BIOMETRICS 56, 160–166
March 2000

A Linear Mixed-Effects Model for Multivariate Censored Data

Wei Pan* **and Thomas A. Louis**

Division of Biostatistics, School of Public Health, University of Minnesota,
Box 303, Mayo Memorial Building, 420 Delaware Street SE, Minneapolis, Minnesota 55455-0378, U.S.A.
**email:* weip@biostat.umn.edu

SUMMARY. We apply a linear mixed-effects model to multivariate failure time data. Computation of the regression parameters involves the Buckley–James method in an iterated Monte Carlo expectation-maximization algorithm, wherein the Monte Carlo E-step is implemented using the Metropolis–Hastings algorithm. From simulation studies, this approach compares favorably with the marginal independence approach, especially when there is a strong within-cluster correlation.

KEY WORDS: Buckley–James method; Generalized estimating equations; Least squares; Metropolis–Hastings algorithm; Monte Carlo expectation-maximization; Restricted maximum likelihood estimation.

1. Introduction

In a litter-matched tumorigenesis study (Lee, Wei, and Ying, 1993), one of the three rats in each litter was drug-treated, whereas the other two were controls. The aim of this study is to investigate whether there is any treatment effect in shortening the time of tumor appearance. This can be accomplished through regression analyses. However, litter effects imply that the failure times (i.e., tumor appearance times) of the rats in the same litter are correlated. In assessing the effects of the drug treatment on failure times, we need to consider appropriately the correlation among the failure times.

Two regression models have enjoyed tremendous popularity and successes in failure time analysis: the Cox proportional hazards model, wherein the failure time is modeled indirectly through its hazard function, and the accelerated failure time (AFT) model, in which a covariate acts to speed up or slow down the expected time to failure. In the logarithmic scale, the AFT model becomes a linear model. These models have been well studied when the observed failure times are independent (Kalbfleisch and Prentice, 1980; Louis, 1981; Tsiatis, 1990). However, in many situations the failure times are dependent, as mentioned in the above tumorigenesis study.

One current approach in the AFT model, the marginal independence approach, is to ignore the dependence in estimating the regression coefficient and then appropriately draw statistical inference, which has a large sample justification (Lee et al., 1993). However, ignoring the dependence generally does not take full advantage of the information in the data and is thus not efficient. To account for the dependence, we propose an AFT mixed-effects model wherein a random effect shared within a cluster induces correlations of failure times. Lee et al. (1993) mentioned this approach but did not pursue it because of the complexities of implementation. For the proportional hazards regression, an analogous mixed-effects (frailty) model has been investigated extensively in recent years (Vaupel, Manton, and Stallard, 1979), but a corresponding approach has not been developed for the AFT model.

2. The Marginal Independence Approach

Suppose T_{ij} is the logarithm of the failure time of the jth individual in cluster i, $j = 1, \ldots, n_i$ and $i = 1, \ldots, n$. Because of censoring, T_{ij} is not always observed. There is a censoring random variable C_{ij} independent of T_{ij}, and we observe $T^*_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = I(T_{ij} \le C_{ij})$ for each j and i, where $I(.)$ is the indicator function. The linear (i.e., AFT) model (Miller, 1981) is

$$T_{ij} = X_{ij}\beta + \epsilon_{ij},$$

where X_{ij} are covariates, and β is the unknown regression coefficient (vector) that is of primary interest. The mean-zero random vectors $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{i,n_i})'$, $i = 1, \ldots, n$, are independent of each other, but for multivariate failure time data, their components generally are dependent. Denote the distribution function of ϵ_{ij} by F_0. If we have a parametric assumption on F_0, the usual maximum likelihood method can be applied. In many biomedical studies, for robustness, nonparametric or semiparametric approaches are preferred. Hence, we restrict our discussion to estimating F_0 nonparametrically.

The marginal independence approach (Lee et al., 1993) estimates β by ignoring the possible correlation among the components of ϵ_i. If there is no censoring, we can estimate β by a least-squares estimation equation:

$$S_n(\beta) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} X_{ij}(T_{ij} - X_{ij}\beta) = 0.$$

Denote the residuals as $e_{ij}(\beta) = T_{ij} - X_{ij}\beta$. For censored data, some T_{ij} are not observed, and a natural choice is to replace $e_{ij}(\beta)$ by $r_{ij}(\beta) = E(T_{ij} \mid T^*_{ij}, \delta_{i,j}, X_{ij}, \beta) - X_{ij}\beta$, which depends on the unknown F_0. Buckley and James (1979) suggested an iterative method: (0) give an initial estimate $\tilde{\beta}$; (1) estimate F_0 from $(T_{ij}^* - X_{ij}\tilde{\beta}, \delta_{ij})$ by the Kaplan–Meier estimator $\hat{F}_0$; (2) estimate r_{ij} using $\hat{F}_0$ and then solve the equation $S_n(\beta) = 0$ to obtain a new estimate $\tilde{\beta}$; (3) repeat (1) and (2) until "convergence."

It is well known that the Buckley–James method may fail to converge and may eventually oscillate. One can take the average of the oscillating iterates as the solution (Miller, 1981). In the univariate case (i.e., $n_i = 1$ for all i), the consistency and the asymptotic normality of the Buckley–James estimator have been established under suitable conditions by Lai and Ying (1991). As a partial justification for the marginal independence approach in the multivariate case, Lee et al. (1993) have shown that $n^{1/2}S_n(\beta_0)$ is asymptotically normal with mean zero, where β_0 is the true value of the regression coefficient.

3. A Linear Mixed-Effects Model

3.1 *Model*

To take into account possible correlations among the components of ϵ_i, we adopt a linear mixed-effects model, as for uncensored data (Laird and Ware, 1982, for the Gaussian case):

$$T_i = X_i\beta + Z_i b_i + \epsilon_i, \tag{1}$$

where b_i is the random effect causing the possible correlation for the components of $T_i = (T_{i1}, \ldots, T_{i,n_i})'$. Now, ϵ_{ij} are assumed to be independent and identically distributed (i.i.d.). By default, Z_i is assumed to be a vector of 1s throughout the article except when specified otherwise in Section 6, wherein we explain why we need this assumption and outline a possible extension for our proposal.

We have two working parametric assumptions. First, as in the usual linear regression or Buckley–James estimation for censored data, we derive the least-squares estimation equation under the assumption of normally distributed errors, ϵ_{ij}. However, in other aspects, such as when imputing censored failure times, we estimate the distribution function F_0 of ϵ_{ij} by the nonparametric Kaplan–Meier estimate. Second, a normality assumption on b_i explicitly accounts for possible correlations among failure times to improve efficiency (as in modeling the covariance structure for generalized estimation equations [GEE], Liang and Zeger, 1986). Specifically, we assume that

$$b_i \overset{\text{i.i.d.}}{\sim} N(0, \tau^2), \quad \epsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2),$$

and that b_i and ϵ_{ij} are independent for $j = 1, \ldots, n_i$ and $i = 1, \ldots, n$.

The normal random-effect assumption may not influence the asymptotic validity of our method. Verbeke and Lesaffre (1997) have shown that for the mixed-effects model with complete data, the maximum likelihood estimates of the regression coefficient and the variance component of the error term obtained under normality assumptions for both the random effect and the error term are consistent and asymptotically normal, even though the true distribution of the random effect is not normal. In fact, the marginal likelihood score equations based on normality assumptions for both the random effect and the error distribution are normal equations, a special case of GEE (Liang and Zeger, 1986). By the general theory of GEE, the resulting estimates of the regression coefficients are consistent and asymptotically normal, even when neither the random effect nor the random error term has a normal distribution. For finite samples, Butler and Louis (1992) have shown that misspecifying the random-effect distribution has little effect on the fixed-effects estimates. However, in our context (i.e., with censoring), further investigation is needed.

3.2 *Estimation with Uncensored Data*

To facilitate our discussion, we first present our estimation procedure based on uncensored data and then adapt it to censored data. The procedure is based on using Monte Carlo expectation maximization (EM) (Tanner, 1996) to estimate β, and using restricted maximum likelihood (REML) to estimate the variance components τ^2 and σ^2 (Diggle, Liang, and Zeger, 1993, p. 67).

If we could observe b_i, $i = 1, \ldots, n$, then the complete data log likelihood is

$$\begin{aligned} L_c(\theta \mid T, X, Z, b) &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} \log f(T_{ij} - X_{ij}\beta - Z_{ij}b_i \mid T_{ij}, X_{ij}, Z_{ij}, b_i, \theta) \\ &\quad + \sum_{i=1}^{n} \log g(b_i \mid \theta), \end{aligned}$$

where $\theta = (\beta, \tau, \sigma)$, f and g are the normal densities for ϵ_{ij} and b_i, respectively, and $T = (T_1', \ldots, T_n')', X = (X_1', \ldots, X_n')'$, $Z = (Z_1', \ldots, Z_n')'$, $b = (b_1', \ldots, b_n')'$. The E-step in the EM computes the expectation $E(L_c)$ under the conditional distribution of b_i, given T_i (and current estimate $\hat{\theta}$). Under the assumptions of normal distributions for b_i and T_i, $b_i \mid T_i$ is also normal. Now, assuming we have a random sample $b_i^{(k)}$, $k = 1, \ldots, K$, from the conditional distribution $b_i \mid (T_i, \hat{\theta})$, we can approximate $E(L_c)$ by

$$Q(\beta, \tau, \sigma) = \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n_i} \left(T_{ij} - X_{ij}\beta - Z_{ij}b_i^{(k)}\right)^2 + C(\tau, \sigma), \tag{2}$$

where $C(\tau, \sigma)$ does not depend on β. Of course, the larger the K, the more accurate the Q. For the M-step, after maximizing Q with respect to β, we obtain

$$\hat{\beta} = \sum_{k=1}^{K} \hat{\beta}^{(k)}/K, \quad \hat{\beta}^{(k)} = (X'X)^{-1}X'\left(T - Zb^{(k)}\right).$$

Following the M-step, we estimate τ and σ by REML, from which we obtain the estimated covariance of T, $\hat{V} = V(\hat{\tau}, \hat{\sigma})$ (Diggle et al., 1993, p. 67). Using this estimated covariance, we readjust the estimated β as $\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}T$. We then return to the E-step with this $\hat{\beta}$; the above EM and REML steps are iterated until convergence to obtain the final estimate of β.

Note that the estimation of τ and σ is parametric in the sense that their asymptotic validity depends on the normality assumption. We do not use these estimates directly in drawing statistical inferences about β. Rather, in our numerical example, we use bootstrap methods.

3.3 *Extending to Censored Data*

Several issues arise with censored data. First, we cannot estimate $\beta^{(k)}$ or (τ, σ) directly. Second, with censored data $\{T_{ij}^*, \delta_{ij}\}$, the conditional distribution of b_i does not have a closed form. We now address these two issues.

As in Section 2, we use the Buckley–James method to estimate $\hat{\beta}^{(k)}$ iteratively for each $1 \leq k \leq K$. The basic idea is to impute T_{ij} from censored T_{ij}^* by the conditional expectations $T_{ij}^{(k)} = \hat{E}(T_{ij} \mid T_{ij}^*, \delta_{ij}, X_{ij}, Z_{ij}, b_i^{(k)}, \hat{\beta}^{(k)})$, using the Kaplan–Meier estimate $\hat{F}$ of the distribution function of ϵ_{ij}. This expectation depends on $\hat{\beta}^{(k)}$ and hence needs to be iterated. After we obtain $\hat{\beta} = \Sigma_{k=1}^K \hat{\beta}^{(k)}/K$, we again impute T_{ij} by $Y_{ij} = \hat{E}(T_{ij} \mid T_{ij}^*, \delta_{ij}, X_{ij}, b = 0, \hat{\beta})$ (using the Kaplan–Meier estimate $\hat{F}$). The REML estimate of (τ, σ) can then be obtained, from which we can readjust the estimate of β. The above two steps for estimating β and (τ, σ) (i.e., the algorithm described in Section 3.2) are iterated until convergence.

Specifically, our algorithm is as follows:

(0) Let $\hat{\beta}$ be an initial estimate.
(1) For $k = 1$ to K:
  (1.0) Let $\beta^{(k)} = \hat{\beta}$.
  (1.1) Generate $b^{(k)}$ from conditional distribution of $b \mid (T^*, \delta)$ (to be described later).
  (1.2) Impute T from T^* to obtain $T^{(k)}$ using $\beta^{(k)}$ and $b^{(k)}$.
  (1.3) Let $\beta^{(k)} = (X'X)^{-1}X'(T^{(k)} - Zb^{(k)})$.
  (1.4) Return to Step 1.2 until convergence.
(2) Let $\bar{\beta} = \Sigma_{k=1}^K \beta^{(k)}/K$.
(3) Impute from T^* to obtain Y using $\bar{\beta}$ and $b = 0$.
(4) Compute the REML estimates $(\hat{\sigma}, \hat{\tau})$, and hence $\hat{V} = V(\hat{\sigma}, \hat{\tau})$, using imputed Y.
(5) Readjust $\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y$.
(6) Go to Step 1 until convergence.

Because of the convergence issue with the Buckley–James method, the "convergence" here (Steps 1.4 and 6) implies either a true convergence of the parameter estimate or that a fixed number of iterations has been reached. The inner loop (Steps 1.2–1.4) is not necessary, but it may speed up the convergence of the outer loop and save CPU time, because the outer loop contains both the time-consuming Step 1.1 and the inner loop. Perhaps a more robust estimate of the covariance V in Step 4, such as the sandwich-type estimate (Huber, 1967), is more desirable. Step 5 is optional and may improve efficiency when the distribution of $Zb + \epsilon$ is close to a normal.

With censored data, we do not have a closed form for the distribution of $b_i \mid (T_i^*, \delta_i)$ in Step 1.1. However, the conditional distribution of $b_i \mid (T_i^*, \delta_i)$ is proportional to

$$\prod_{j=1}^{n_i} f(T_{ij}^* - X_{ij}\beta - Z_{ij}b_i)^{\delta_{ij}} \times \left[1 - F(T_{ij}^* - X_{ij}\beta - Z_{ij}b_i)\right]^{1-\delta_{ij}} g(b_i),$$

where f and g are the normal densities for ϵ_{ij} and b_i, respectively, and F is the cumulative distribution function for ϵ_{ij}, with current estimates of (β, τ, σ) as their parameters. Hence, we can use the Metropolis–Hastings algorithm (Carlin and Louis, 1995) to generate a sample from $b_i \mid (T_i^*, \delta_i)$ (with current estimates of (β, τ, σ)). This procedure is computationally intensive.

We describe an approximation to the conditional distribution $b_i \mid (T_i^*, \delta_i)$ when the percentage of censoring is not high. For simplicity, suppose $n_i = 2, i = 1, \ldots, n$. If neither T_{i1} nor T_{i2} is censored, the distribution of $b_i \mid (T_{i1}, T_{i2})$ is normal as before. When only one of them is censored, we use $b_i \mid T_{i1}$ (if T_{i1} is not censored), or $b_i \mid T_{i2}$ (otherwise) to approximate $b_i \mid (T_{i1}, T_{i2})$. Otherwise we use the current estimate of the unconditional distribution of b_i. However, with larger n_i or heavier censoring, the Metropolis–Hastings algorithm is generally preferred.

We denote the mean and covariance matrix (or variance) of the above approximation to the conditional distribution of $b_i \mid (T_i^*, \delta_i)$ as μ_i and Σ_i, respectively. In the Metropolis–Hastings algorithm, the transition probability function $q(x, y) = q(y)$ (from the current state x to the next state y) in the Markov chain is Gaussian, with mean μ_i and covariance matrix $c\Sigma_i$, where c is taken to be $2.4/d^{1/2}$ and d is the dimension of b_i (Gelman et al., 1995). In our simulations, for each Markov chain, we discarded the first 200 runs as "burn-in" period and used the next 200 runs.

3.4 *Statistical Inference*

Lee et al. (1993) suggested the use of $S_n(\tilde{\beta})$ (where $\tilde{\beta}$ is the estimate from the marginal independence approach) to draw inference about the true value β_0. The confidence interval for β_0 is obtained by inverting an asymptotic χ^2 statistic based on $S_n(\tilde{\beta})$. This procedure is justified by the asymptotic normal distribution of $n^{1/2}S_n(\beta_0)$. Hence, with a better estimate $\hat{\beta}$, inference based on $S_n(\hat{\beta})$ presumably would be more effective, but we will not pursue it here.

Another choice is to use bootstrap methods (Efron and Tibshirani, 1993), which are illustrated in our example. To conserve the correlation structure within the cluster, in our bootstrap the resampling unit is the cluster and not each individual failure time. This is similar to cross-validation in choosing smoothing parameters for nonparametric curve estimation with longitudinal data (Rice and Silverman, 1991). A rigorous justification for the bootstrap approach here may depend on the asymptotic properties of $\hat{\beta}$. As shown in Section 4, the bootstrap works well in estimating the standard error of $\hat{\beta}$.

4. Simulations

We now present our simulation results under different configurations. The data were generated as in model (1) with

$$T_{ij} = X_{ij} + b_i + \epsilon_{ij},$$

where $j = 1, 2, 3$; $i = 1, \ldots, n$. Therefore, the true value of β is always 1. We examined sample sizes $n = 50$ and $n = 100$. The covariates X_{ij} are always i.i.d. from a uniform distribution between 0 and 1. We used the following combinations of the normal and non-normal distributions for the random-effects b_i and random errors ϵ_{ij} to investigate the dependence of our method on the normality assumption:

(1) *Normal (N):* $N(0, \sigma^2)$ with mean 0, standard deviation σ, and $\sigma = 1$ or 2.

(2) *Contaminated normals* (CN)*:* a mixture of two normals, $0.6N(0,1)+0.4N(0,3^2)$.
(3) *Extreme value distribution* (Ex)*:* with a transformation such that its mean and variance are 0 and 1, respectively.
(4) *Two-point distribution* ($D2$)*:* equal to 1 and -1 with an equal probability 0.5.
(5) *Three-point distribution* ($D3$)*:* equal to -1, 0.5, and 1.5 with probabilities 0.5, 0.25, and 0.25, respectively.

We used two levels of censoring: low (around 20%) and high (around 60%). By changing the variances of b_i and ϵ_{ij}, we had various within-cluster correlations.

We fitted a model with an intercept term. In our simulations, we chose $K = 200$ for Monte Carlo EM, which seemed large enough for our sample size. The results are presented in Table 1. The mixed-effects model often yields better results than the marginal independence approach in terms of smaller mean squared errors, primarily because of its smaller variability. In particular, both estimators have, at most, very small bias. This (almost) unbiasedness is surprising because in the Buckley–James method, the Kaplan–Meier survival estimate is forced to be 0 at the largest observed value because the Kaplan–Meier estimate is undefined if the largest observation is censored; this may give an underestimated imputed value for a censored observation.

Not surprisingly, when the normality assumptions on the random error and the random effect hold, the mixed-effects model is substantially more efficient relative to the marginal independence approach. Efficiency is more than twice that of the marginal independence approach when the within-cluster correlation is 0.8. Even when the normality assumption is violated, the mixed-effects model still has an edge. If we increase the sample size, the performance of both estimators improves and their difference narrows. This suggests large-sample consistency for both estimators.

To assess the inferential performance of the bootstrap methods in the mixed-effects approach, we performed a simulation study to investigate the performance of the bootstrap standard error of the estimated regression coefficient. The setup is the same as for the above case. To save CPU time, we used only 50 bootstrap replications for each of the first 200 independent samples in three representative configurations. The results are shown in Table 2, from which we can see that the bootstrap works well in estimating the standard error.

5. An Example

For illustration, we apply the methods to the litter-matched tumorigenesis data set (Lee et al., 1993). Only 36 failures were observed among 150 rats (in 50 litters); all others were right-censored because of death. We consider only the drug indicator covariate: $x_{ij} = 0$ if the jth rat in the ith litter was drug-treated; otherwise $x_{ij} = 1$. We investigated whether there was any treatment effect in shortening the time to tumor appearance. Note that in Table 3, our point estimate of β from the marginal independence approach is different from the value of 0.165 obtained by Lee et al. (1993). This difference may be caused by different handling of ties in the Kaplan–Meier estimator or by large censored observations in the Buckley–James method. We note that our point estimate is closer to the value of 0.115 produced from the Wilcoxon log-rank approach in Lee et al. (1993). To draw an inference, we applied two bootstrap methods (with 1000 bootstrap replications) to construct the percentile and BC$_a$ (bias corrected and accelerated) confidence intervals, which have first-order and second-order accuracy, respectively (Efron and Tibshirani, 1993). The bootstrap CIs are shifted to the left when compared with those obtained in Lee et al. (1993) by inverting an asymptotic χ^2 statistic based on $S_n(\tilde{\beta})$ (i.e., 90% CI $[-0.01, 0.36]$ and 95% CI $[-0.03, 0.38]$).

Table 1

Monte Carlo mean, standard deviation, and mean squared error (MSE) of the estimates of the regression coefficient $\beta_0 = 1$ from 1000 simulations, by the marginal independence and mixed-effects approaches. The maximum Monte Carlo standard error of the estimates of β_0 is 0.03.

Configuration					Marginal independence			Mixed-effects			RE[c]
n	r[a]	cen.[b]	b	ϵ	Mean($\tilde{\beta}$)	SD($\tilde{\beta}$)	MSE	Mean($\hat{\beta}$)	SD($\hat{\beta}$)	MSE	(%)
50	0.50	l	N	N	1.01	0.42	0.174	1.00	0.35	0.119	68
	0.80	l	N	N	1.00	0.65	0.423	1.00	0.39	0.149	35
	0.33	l	N	CN	1.00	0.70	0.485	1.02	0.65	0.425	88
	0.50	l	N	Ex	1.01	0.42	0.172	1.00	0.36	0.129	75
	0.50	h	N	N	1.00	0.51	0.258	0.98	0.46	0.209	81
	0.80	h	N	N	1.03	0.74	0.546	0.99	0.47	0.223	41
	0.33	h	N	CN	0.98	0.83	0.686	0.98	0.80	0.643	94
	0.50	h	N	Ex	1.01	0.53	0.276	1.00	0.47	0.223	81
	0.50	h	D2	N	0.99	0.49	0.243	0.97	0.44	0.191	79
	0.50	h	D3	CN	1.01	0.94	0.879	0.99	0.82	0.671	76
100	0.50	h	D2	N	1.02	0.36	0.130	1.01	0.32	0.104	80
	0.50	h	D3	CN	0.97	0.67	0.444	0.97	0.60	0.357	81

[a] Correlation coefficient within clusters.
[b] Censoring level (l = low and h = high).
[c] This is the ratio of the two MSEs.

Table 2
Performance of the bootstrap standard error of the regression coefficient estimate from the mixed-effects approach, $SE(\hat{\beta})$. We use mean$(\hat{\theta})$ and $SD(\hat{\theta})$ to denote the (Monte Carlo) mean and standard deviation of a statistic θ from 200 simulations. The maximum Monte Carlo standard error of the bootstrap estimates $SE(\hat{\beta})$ is 0.01.

Configuration					$SD(\hat{\beta})$	Mean$(SE(\hat{\beta}))$	$SD(SE(\hat{\beta}))$
50	0.50	l	N	N	0.35	0.34	0.06
	0.33	l	N	CN	0.65	0.65	0.10
	0.50	h	D3	CN	0.82	0.84	0.18

We conclude that there is no strong statistical evidence to support a treatment effect.

6. Remarks

A referee and an associate editor raised our attention to an unpublished work by Hornsteiner and Hamerle (1996), wherein a multivariate extension of the univariate AFT model is also proposed. Their approach is a combination of the GEE approach (Liang and Zeger, 1986) and the Buckley–James method. It can be regarded as a more general marginal approach. We review and comment on it briefly.

If we can observe the exact failure times T, the GEE is a set of normal equations:

$$\sum_{i=1}^{n} X_i' V_i^{-1}(T_i - X_i\beta) = 0,$$

where V_i is a working covariance matrix of T_i. For instance, according to the random-effects model, $V_i = V(\sigma, \tau) = \sigma^2 I + \tau^2 11'$. With censoring, we observe T^* instead of T. Using the Buckley–James method, based on our current estimate $\hat{\beta}_{(u)}$ in the uth iteration, we can impute T using the conditional expectations as before. Let us denote the imputed failure times as Y. Hornsteiner and Hamerle proposed to update the estimate of β as

$$\hat{\beta}_{(u+1)} = \hat{\beta}_{(u)} + \left(\sum_{i=1}^{n} X_i'\hat{V}_i^{-1}\frac{\partial Y_i}{\partial \beta}\right)^{-1} \times \left(\sum_{i=1}^{n} X_i'\hat{V}_i^{-1}\left(Y_i - X_i\hat{\beta}_{(u)}\right)\right), \qquad (3)$$

and estimates in the working covariance V_i were also updated accordingly ($u = 1, 2, \ldots$ until convergence). They also gave a sandwich estimator of the covariance matrix of $\hat{\beta} = \hat{\beta}_{(\infty)}$:

$$\widehat{\text{cov}}(\hat{\beta}) = \left(\sum_{i=1}^{n} X_i'\hat{V}_i^{-1}\frac{\partial Y_i}{\partial \beta}\right)^{-1} \times \left(\sum_{i=1}^{n} X_i'\hat{V}_i^{-1}\widehat{\text{cov}}(Y_i)\hat{V}_i^{-1}X_i\right) \times \left(\sum_{i=1}^{n} X_i'\hat{V}_i^{-1}\frac{\partial Y_i}{\partial \beta}\right)^{-1}, \qquad (4)$$

where $\widehat{\text{cov}}(Y_i) = (Y_i - X_i\hat{\beta})(Y_i - X_i\hat{\beta})'$.

It is not clear to us as to how the updates in (3) are derived. More importantly, it appears difficult to calculate $\partial Y_i/\partial\beta$. Note that $Y_{ij} = \delta_{ij}T_{ij}^* + (1-\delta_{ij})\hat{E}(T_{ij} \mid T_{ij} > T_{ij}^*)$ is a random variable and may depend on all other (T_{ij}^*, δ_{ij})'s. However, the idea of Hornsteiner and Hamerle is novel and important. Based on their idea and GEE, we propose to use the following estimating equations:

$$\sum_{i=1}^{n} X_i' V_i^{-1}(Y_i - X_i\beta) = 0$$

from which an explicit solution of β is

$$\hat{\beta}_{(u+1)} = \left(\sum_{i=1}^{n} X_i'\hat{V}_i^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'\hat{V}_i^{-1}Y_i\right). \qquad (5)$$

Based on $\hat{\beta}_{(u+1)}$, we can use the method of moments to update the parameters in V_i to obtain new $\hat{V}_i$ and impute T_i to obtain new Y_i. This process is iterated until convergence. A sandwich estimate of the covariance matrix can be also derived:

$$\widehat{\text{cov}}(\hat{\beta}) = \left(\sum_{i=1}^{n} X_i'\hat{V}_i^{-1}X_i\right)^{-1} \times \left(\sum_{i=1}^{n} X_i'\hat{V}_i^{-1}\widehat{\text{cov}}(Y_i)\hat{V}_i^{-1}X_i\right) \times \left(\sum_{i=1}^{n} X_i'\hat{V}_i^{-1}X_i\right)^{-1}, \qquad (6)$$

where $\widehat{\text{cov}}(Y_i) = (Y_i - X_i\hat{\beta})(Y_i - X_i\hat{\beta})'$. But (6) may be biased since it is based on a single imputation.

Table 3
Estimates and confidence intervals of the regression coefficient from the tumorigenesis data set

Method	Point estimate	Level (%)	Percentile CI	BC_a CI
Marginal independence	0.102	90	$(-0.069, 0.267)$	$(-0.106, 0.257)$
		95	$(-0.093, 0.344)$	$(-0.178, 0.264)$
Mixed effects	0.103	90	$(-0.068, 0.271)$	$(-0.046, 0.292)$
		95	$(-0.095, 0.327)$	$(-0.077, 0.362)$

Because of the close connection between the mixed-effects model and the GEE in the linear regression (for complete data), the performance of the marginal approach after accommodating the within-cluster correlation and that of the mixed-effects model approach is expected to be close. The simulation results of Hornsteiner and Hamerle are consistent with ours; taking the correlation appropriately into account for the multivariate AFT model may improve the estimation efficiency.

However, in imputing from censored data there is an important difference between our method proposed in Section 3.3 and the marginal approach. In the marginal approach, when we calculate $\hat{E}(T_{ij} \mid T_{ij} > T^*_{ij})$, it is based on the Kaplan–Meier estimate of the distribution function of $Z_{ij}b_i + \epsilon_{ij}$. This explains why we need to assume that all Z_{ij} are 1s (or 0s trivially) to make sure that all $Z_i b_i + \epsilon_i$ have a common marginal distribution, which is a prerequisite of any marginal approach (Lee et al., 1993; Hornsteiner and Hamerle, 1996). This implies that the Kaplan–Meier estimate in the marginal approach is based on dependent residuals $(T^* - X\beta, \delta)$ (Ying and Wei, 1994). In our imputation scheme in Step 1.2, it is based on the estimated distribution of i.i.d. ϵ_{ij}, after conditioning on b. This was our main motivation for using the Monte Carlo EM. It also appears possible to extend our proposal to accommodate more general Z_i in model (1). Steps 0–2 are the same, but we modify Step 3 as follows: Compute the ML estimates $(\hat{\sigma}, \hat{\tau})$ based on (2). We also delete Steps 4 and 5. Future studies are needed in this direction.

7. Discussion

We have shown by simulation that a linear mixed-effects model may improve the performance of the Buckley–James estimator of the regression coefficient in the multivariate accelerated failure time model when compared with the marginal independence approach. This improvement is more obvious when there is a strong correlation among failure times. Although we adopt two working normality assumptions, we expect our method, like the Buckley–James method for univariate data, to be essentially semiparametric because of the nonparametric least-squares estimation and the Kaplan–Meier estimator.

Another approach for the univariate AFT model is based on linear rank statistics (Louis, 1981; Tsiatis, 1990). Ritov (1990) has shown its asymptotic equivalence to the Buckley–James estimator. However, there is still no efficient computational algorithm (Tsiatis, 1990), except for a random search based on simulated annealing (Lin and Geyer, 1992). Although conceptually it is possible to extend this approach with a mixed-effects model, further studies are needed in this direction.

Acknowledgements

The authors wish to thank John Connett for helpful suggestions on an earlier draft. The authors are also grateful to the referee, the associate editor, and the editor for their constructive comments, particularly for bringing to our attention an earlier reference.

Résumé

Nous appliquons un modèle linéaire à effets mixtes à des données de durée multivariées. Le calcul des paramètres de régression utilise la méthode de Buckley-James dans un algorithme de Monte-Carlo itératif de type EM., où l'étape E de l'algorithme de Monte-Carlo est implémentée en utilisant un algorithme de Metropolis-Hastings. A partir d'études de simulation, cette approche se compare favorablement avec l'approche d'indépendance marginale, particulièrement quand il existe une forte corrélation intra-grappe.

References

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66,** 429–436.

Butler, S. M. and Louis, T. A. (1992). Random effects models with non-parametric priors. *Statistics in Medicine* **11,** 1981–2000.

Carlin, B. P. and Louis, T. A. (1995). *Bayes and Empirical Bayes Methods for Data Analysis.* London: Chapman and Hall.

Diggle, P. J., Liang, K.-Y., and Zeger, S. L. (1993). *Analysis of Longitudinal Data.* New York: Oxford.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* London: Chapman and Hall.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. R. (1995). *Bayesian Data Analysis.* London: Chapman and Hall.

Hornsteiner, U. and Hamerle, A. (1996). A combined GEE/Buckley–James method for estimating an accelerated failure time model of multivariate failure times. Discussion Paper 47, Ludwig–Maximillians Universitat, Munchen. Also available from `http://stat.uni-muenchen.de/sfb386/publikation.html`.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the 5th Berkeley Symposium* **1,** 221–233.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Lai, T. L. and Ying, Z. (1991). Large sample theory of a modified Buckley–James estimator in regression analysis with censored data. *Annals of Statistics* **19,** 1370–1402.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38,** 963–974.

Lee, C. W., Wei, L. J., and Ying, Z. (1993). Linear regression analysis for highly stratified failure time data. *Journal of the American Statistical Association* **88,** 557–565.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Lin, D. Y. and Geyer, C. J. (1992). Computational methods for semi-parametric linear regression with censored data. *Journal of Computational and Graphical Statistics* **1,** 77–90.

Louis, T. A. (1981). Non-parametric analysis of an accelerated failure time model. *Biometrika* **68,** 381–390.

Miller, R. (1981). *Survival Analysis.* New York: Wiley.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Section B* **53,** 233–243.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Annals of Statistics* **18,** 303–328.

Tanner, M. A. (1996). *Tools for Statistical Inference.* New York: Springer.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics* **18,** 354–372.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty and the dynamics of mortality. *Demography* **16,** 439–454.

Verbeke, G. and Lesaffre, E. (1997). The effects of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis* **23,** 541–556.

Ying, Z. and Wei, L. J. (1994). The Kaplan–Meier estimate for dependent failure time observations. *Journal of Multivariate Analysis* **50,** 17–29.

Received March 1998. *Revised April* 1999.
Accepted May 1999.