelihood estimating equations and the Fisher information matrix are derived in Appendix B of the supplementary material available at Biostatistics online.

The maximum likelihood estimates are given in Table 2. The approximate confidence intervals are obtained by inverting the Fisher information matrix and using as standard error the square root of the diagonal entries. The estimated drop in CD4 count at seroconversion is 5.28 that implies a rapid drop of a CD4 count by inverting the Fisher information matrix and using as standard error the square root of the diagonal entries. The estimated drop in CD4 count at seroconversion is 5.28 that implies a rapid drop of a CD4 count by inverting the Fisher information matrix and using as standard error the square root of the diagonal entries.

The value of \( \kappa = 0.34 \) corresponds to a half-life of the covariate memory effect of around 2 years.

3. Joint likelihood for longitudinal and event-time data

The likelihood \( \mathcal{L}_V \) is based on the longitudinal data for the 595 seroconverters and the model specified by (2.3), (2.7), and (2.8). There is additional information in the MACS data clearly relevant to CD4 counts, most obviously the year of death and cause of death, since CD4 count is closely related to the onset of AIDS. Assume that the probability of death due to AIDS in the small time interval \((t, t + \Delta t)\)
conditioned on survival to time $t$ and $Y(\tau)$ is $\lambda \exp\{-\gamma Y(\tau)\} \Delta t$, where $\lambda$ governs the baseline intensity and $\gamma$ controls the rate of increase of the hazard as the CD4 count decreases. Given the complexities associated with imputing the missing data, we decided to use a simple constant baseline hazard and only the covariate $Y(\tau)$. More complex alternatives are also possible but computationally more difficult due to the large amount of missing data.

Suppose the event–time data for one individual are $(\tau, \delta)$, where $\tau$ is the time of death or lost to follow up if the individual is censored, and $\delta = 1$ indicates a death due to AIDS at time $\tau$. The contribution to the likelihood of the event–time data, provided that the process $Y$ is observed after seroconversion, is

$$
\mathcal{L}_{\text{ET}} = \int_0^\tau \exp\left\{ -\lambda \int_0^\tau e^{-\gamma Y(u)} \, du \right\} \lambda^{\delta - 1} \exp\{ -\gamma Y(\tau) \} \Delta t.
$$

We begin the integral at time $t = 0$ since death due to AIDS is assumed to be impossible before seroconversion at $t = 0$. Expression (3.1) is the conditional likelihood given the complete path of the process $Y(\tau)$. However, this path is not fully observed. CD4 counts are only observed with measurement error at specific times. Because the longitudinal parameters are present in both $\mathcal{L}_V$ and implicitly in $\mathcal{L}_{\text{ET}}$ (they influence the unobserved path $Y(u), u > 0$), estimation of the parameters is more difficult. A standard alternative which better separates the parameters is to use the “joint likelihood” of the observed $V(t)$ and the path $Y(\tau)$, where $Y(\tau)$ is regarded as incompletely observed.

For one individual, we let $V(t) = \{V(t_1), \ldots, V(t_d)\}^T$ be the $d \times 1$ vector of observed transformed responses and $t_k = (t_1^k, \ldots, t_d^k)$ be a grid of $n$ time points with $t_1^k < t_2^k < \cdots < t_n^k = \tau$. The set of points $t_k$ is chosen to include all the time points at which covariate information is available as well as the seroconversion time $t = 0$ and augmented so that $t_i^k - t_{i-1}^k$ is small. Let $Y(t) = \{Y(t_1), \ldots, Y(t_d)\}^T$ and $Y(t_k) = \{Y(t_1^k), \ldots, Y(t_d^k)\}^T$ be the vectors of true values of the response at the times $t$ and $t_k$,

<table>
<thead>
<tr>
<th></th>
<th>Longitudinal model</th>
<th>Longitudinal model</th>
<th>Joint model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>complete data</td>
<td>incomplete data</td>
<td>incomplete data</td>
</tr>
<tr>
<td>$b_1$ (seroconversion indicator)</td>
<td>$-5.28 (-5.73, -4.83)$</td>
<td>$-5.25 (-5.69, -4.82)$</td>
<td>$-5.57 (-6.15, -5.04)$</td>
</tr>
<tr>
<td>$b_2$ (constant)</td>
<td>$13.98 (13.37, 14.59)$</td>
<td>$13.88 (13.37, 14.4)$</td>
<td>$13.31 (12.74, 13.88)$</td>
</tr>
<tr>
<td>$b_3$ (smoking)</td>
<td>$0.56 (0.36, 0.77)$</td>
<td>$0.51 (0.33, 0.68)$</td>
<td>$0.52 (0.34, 0.77)$</td>
</tr>
<tr>
<td>$b_4$ (recreational drugs)</td>
<td>$0.20 (-0.21, 0.60)$</td>
<td>$0.38 (0.04, 0.72)$</td>
<td>$0.42 (0.05, 0.79)$</td>
</tr>
<tr>
<td>$b_5$ (antiviral drugs)</td>
<td>$2.33 (1.76, 2.91)$</td>
<td>$2.50 (1.97, 3.04)$</td>
<td>$2.91 (2.29, 3.60)$</td>
</tr>
<tr>
<td>$b_6$ (non-antiviral drugs)</td>
<td>$-3.66 (-4.27, -3.06)$</td>
<td>$-3.91 (-4.43, -3.38)$</td>
<td>$-4.22 (-4.83, -3.64)$</td>
</tr>
<tr>
<td>$b_7$ (depressive symptoms)</td>
<td>$-1.00 (-1.40, -0.61)$</td>
<td>$-0.82 (-1.17, -0.48)$</td>
<td>$-0.86 (-1.23, -0.48)$</td>
</tr>
<tr>
<td>$b_8$ (no. sexual partners)</td>
<td>$0.45 (0.18, 0.73)$</td>
<td>$0.41 (0.16, 0.66)$</td>
<td>$0.64 (0.37, 0.92)$</td>
</tr>
<tr>
<td>$\sigma^2_\gamma$ (measurement error)</td>
<td>$0.89 (0.80, 0.98)$</td>
<td>$0.82 (0.75, 0.89)$</td>
<td>$0.84 (0.78, 0.91)$</td>
</tr>
<tr>
<td>$\kappa$ (kappa)</td>
<td>$0.34 (0.30, 0.38)$</td>
<td>$0.33 (0.29, 0.37)$</td>
<td>$0.28 (0.24, 0.32)$</td>
</tr>
<tr>
<td>$\sigma^2_\kappa$ (variance kappa)</td>
<td>$4.63 (4.21, 5.05)$</td>
<td>$4.69 (4.32, 5.05)$</td>
<td>$5.34 (4.82, 5.95)$</td>
</tr>
<tr>
<td>$\sigma^2_\lambda$ (random effects)</td>
<td>$2.44 (1.81, 3.06)$</td>
<td>$2.24 (1.75, 2.72)$</td>
<td>$2.38 (1.81, 3.00)$</td>
</tr>
<tr>
<td>$\lambda$ (baseline intensity)</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>$\gamma$ (hazard rate)</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
</tbody>
</table>

na = not applicable.
respectively. The joint distribution of $V(t)$ and $Y(t_g)$ (see (2.9) and (2.10)) is

$$
\begin{align*}
\begin{bmatrix} V(t) \\ Y(t_g) \end{bmatrix} & \sim \text{MVN} \left[ \begin{bmatrix} B_k(t)\beta \\ B_k(t_g)\beta \end{bmatrix}, \begin{bmatrix} \Sigma_V(t) & \Sigma(t, t_g) \\ \Sigma(t_g, t) & \Sigma_Y(t_g) \end{bmatrix} \right],
\end{align*}
$$

where $\Sigma_Y(u) = \sigma^2_k \Lambda_k(u) + \sigma^2 I_m(1_m)^T$ and $\Sigma(t, t_g) = \text{cov}\{V(t), Y(t_g)\}$ is the $d \times n$ matrix whose rows are equal to the rows of $\Sigma_Y(t_g)$ which correspond to the time points $t$. The conditional distribution of $V(t)$ given $Y(t_g)$ is $\text{MVN}(Y(t), \sigma^2 \mathbf{I}_d)$.

The joint likelihood of the observed $V^{(i)}(t^{(i)})$ and the unobserved $Y^{(i)}(u)$ for all individuals is

$$
\mathcal{L}_{V,Y} = \prod_{i=1}^N \mathcal{L}_{V^{(i)}|Y} \mathcal{L}_{Y^{(i)}},
$$

where $\mathcal{L}_{Y^{(i)}}$ is given by (3.1), $\mathcal{L}_{V^{(i)}|Y} = \Phi\{Y^{(i)}(t_g^{(i)}) ; B_k^{(i)}(t_g^{(i)})\beta, \Sigma_Y^{(i)}(t_g^{(i)})\}$ and $\mathcal{L}_{V^{(i)}} = \Phi\{V^{(i)}(t^{(i)}) ; Y^{(i)}(t^{(i)}), \sigma^2 \mathbf{I}_d\}$.

### 3.1 Bayes estimation and MCMC

We used Metropolis within Gibbs (see, e.g. Robert and Casella, 2004) to generate observations from a complex (posterior) distribution in a number of simpler steps. There are many unknowns: missing covariates (including seroconversion times); missing responses (CD4 counts); the unknown parameters $\theta = (\beta, \sigma^2_k, \kappa, \sigma^2, \gamma)$; and the unobserved path $Y(t)$. MCMC draws from the joint posterior distribution of all of the unknowns, including posterior distributions for the parameters and the missing values. McNeil and Gore (1996) used Gibbs sampling to obtain posterior distributions for parameters for a smaller dataset without missing observations and a simpler model.

The likelihood function (3.2) is a product of independent contributions of similar form from $N$ individuals. We assume proper uniform priors for $s$, the unknown seroconversion time, and for $x_{\text{miss}}$, the vector of missing covariates. For each individual, we generate values of the seroconversion time and missing covariates given the parameter values and generated paths from the previous simulation. Let $\Gamma(\alpha, \lambda)$ be the Gamma distribution with shape $\alpha$ and scale $1/\lambda$. The algorithm proceeds as follows:

1. **Update ($s, x_{\text{miss}}$)**: Assume a proper uniform proposal for the conditional distribution of the seroconversion time and missing covariates given $Y(t_g)$ and $\theta$. Generate new values of ($s, x_{\text{miss}}$) from a proper uniform distribution for individual $i$ and accept the move to these new values with probability proportional to the ratio $\mathcal{L}_{V^{(i)}}^{(i)}(s_{\text{new}}, x_{\text{miss}}^{\text{new}}) / \mathcal{L}_{V^{(i)}}^{(i)}(s_{\text{old}}, x_{\text{miss}}^{\text{old}})$, otherwise keep the old values.

2. **Update $Y(t_g)$**: The conditional distribution of $Y(t_g)$ given $V(t)$, ($s, x_{\text{miss}}$), $\theta$ and the event time data is $\Phi\{Y(t_g) ; \mu_{Y|V}(t_g, t), \Sigma_{Y|V}(t_g, t)\} \mathcal{L}_{ET}$, where

$$
\begin{align*}
\mu_{Y|V}(t_g, t) &= B_k(t_g)\beta + \Sigma(t_g, t)\{\Sigma_V(t)\}^{-1}[V(t) - B_k(t)\beta],
\end{align*}
$$

and

$$
\Sigma_{Y|V}(t_g, t) = \Sigma_{Y}(t_g) - \Sigma(t_g, t)\{\Sigma_V(t)\}^{-1}\Sigma(t, t_g).
$$

Generate $Y(t_g)$ from a $\text{MVN}(\mu_{Y|V}, \Sigma_{Y|V})$ distribution and accept the proposal with probability proportional to $\mathcal{L}_{ET}$. New proposals are generated until one is accepted. The integral in $\mathcal{L}_{ET}$ was approximated using the trapezoidal rule.

3. **Update $\lambda$**: Using a uniform prior, generate $\lambda$ from its conditional distribution given everything else, a $\Gamma(D + 1, \sum_{i=1}^N h_i)$ distribution, where $D = \sum_{i=1}^N \delta_i$ and $h_i = \int_0^T e^{-\gamma}Y^{(i)}(w) \, dw$. 


4. **Update** $\sigma^2_e$: The only term in the likelihood which depends on $\sigma^2_e$ is $\prod_{i=1}^N L^{(i)}_{Y(i)}$. Assuming that $\sigma^2_e$ has a proper uniform prior, we generate $\sigma^2_e$ by letting $\sigma^2_e = 1/u$ where $u$ is generated from a $\Gamma(c_1, c_2)$ distribution with $c_2 = 0.5\sum_{i=1}^N \{V^{(i)}(t^{(i)}) - Y^{(i)}(t^{(i)})\}^T \{V^{(i)}(t^{(i)}) - Y^{(i)}(t^{(i)})\}$ and $c_1 = D/2 - 1$.

5. **Update** $\beta$: With an improper uniform prior for $\beta$, we generate $\beta$ from its conditional distribution given everything else, the MVN($\mu_\beta$, $\Sigma_\beta$) distribution, where

$$
\Sigma_\beta = \sum_{i=1}^N \{B_K(t^{(i)}_g)\}^T \{\Sigma_\gamma(t^{(i)}_g)\}^{-1} \{B_K(t^{(i)}_g)\}
$$

and

$$
\mu_\beta = (\Sigma_\beta)^{-1} \sum_{i=1}^N \{B_K(t^{(i)}_g)\}^T \{\Sigma_\gamma(t^{(i)}_g)\}^{-1} Y^{(i)}(t^{(i)}).
$$

6. **Update** $(\kappa, \sigma^2_k, \sigma^2_\tau)$: Assuming uniform priors for the parameters $(\kappa, \sigma^2_k, \sigma^2_\tau)$, we use one step of a Metropolis algorithm. We generate $(\kappa, \sigma^2_k, \sigma^2_\tau)$ from a multivariate normal proposal distribution with mean equal to the old parameter values. We accept the move with probability proportional to the ratio of $\prod_{i=1}^N L^{(i)}_{Y}$ evaluated at the new values to $\prod_{i=1}^N L^{(i)}_{Y}$ evaluated at the old values.

7. **Update** $\gamma$: Assuming a uniform prior for $\gamma$, we use one step of a Metropolis algorithm. We generate $\gamma$ using a normal distribution with mean equal to the old parameter value. We accept the move with probability proportional to the ratio of $\prod_{i=1}^N L^{(i)}_{ET}$ evaluated at the new value to $\prod_{i=1}^N L^{(i)}_{ET}$ evaluated at the old value.

MCMC generates a sequence of values of the parameters and missing values that, in this case, approaches the joint posterior distribution reasonably quickly, and so provides an estimate of the posterior mean and the posterior covariance matrix of the parameters given the data. For convenience, we initiated the vector $\theta = (\beta, \sigma^2_e, \kappa, \sigma^2_\kappa, \sigma^2_\tau, \lambda, \gamma)$ using the maximum likelihood estimates obtained from the complete cases (see Section 2). Different starting values for $\theta$ gave similar results. We discarded the first thousand points in the sequence to allow for a break-in period, and averaged the next 47,000 points to obtain the Bayes estimates of the parameters. Plots found in Appendix C of the supplementary materials available at *Biostatistics* online provide evidence that the MCMC run of 47,000 is sufficiently long for adequate mixing.

One major advantage in having repeated draws from the posterior distribution is the ability to construct interval estimates or credible regions for the parameters. The interval estimates for the parameters are given in Table 2. For comparison we also ran a similar MCMC, on the longitudinal model alone, that is, without the event–time data. The results are also displayed in Table 2 for 35,100 steps of a Metropolis-in-Gibbs algorithm. Evidently, all of the covariates are significant, although some (e.g. $\beta_4$), only marginally so. In some cases the sign of the parameters is surprising (e.g. $\beta_3$, $\beta_6$, $\beta_8$), possibly because of a confusion of cause and effect and confounding with other factors. There are no substantial differences between the estimates in the longitudinal and joint models, although the variance parameters in the joint model are somewhat larger. This is not surprising since the residual sum of squares is effectively minimized in fitting the longitudinal model which leads to smaller values of the variance parameters. The values of $\kappa$ correspond to a half-life of the covariate memory effect of around 2.1 years for the longitudinal model to 2.5 years in the joint model. This indicates that a model which involves not only the current values of the covariates but their values over the past couple of years is important. The apparently modest value of $\beta_1$ is deceiving because the measurements have been transformed. It corresponds to a fairly rapid drop in a CD4 count of 1000, say, prior to seroconversion to about 340 after. The estimates based on the complete and incomplete data are also similar, particularly considering that one analysis is Bayesian and based on
MCMC while the other is based on maximum likelihood and the information matrix using the complete cases only. There are some differences. The value of $\beta_4$ under the complete case analysis is smaller and not significant. It is surely reassuring that, in this example, whether we used the longitudinal or joint model, whether we used Bayesian or maximum likelihood methodology, and whether we use a complete case analysis or accommodate the missing values, we generally obtain qualitatively similar conclusions for the effects of most covariates.

An advantage of the Bayes approach is the ability to obtain an easy measure of the dependence among various parameter estimators. An estimate of the correlation matrix of the posterior distribution of the parameters from our simulations is given in Table 3. The large negative correlation ($-0.79$) between the estimate of $\kappa$ and that of $\sigma^2_\kappa$ is largely caused by our choice of parameterization since the diffusion coefficient is the product $\sqrt{2\kappa}\sigma_\kappa$. This gives a possible explanation for the corresponding increase in the estimate of $\sigma^2_\kappa$ when the event time data are included.

### 3.2 Fit of the model

It is very difficult to formally test the fit of the model for several reasons. The large number of incomplete observations makes it necessary to impute missing values before testing fit, and this imputation must be taken into account in the distribution of the test statistic. Incomplete observations in continuous time also make it difficult to assess the correlation structure of the underlying diffusion. To provide some indication of fit, we calculated standardized residuals $R_i = \{\Sigma^{(i)}_V(t^{(i)})\}^{-1/2}\{V^{(i)}(t^{(i)}) - B^{(i)}_\kappa(t^{(i)})\beta\}$ based on the observed data and using the parameter estimates shown in Table 2. Assuming a correct specification of the parameters and the covariance structure, the elements of $R = \{R^T_1, R^T_2, \ldots, R^T_N\}$ should be approximately independent standard normal random variables. To produce $R$, we must draw values for missing covariates and seroconversion times from their predictive distributions and therefore the plot may depend on this draw but we found that the plots based on different draws differed very little.

Appendix D of the supplementary material (available at Biostatistics online) contains a plot of the autocorrelation function of the transformed residuals, and a QQ plot and histogram of these residuals. The plots indicate that the residuals are uncorrelated and suggest a heavier tailed distribution than the normal, not surprising with over 10 000 observations. Although a heavier tailed distribution might have

<table>
<thead>
<tr>
<th></th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
<th>$\beta_7$</th>
<th>$\beta_8$</th>
<th>$\sigma^2_\kappa$</th>
<th>$\kappa$</th>
<th>$\sigma^2_\kappa$</th>
<th>$\sigma^2_\gamma$</th>
<th>$\lambda$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.21</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.03</td>
<td>-0.22</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.15</td>
<td>-0.21</td>
<td>-0.13</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>-0.68</td>
<td>-0.14</td>
<td>0.07</td>
<td>0.04</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>0.18</td>
<td>-0.05</td>
<td>-0.03</td>
<td>0.03</td>
<td>-0.55</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>0.00</td>
<td>-0.10</td>
<td>-0.07</td>
<td>-0.02</td>
<td>-0.05</td>
<td>-0.05</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_8$</td>
<td>0.07</td>
<td>-0.76</td>
<td>0.06</td>
<td>-0.30</td>
<td>0.10</td>
<td>0.06</td>
<td>-0.01</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_\kappa$</td>
<td>-0.24</td>
<td>-0.06</td>
<td>0.03</td>
<td>-0.02</td>
<td>0.23</td>
<td>-0.13</td>
<td>-0.02</td>
<td>0.06</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.72</td>
<td>0.12</td>
<td>-0.07</td>
<td>0.04</td>
<td>-0.64</td>
<td>0.30</td>
<td>0.04</td>
<td>-0.11</td>
<td>-0.36</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_\gamma$</td>
<td>-0.58</td>
<td>-0.08</td>
<td>0.06</td>
<td>-0.03</td>
<td>0.51</td>
<td>-0.23</td>
<td>-0.02</td>
<td>0.08</td>
<td>-0.08</td>
<td>-0.79</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_\gamma$</td>
<td>0.18</td>
<td>-0.03</td>
<td>-0.04</td>
<td>0.08</td>
<td>-0.18</td>
<td>0.16</td>
<td>-0.06</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.30</td>
<td>-0.36</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>-0.13</td>
<td>-0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>0.09</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.04</td>
<td>0.08</td>
<td>-0.13</td>
<td>0.08</td>
<td>-0.01</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-0.16</td>
<td>-0.06</td>
<td>0.00</td>
<td>0.01</td>
<td>0.11</td>
<td>-0.04</td>
<td>0.00</td>
<td>0.06</td>
<td>0.05</td>
<td>-0.15</td>
<td>0.13</td>
<td>-0.00</td>
<td>0.89</td>
<td>1</td>
</tr>
</tbody>
</table>
been used, this would also make the theoretical properties of the model very difficult to ascertain. Of course, we assumed a normal distribution for the residuals, “conditional on known values of the parameters and the missing covariates” whereas the QQ plot gives the “marginal distribution” affected by the uncertainty in the parameter $\beta$ and the imputation but this apparently does not fully explain the heavy tails.

3.3 Estimation of population mean and individual trajectories

The first objective in Diggle and others (2005) was to estimate the average time of CD4 cell depletion. Since we do not have a parametric specification of the posterior distribution of $Y(u)$ given the data, we use Gibbs sampling as in Section 3.1 for estimating the posterior means and marginal distributions for the parameters given the data. After running the 47,000 simulations and generating the estimates in Table 2, we ran an additional 1000, generating paths $Y(t_g)$ for each of the 595 individuals using a probability density proportional to $\Phi(Y(t_g); \mu_{Y|V(t_g,t)}, \Sigma_{Y|V(t_g,t)})L_{ET}$, where $t_g$ was chosen to be a common set of time points between $-5$ and $15$. The mean over all individuals alive at time $t$ is displayed in Figure 1 together with a scatterplot of the observed CD4$^\xi$ values. The observed values of CD4$^\xi$ with $\xi = 0.4$ are plotted against the time since (imputed) seroconversion time. The mean plot is quite natural and continuous, flat before seroconversion and decreasing thereafter. This methodology and the inclusion of the event–time data has eliminated the apparent increase in CD4 count after 4 years in the mean plot, Figure 5.1, of Diggle and others (2005). This analysis apparently mitigates the effect of
the survivorship bias caused because CD4 counts of long-term surviving individuals tend to be higher than those who dropped out.

The second objective in Diggle and others (2005) was to estimate the time course for individuals taking account measurement error. We chose 2 individuals whose trajectories are also displayed in Figure 5.13 of Diggle and others (2005), one who was alive in 2001 and the other who died of AIDS in 1996. We then used the 1000 paths which were used to generate the mean plot above to estimate the trajectories for these 2 individuals, conditional on the observed values. For example in Figure 2, corresponding to an individual who died due to AIDS, the conditional expected value of CD4$^{0.4}$ given the data is plotted together with the actual data. The conditional expected value provides a reasonable estimate of the actual CD4 counts with measurement error removed for a given individual. The dotted lines, are the means ± 2 standard errors and so provide an approximate 95% credible region. We have used different symbols plotted at CD4$^{0.4} = 0$ to indicate visits with missing covariates and missing responses. Clearly, standard errors increase in periods with a number of missing visits.

4. Conclusions

We have fit a particular diffusion model to transformed CD4 counts from the MACS data set, where the diffusion parameters are driven by covariates. Gibbs sampling is used to accommodate the large amount of incomplete data. We demonstrate the feasibility of a Bayesian approach to this problem and show that, in this case at least, dropping the incomplete cases does not strongly influence the parameter estimates. This continuous time model provides a reasonable fit to the CD4 data, even in the presence of large amounts of incomplete information and censoring.
SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENT

Conflict of Interest: None declared.

FUNDING

Natural Sciences and Engineering Research Council of Canada (NSERC-RGPIN8335-06 to D.L.M.).

REFERENCES


[Received July 1, 2010; revised November 15, 2010; accepted for publication December 2, 2010]