Biometrika (2000), **87,** 4, *pp.* 849–866

Printed in Great Britain

Covariate measurement errors in frailty models for clustered survival data

BY YI LI

Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, Massachusetts 02115, U.S.A.

yili@jimmy.harvard.edu

AND XIHONG LIN

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.

xlin@umich.edu

SUMMARY

We propose a new class of frailty measurement error models for clustered survival data when covariates are measured with error. We show that the induced hazard function conditional on the observed covariates also follows a frailty model but of a more complicated form. We study the asymptotic bias in regression coefficients and variance components when measurement error is ignored, and the impact of censoring on this asymptotic bias. We show that the naive estimator of the regression coefficient is attenuated and the naive estimator of the variance component is inflated when measurement error is ignored. As the censoring proportion increases, the asymptotic bias in the former becomes larger, but the asymptotic bias in the latter interestingly becomes smaller. We develop a structural approach for parameter estimation using the nonparametric maximum likelihood method, where the baseline hazard is estimated nonparametrically. We prove model identifiability and the existence of the nonparametric maximum likelihood estimators. An EM algorithm is developed for calculating the nonparametric maximum likelihood estimates. The method is applied to the western Kenya parasitaemia data and its performance is evaluated through simulations.

Some key words: Asymptotic bias; EM algorithm; Frailty measurement error model; Identifiability; Nonparametric maximum likelihood estimation; Variance components.

1. INTRODUCTION

Frailty models (Clayton & Cuzick, 1985) are increasingly popular for analysing clustered survival data, where frailties or random effects often enter into the baseline hazard multiplicatively to model the correlation among observations within the same cluster. A challenge in fitting frailty models is that the standard partial likelihood approach is not applicable and one has to estimate the regression coefficients, the variance components and the nonparametric baseline hazard simultaneously. Nielsen et al. (1992) discussed the use of the EM algorithm for estimation in frailty models, and Murphy (1994, 1995) and Parner

(1998) studied the theoretical properties of the models. For a review of frailty models, see Oakes (1989).

A common problem in survival analysis is the presence of covariate measurement error. For example, in clinical trials many biomarkers, such as blood pressure (Carroll, Ruppert & Stefanski, 1995, Ch. 1) and CD4 counts (Tsiatis, De Gruttola & Wulfsohn, 1995), are subject to measurement error. In nutritional studies, fat intake is often measured with error (Carroll et al., 1995, Ch. 1). For independent survival data, several authors have studied the measurement error problem within the proportional hazard model framework. Prentice (1982) and Pepe, Self & Prentice (1989) showed that the induced hazard function conditional on the observed covariates is also a multiplicative hazard model, but of a complicated form. Hughes (1993) studied the asymptotic bias of the partial likelihood estimator when measurement error is ignored and showed that the naive estimator of the regression coefficient is attenuated and that increasing the level of censoring results in a more attenuated naive estimator. Hu, Tsiatis & Davidian (1998) compared several estimation procedures.

However, little has been done for clustered censored survival data with measurement error in covariates. In this paper, we propose a new class of frailty measurement error models, which model correlation using frailties and measurement error simultaneously, for clustered survival data. We state the model in § 2. We then study in § 3 the impact of covariate measurement error in frailty models by performing asymptotic bias analysis when measurement error is ignored and by examining how censoring and cluster size influence the effect of measurement error. We investigate in § 4 the theoretical properties of the model including model identifiability and the existence of the nonparametric maximum likelihood estimators. We develop in § 5 an EM algorithm for calculating the nonparametric maximum likelihood estimates. We evaluate the performance of the proposed method through simulations in § 6 and apply it to the western Kenya parasitaemia data in § 7, followed by discussion in § 8.

2. THE FRAILTY MEASUREMENT ERROR MODEL

We assume that survival times are subject to right censoring and that censoring is noninformative. Let $t_{ij} = \min(v_{ij}, c_{ij})$ be the observed time for the jth subject ($j = 1, \ldots, n_i$) in the ith cluster, such as a family, ($i = 1, \ldots, m$), where v_{ij} is the true survival time and c_{ij} is the censoring time. Let $\delta_{ij} = I(v_{ij} \leqslant c_{ij})$ be the noncensoring indicator, which takes value 1 if a failure is observed and 0 otherwise. Let X_{ij} be the true unobserved covariate and W_{ij} be the observed X_{ij}-related error-prone covariate. For simplicity, we here assume that X_{ij} is a scalar. Let Z_{ij} ($p \times 1$) be the other covariates which are accurately measured. Conditional on the cluster-specific frailty vector b_i ($q \times 1$), the observations (t_{ij}, δ_{ij}) are independent with conditional proportional hazard functions

$$\lambda_{ij}(t \mid X_{ij}, Z_{ij}, b_i) = \lambda_0(t) e^{X_{ij}\beta_x + Z'_{ij}\beta_z + B'_{ij}b_i}, \tag{1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard, β_x and β_z ($p \times 1$) are fixed effects, (X_{ij}, Z_{ij}) and B_{ij} ($q \times 1$) are covariates associated with the fixed effects and the frailty respectively, and B_{ij} is measured without error. Here we assume that the frailties b_i are independent of X_{ij} and Z_{ij} and are independent and identically distributed as $N\{0, D(\theta)\}$, where $D(\theta)$ is a positive definite matrix depending on θ, and θ is an $l \times 1$ vector of variance components.

Several authors considered the single frailty case where $B_{ij} = 1$ and assumed that the scalar frailty b_i follows a log-Gamma distribution, or equivalently $\exp\{b_i\}$ follows a

Gamma distribution (Clayton & Cuzick, 1985; Murphy, 1995). We here assume that the frailty vector b_i has a multivariate normal distribution. Advantages of this assumption are that it can easily accommodate multiple frailties and is convenient for introducing a measurement error model for X_{ij}. Note that the log-Gamma distribution and the normal distribution are often similar.

Define $T_i = (t_{i1}, \ldots, t_{in_i})'$, $\Delta_i = (\delta_{i1}, \ldots, \delta_{in_i})'$ and X_i, W_i, Z_i and B_i similarly. The integrated likelihood of (T_i, Δ_i) given X_i and Z_i in the ith cluster is

$$L_i(T_i, \Delta_i | X_i, Z_i) = \int \prod_{j=1}^{n_i} \{\lambda_0(t_{ij}) e^{X_{ij}\beta_x + Z'_{ij}\beta_z + B'_{ij}b_i}\}^{\delta_{ij}} e^{-\Lambda_0(t_{ij}) \exp(X_{ij}\beta_x + Z'_{ij}\beta_z + B'_{ij}b_i)} \, d\Phi(b_i; D), \tag{2}$$

where $\Phi(.; D)$ is the distribution function of $N\{0, D(\theta)\}$ and $\Lambda_0(t)$ is the cumulative baseline hazard. Note that (2) does not have a closed form.

The frailty measurement error model is completed by adding measurement error to X_{ij}:

$$W_{ij} = X_{ij} + U_{ij}, \tag{3}$$

where the measurement error terms U_{ij} are independent and identically distributed as $N(0, \sigma_u^2)$. We here assume that measurement error is non-differential, that is $L(T_i, \Delta_i | X_i, W_i, Z_i) = L(T_i, \Delta_i | X_i, Z_i)$, which implies that, conditional on the true unobserved covariate X_i, the observed covariate W_i does not contain additional information about (T_i, Δ_i). The likelihood of the observed data (T_i, Δ_i, W_i) is

$$L(T_i, \Delta_i, W_i | Z_i) = \int L(T_i, \Delta_i | X_i, Z_i) L(W_i | X_i, Z_i) L(X_i | Z_i) \, dX_i, \tag{4}$$

where $L(W_i | X_i, Z_i) = L(W_i | X_i)$ under (3), and $L(X_i | Z_i)$ is the likelihood of X_i.

In the classical measurement error literature, it is common to assume the unobserved X_{ij} to be independent and normally distributed (Carroll et al., 1995, p. 7). However, for clustered data, the X_{ij} within the same cluster are likely to be correlated. Hence we consider a linear mixed model (Laird & Ware, 1982) for the unobserved X_i:

$$X_i = 1_i \mu_x + Z_i \mu_z + A_i a_i + \varepsilon_i, \tag{5}$$

where 1_i is an $n_i \times 1$ vector of ones, (μ_x, μ_z) is a vector of fixed effects and a_i is a vector of random effects following $N(0, \Sigma_{x\mu})$, A_i is the design matrix associated with a_i, and ε_i is a residual vector independent of a_i and following $N(0, \sigma_x^2 I_i)$. Here I_i is an $n_i \times n_i$ identity matrix. A special case of model (5) is the fixed effects model,

$$X_i = 1_i \mu_x + Z_i \mu_z + \varepsilon_i, \tag{6}$$

which assumes that the X_{ij} are independent and is appropriate when the X_{ij} are cluster-level covariates.

Define $\Sigma_{i,xx} = \mathrm{cov}(X_i) = A_i \Sigma_{x\mu} A'_i + \sigma_x^2 I_i$ and $\Gamma_i = \Sigma_{i,xx}(\Sigma_{i,xx} + \sigma_u^2 I_i)^{-1}$. Using equations (3) and (5), one can easily calculate the conditional distribution of X_i given W_i and write $X_i = (I_i - \Gamma_i)(1_i \mu_x + Z_i \mu_z) + \Gamma_i W_i + \tilde{b}_i$, where $\tilde{b}_i$ can be shown to be independent of W_i conditional on Z_i and follows $N\{0, (I_i - \Gamma_i)\Sigma_{i,xx}\}$. Some calculations then show that, conditional on the observed covariates (W_i, Z_i), the outcomes (T_i, Δ_i) also follow a frailty model, but having a much more complicated form:

$$\begin{aligned}\lambda_{ij}(t | W_i, Z_i, b_i, \tilde{b}_i) = \lambda_0(t) \exp\{&e'_{ij}(I_i - \Gamma_i) 1_i \mu_x \beta_x + e'_{ij} \Gamma_i W_i \beta_x + e'_{ij}(I_i - \Gamma_i) Z_i \mu_z \beta_x \\ &+ Z'_{ij}\beta_z + B'_{ij} b_i + \beta_x e'_{ij} \tilde{b}_i\},\end{aligned} \tag{7}$$

where $(b_i, \tilde{b}_i)$ are frailties and are independent, and e_{ij} is an $n_i \times 1$ indicator vector whose jth element is 1 and 0 otherwise. It is difficult to fit the frailty model (7) because of its complicated form.

Equation (7) suggests that ignoring measurement error by naively replacing the X_{ij} with the W_{ij} in (1) results in misspecification of both the fixed effects and the frailty structures. Our first goal in this paper is to understand the impact of measurement error on parameter estimation in frailty measurement error models; see § 3. Specifically, we study the asymptotic biases in regression coefficient and variance component estimators when measurement error is ignored, and how cluster size and censoring influence these biases. Our second goal is to develop nonparametric maximum likelihood estimation under the structural model in equations (1), (3) and (5), where the baseline hazard $\lambda_0(t)$ is estimated nonparametrically; see § 4–5.

3. Asymptotic bias analysis

3·1. *Preamble*

Wang et al. (1998) studied the asymptotic bias in naive regression when measurement error is ignored within the framework of generalised linear mixed measurement error models. They restricted their attention to clustered Gaussian, Binomial and Poisson outcomes. However, little is known about the bias in naive regression in frailty measurement error models for survival outcomes, where extra complications arise. First, instead of modelling the mean function we model the hazard function, and hence the impact of measurement error is not clear and is more complicated. Secondly, survival data involve censoring. It is of substantial interest to investigate how censoring affects the asymptotic bias in naive regression when measurement error is ignored.

Our asymptotic bias analysis detailed below shows that ignoring measurement error results in an attenuated regression coefficient estimator and an inflated variance component estimator. When the unobserved covariate X is a subject-level covariate, as the censoring proportion increases, the asymptotic bias in the naive regression coefficient estimator increases, but the asymptotic bias in the naive variance component estimator decreases.

3·2. *The specific model considered in bias analysis*

Asymptotic bias analysis under the general model (1) is difficult, and here we concentrate on a simple but representative frailty measurement error model. Specifically, we assume a constant cluster size $n_i = n$ and consider a simple random intercept frailty model (Clayton & Cuzick, 1985) with a single covariate X:

$$\lambda_{ij}(t \mid X_{ij}, b_i) = \lambda_0(t) e^{\beta_x X_{ij} + b_i}, \tag{8}$$

where $b_i \sim N(0, \theta)$. The likelihood of (8) is

$$L(T_i, \Delta_i \mid X_i) = \int \prod_{j=1}^{n} \{\lambda_0(t_{ij}) e^{\beta_x X_{ij} + b_i}\}^{\delta_{ij}} e^{-\Lambda_0(t_{ij}) \exp(\beta_x X_{ij} + b_i)} \, d\Phi(b_i; \theta). \tag{9}$$

A simple model for X_{ij} is a random intercept model, which allows possible correlation among the X_{ij} within the same cluster:

$$X_{ij} = \mu_x + a_i + \varepsilon_{ij}, \tag{10}$$

where $a_i \sim N(0, \sigma^2_{x\mu})$ and $\varepsilon_{ij} \sim N(0, \sigma^2_x)$. Note that (10) is a special case of equation (5).

Using equations (3) and (10), one can easily calculate the conditional distribution of $X_i|W_i$. The frailty model (7) of the observed data $(T_i, \Delta_i|W_i)$ can then be simplified as

$$\lambda_{ij}(t|W_i, b_i^*, b_{ij}^{**}) = \lambda_0(t)e^{\beta_0^* + \beta_x^* W_{ij} + \beta_2^* \bar{W}_{i\cdot} + b_i^* + b_{ij}^{**}}, \tag{11}$$

where

$$\beta_0^* = (1-\alpha)\tilde{\alpha}\mu_x\beta_x, \quad \beta_x^* = \alpha\beta_x, \quad \beta_2^* = (1-\alpha)(1-\tilde{\alpha})\beta_x,$$

$$\alpha = \frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2}, \quad \tilde{\alpha} = \frac{\sigma_x^2+\sigma_u^2}{\sigma_x^2+\sigma_u^2+n\sigma_{x\mu}^2}, \quad \bar{W}_{i\cdot} = n^{-1}\sum_{j=1}^{n} W_{ij},$$

b_i^* is a cluster-level frailty following $N(0, \theta + n^{-1}\beta_x\beta_2^*\sigma_u^2)$, b_{ij}^{**} is a subject-level frailty following $N(0, \gamma = \alpha\sigma_u^2\beta_x^2)$, and b_i^* and b_{ij}^{**} are independent of each other and of W_i. Some calculations show that equation (11) can be rewritten as a single frailty model; compare Prentice (1982, eqn (3)), with

$$\lambda_{ij}(t|W_i, b_i^*) = \lambda_0(t)e^{\beta_0^* + \beta_x^* W_{ij} + \beta_2^* \bar{W}_{i\cdot} + b_i^*} E(e^{b_{ij}^{**}}|t_{ij} \geqslant t, W_i, b_i^*), \tag{12}$$

where $E(e^{b_{ij}^{**}}|t_{ij} \geqslant t, W_i, b_i^*)$ takes a complicated form involving $\Lambda_0(t)$:

$$E(e^{b_{ij}^{**}}|t_{ij} \geqslant t, W_i, b_i^*) = \frac{\int e^{b_{ij}^{**} - \Lambda_0(t)\exp(\beta_0^* + \beta_x^* W_{ij} + \beta_2^* \bar{W}_{i\cdot} + b_i^* + b_{ij}^{**})} d\Phi(b_{ij}^{**}; \gamma)}{\int e^{-\Lambda_0(t)\exp(\beta_0^* + \beta_x^* W_{ij} + \beta_2^* \bar{W}_{i\cdot} + b_i^* + b_{ij}^{**})} d\Phi(b_{ij}^{**}; \gamma)}.$$

The naive estimator is the estimator under the model that ignores measurement error by simply replacing X_{ij} by W_{ij} in (8). Thus

$$\lambda_{ij,\text{naive}}(t|W_{ij}, b_{i,\text{naive}}) = \lambda_{0,\text{naive}}(t)e^{W_{ij}\beta_{x,\text{naive}} + b_{i,\text{naive}}}, \tag{13}$$

where $b_{i,\text{naive}} \sim N(0, \theta_{\text{naive}})$. A comparison of (11) and (13) suggests that the naive model misspecifies both the fixed effect structure and the frailty structure. This makes asymptotic bias analysis complicated and closed-form expressions are often not available.

When the X_{ij} are cluster-level covariates, so that $X_{ij} = X_i$, asymptotic bias analysis is simple and the naive estimators $\{\beta_{x,\text{naive}}, \theta_{\text{naive}}, \lambda_{0,\text{naive}}(t)\}$ have a closed form. In this case, it is more reasonable to assume that each X_i follows $N(\mu_x, \sigma_x^2)$ instead of the random intercept model (10). The frailty model of the observed data $(T_i, \Delta_i|W_i)$ hence takes the simpler form

$$\lambda_{ij}(t|b_i^{**}) = \lambda_0(t)e^{\beta_0^{**} + \beta_x^* W_{ij} + b_i^{**}},$$

where $\beta_0^{**} = (1-\alpha)\mu_x\beta_x$ and $b_i^{**} \sim N(0, \theta + \alpha\sigma_u^2\beta_x^2)$. The naive model (13) becomes

$$\lambda_{ij,\text{naive}}(t|W_i, b_{i,\text{naive}}) = \lambda_{0,\text{naive}}(t)e^{W_i\beta_{x,\text{naive}} + b_{i,\text{naive}}}.$$

Since the two models take the same form, one can use parameter correspondence to calculate the asymptotic limits of the naive estimators as follows:

$$\beta_{x,\text{naive}} = \alpha\beta_x, \quad \theta_{\text{naive}} = \theta + \alpha\sigma_u^2\beta_x^2, \quad \lambda_{0,\text{naive}}(t) = \lambda_0(t)e^{(1-\alpha)\mu_x\beta_x}.$$

This result suggests that the naive estimator of the regression coefficient β_x is attenuated in the conventional way, and the naive estimator of the variance component θ overestimates θ. This result applies to both uncensored and censored data and to any cluster size n.

When X is a subject-level covariate, the bias analysis is much more complicated. There is generally no closed-form expression for the asymptotic bias, and numerical calculations are used.

3·3. *Asymptotic bias in the naive estimator when X is a subject-level covariate*

We assume that X follows the random intercept model (10). A comparison of the naive model (13) with the observed data $(T_i, \Delta_i | W_i)$ model in (12) shows that the naive model misspecifies the true $(T_i, \Delta_i | W_i)$ model by a complicated term $e^{\beta_2^* \bar{W}_{i\cdot}} E(e^{b_{ij}^{**}} | t_{ij} \geqslant t, W_i, b_i^*)$.

When the event is rare and the data are independent, Prentice (1982) showed that a closed-form expression for the bias is available in the standard Cox model with measurement error. We will show that this is not the case in frailty measurement error models. When the event is rare, we have that

$$E(e^{b_{ij}^{**}} | t_{ij} \geqslant t, W_{ij}, b_i) \doteq E(e^{b_{ij}^{**}} | W_{ij}, b_i) = e^{\gamma^2/2}.$$

Hence the $(T_i, \Delta_i | W_i)$ model in (12) becomes

$$\lambda_{ij}(t | W_i, b_i^*) \doteq \lambda_0(t) e^{\beta_0^* + \gamma^2/2 + \beta_x^* W_{ij} + \beta_2^* \bar{W}_{i\cdot} + b_i^*}. \tag{14}$$

A comparison of (14) with the naive model (13) reveals that the naive model misspecifies the fixed effects structure. Therefore, unlike in the standard Cox model with measurement error, we still do not have closed-form expressions for the asymptotic biases of the naive estimators. Numerical calculations are necessary.

In what follows, we conduct bias analysis without making the rare-event assumption. We consider the censoring mechanism as type I censoring, where all subjects are followed to either a failure or some maximum time T_o, whichever is shorter. In other words, $c_{ij} = T_o$ for all i and j. It is difficult to calculate the asymptotic bias numerically with the baseline hazard $\lambda_0(t)$ unspecified. In our numerical study, we assume for simplicity that the baseline hazard function is a constant, λ_0.

If we suppress the index i, the expected censoring proportion can be calculated as

$$\mathrm{pr}(v_j > T_o) = \int S_j(T_o | X_j, b) f(X_j)\, dX_j\, d\Phi(b; \theta),$$

where $S_j(.)$ is the conditional survival function under the frailty model (8) and $f(X_j)$ is the marginal density of X_j under (10). Values of T_o can hence be chosen to obtain desired censoring proportions, with $T_o = \infty$ corresponding to noncensoring.

Denote the observed times by $t_j = v_j^{\delta_j} T_o^{1-\delta_j}$ $(j = 1, \ldots, n)$, or, in vector notation, $T = V^{\cdot\Delta} \cdot T_o^{1-\Delta}$, where $V = (v_1, \ldots, v_n)'$ is the true survival time vector and the operator $\cdot$ denotes an elementwise multiplication. Denote by $\Theta = \{\beta_x, \theta, \lambda_0(t)\}'$ the true value and by Θ_{naive} the asymptotic limit of the naive estimator of Θ as $m \to \infty$. It follows that Θ_{naive} maximises

$$\begin{aligned} &E\{l_{\text{naive}}(T, \Delta | W; \Theta_{\text{naive}})\} \\ &\quad = \sum_{k=0}^{n} \sum_{\delta_1 + \ldots + \delta_n = k} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{0}^{T_o} \cdots \int_{0}^{T_o} l_{\text{naive}}(V^{\cdot\Delta} \cdot T_o^{1-\Delta}, \Delta | X + U; \Theta_{\text{naive}}) \\ &\qquad \times \left[\prod_{j=1}^{n} \{f_j(v_j | X_j, b)\, dv_j\}^{\delta_j} \{S_j(T_o | X_j, b)\}^{1-\delta_j}\, d\Phi(U_j; \sigma_u^2) \right] \\ &\qquad \times d\Phi(b; \theta)\, d\Phi(X - \mu_x 1; \sigma_{x\mu}^2 11' + \sigma_x^2 I), \end{aligned}$$

where $l_{\text{naive}}(T, \Delta | W; \Theta_{\text{naive}})$ is the logarithm of equation (9) with X replaced by W. Here $f(t_j | X_j, b; \Theta)$ $(j = 1, \ldots, n)$ is the true density of the survival time t_j conditional on X_j and the frailty b and is equal to $f(t_j | X_j, b) = \lambda_{0j} \exp(-t_j \lambda_{0j})$ with $\lambda_{0j} = \lambda_0 \exp(\beta_x X_j + b)$.

Maximisation is with respect to Θ_{naive} for fixed Θ. Since $E\{l_{\text{naive}}(T, \Delta | W; \Theta_{\text{naive}})\}$ involves multidimensional integration, we use combination of the methods of Gauss–Hermite quadrature and Monte Carlo simulations to evaluate it. The Newton–Raphson algorithm is then used for maximisation.

The parameter values used in our numerical calculations were cluster size $n = 2$, $\lambda_0 = \exp(-2)$, $\beta_x = 2$, $\theta = 0{\cdot}5$, $\mu = 1$, $\sigma^2_{x\mu} = 1$ and $\sigma^2_x = 1$. We varied the measurement error variance σ^2_u from 0 to 1. We chose T_o to obtain censoring proportions equal to 0%, 30%, 50% and 80%. Figures 1(a), (b) give the asymptotic biases in the naive estimators $\beta_{x,\text{naive}}$ and θ_{naive} assuming different censoring proportions. Figure 1 suggests that, in the presence of censoring, $\beta_{x,\text{naive}}$ underestimates β_x and θ_{naive} overestimates θ. As the censoring proportion becomes higher, the bias in $\beta_{x,\text{naive}}$ increases, whereas the bias in θ_{naive} interestingly

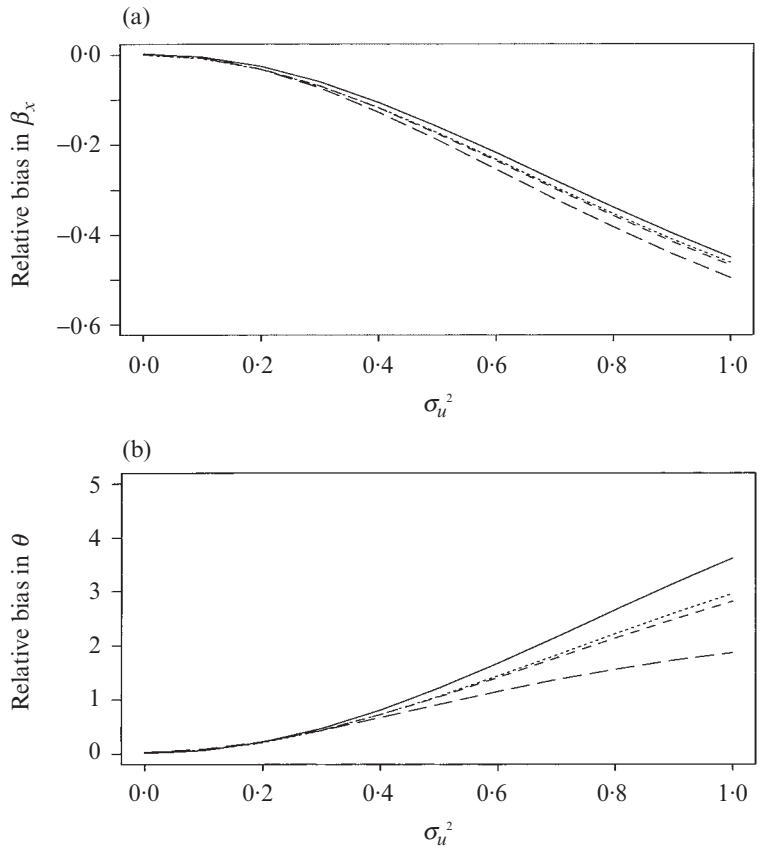

Fig. 1. Asymptotic relative biases in naive estimators against the measurement error variance σ^2_u when the cluster size $n = 2$. The true parameter values are $\lambda_0 = \exp(-2)$, $\beta_x = 2$, $\theta = 0{\cdot}5$, $\mu_x = 1$, $\sigma^2_{x\mu} = 1$ and $\sigma^2_x = 1$. The four curves in each plot correspond to censoring proportions, 0%, 30%, 50% and 80%, with the solid line for 0%, no censoring, and the longest dashed line for 80%.

decreases. The bias results concerning θ_{naive} are consistent with the results in Nielsen et al. (1992) in frailty models without measurement error. They found in their simulation studies that the finite sample bias in the maximum likelihood estimator of θ is much smaller when there is censorship than when there is no censorship.

To study the effect of the cluster size n on the asymptotic bias in the presence of

censoring, we consider the cases where $n=2$ and $n\to\infty$. When $n\to\infty$, we have $\tilde{\alpha}\to 0$, $\beta_2^*\to(1-\alpha)\beta_x$ and $\bar{W}_{i\cdot}=\mu_x+a_i+\bar{U}_{i\cdot}+\bar{e}_{i\cdot}\to\mu_x+a_i$. It follows that the observed data $(T_i, \Delta_i|W_i)$ model becomes

$$\lambda_{ij}(t|W_i, b_{i*})=\lambda_0(t)e^{\beta\delta^*+\beta_x^*W_{ij}+b_{i*}}E(e^{b_{ij}^{**}}|t_{ij}\geqslant W_{ij}, b_{i*}), \tag{15}$$

where $b_{i*}\sim N\{0, \theta+(1-\alpha)^2\sigma_{x\mu}^2\beta_x^2\}$. Asymptotic bias of the naive estimator is hence calculated by maximising $E\{l_{\text{naive}}(T_i, \Delta_i|W_i; \Theta_{\text{naive}})\}$, where the expectation is taken under (15). Figures 2(a), (b) compare the asymptotic biases in naive estimators of β_x and θ when $n=2$ and $n=\infty$ and the censoring proportions are 0% and 50%. These plots show that, as the cluster size increases, the biases become more severe. Similar results hold in linear mixed models and logistic mixed models (Wang et al., 1998).

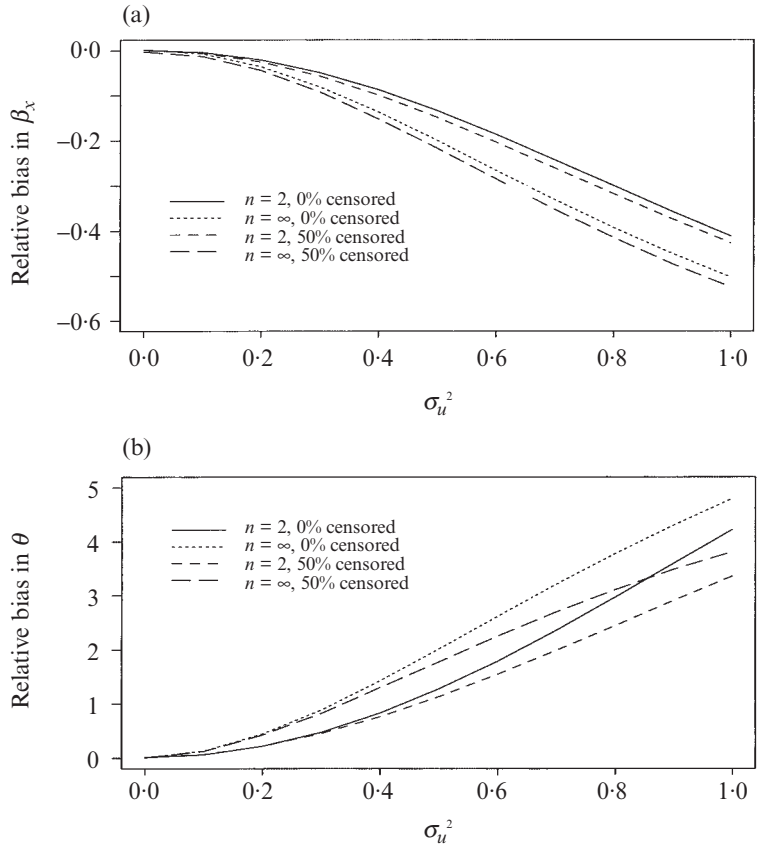

Fig. 2. Asymptotic relative biases in naive estimators against the measurement error variance σ_u^2 for two cluster sizes, $n=2$ and $n=\infty$, and for two censoring proportions, 0% and 50%. The true parameter values are $\lambda_0=\exp(-2)$, $\beta_x=2$, $\theta=0{\cdot}5$, $\mu_x=1$, $\sigma_{x\mu}^2=1$ and $\sigma_x^2=1$.

4. THEORETICAL PROPERTIES OF NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATION

In this section we propose estimation for the general frailty measurement error model in equations (1) and (3) using the nonparametric maximum likelihood method. We assume that X_i follows the linear mixed model (5) and that the baseline hazard $\lambda_0(t)$ is unspecified.

Let $\Omega = \{\beta_x, \beta_z', \mu', \theta, \mathrm{vec}(\Sigma_{x\mu}), \sigma_x^2\}'$ and $\Theta = \{\Omega, \Lambda_0(t)\}$. The nonparametric maximum likelihood estimator of Θ maximises the observed data likelihood (4).

We study the theoretical properties of the model using counting process theory; the proofs of the theorems are given in the Appendix. These results motivate an EM algorithm in § 5 for calculating the nonparametric maximum likelihood estimator of Θ, where the estimated $\Lambda_0(t)$ is a step function with jumps at distinct failure times.

The outcomes (t_{ij}, δ_{ij}) can be written in terms of the counting and at-risk processes

$$N_{ij}(t) = I(t_{ij} \leqslant t, \delta_{ij} = 1), \quad Y_{ij}(t) = I(t_{ij} \geqslant t),$$

respectively. For simplicity, we assume the cluster sizes n_i to be the same, $n_i = n$, in our proof. Let

$$N_i(t) = \{N_{i1}(t), \ldots, N_{in}(t)\}', \quad Y_i(t) = \{Y_{i1}(t), \ldots, Y_{in}(t)\}'.$$

Denote $(N_1, \ldots, N_m)'$ by N and define Y and W similarly. Introduce the right continuous filtration $\{\mathscr{F}_t : t \geqslant 0\}$, where

$$\mathscr{F}_t = \sigma\{X_i, Z_i, N_i(u), Y_i(u+) : 0 \leqslant u \leqslant t\}$$

and $\sigma\{.\}$ denotes a σ-algebra. Given (X_i, Z_i) and the frailty vector b_i, let $(\Xi, \{\mathscr{F}_t\}_{t \geqslant 0}, P)$ be a filtered probability space such that, under P, $N_i(t)$ is a multivariate counting process. We assume that there is no possibility for ties in failure times. The loglikelihood (4) of the observed data $\{N_i(t), Y_i(t), W_i\}$ can be written using counting process notation as

$$\begin{aligned} l_m(N, Y, W; \Omega, \Lambda_0) &= \sum_{i=1}^{m} l(N_i, Y_i, W_i; \Omega, \Lambda_0) \\ &= \sum_{i=1}^{m} \log \int e^{\sum_{j=1}^{n} \int_0^\infty \log\{Y_{ij}^*(t)\lambda_0(t)\}\, dN_{ij}(t) - Y_{ij}^*(t)\, d\Lambda_0(t)} L(W_i | X_i) L(X_i | Z_i)\, dX_i\, d\Phi(b_i; D), \end{aligned} \tag{16}$$

where $Y_{ij}^*(t) = Y_{ij}(t) \exp(X_{ij}\beta_x + Z_{ij}'\beta_z + B_{ij}'b_i)$.

Let $\Theta_0 = \{\Omega_0, \Lambda_0^0(t)\}$ be the true value of Θ and let $\{(N_i, Y_i, W_i, Z_i, B_i)\}$ be a sequence of identically and independently distributed replicates. We postulate the following regularity conditions.

Regularity Condition 1. We require $\mathrm{pr}_{\Theta_0}\{\sum_{j=1}^{n} Y_{ij}(u) \geqslant 1, \text{ for all } u \in [0, \infty)\} > 0$.

Regularity Condition 2. We require $\mathrm{pr}_{\Theta_0}\{\sum_{i=1}^{n} Y_{ij}(0) \geqslant 2\} > 0$.

Regularity Condition 3. If $cG(W_{ij}) = 0$, where $G(.)$ is any nondegenerate function, then $c = 0$.

Regularity Condition 4. If $c'Z_{ij} = G(W_{ij})$, where $G(.)$ is any function, then $c = 0$.

Regularity Condition 1 is a standard assumption in the conventional proportional hazards model to ensure that we can observe failures in the entire interval and therefore can estimate $\Lambda_0(t)$ in the entire interval. Regularity Condition 2 excludes the case where there is only one subject per cluster and ensures the identifiability of the variance components θ (Nielsen et al., 1992). Regularity Conditions 3 and 4 exclude the trivial situation where the observed covariate W_i is a constant or is collinear with Z_i.

THEOREM 1 (*Identifiability*). *If $\Lambda_0(t)$ is a nonnegative nondecreasing continuous function*

over $[0, \infty)$ *with* $\Lambda_0(0) = 0$, *then, under Regularity Conditions* 1–4, *the Kullback–Leibler information is strictly positive for* $\Theta \neq \Theta_0$.

THEOREM 2 (*Unboundedness*). *The loglikelihood* $l_m\{N, Y, W; \Omega, \Lambda_0\}$ *in* (16) *is unbounded in the Euclidean space* $\times C^+[0, \infty)$, *where* $C^+[0, \infty)$ *is the set of nonnegative nondecreasing continuous functions on* $[0, \infty)$ *taking value* 0 *at* $t = 0$.

Theorem 1 shows that, under Regularity Conditions 1–4, the frailty measurement error model is identifiable. Theorem 2 implies that, when $\Lambda_0(t)$ lies in the space of continuous functions, the maximiser of the likelihood (16) does not exist. This suggests that we need to restrict the parameter space of $\Lambda_0(t)$ to include only discrete functions. In other words, one needs to replace $\lambda_0(t)$, the continuous derivative of $\Lambda_0(t)$, in the likelihood function (16) by $\Delta\Lambda_0(t)$, the jump of $\Lambda_0(t)$ at time t. The resulting maximiser is called the nonparametric maximum likelihood estimator. Since direct maximisation of the integrated likelihood (16) is difficult, in § 5 we propose EM algorithms for calculating this estimator; a simpler EM algorithm was used by Nielsen et al. (1992) and Murphy (1995) for estimation in frailty models without measurement error. Theorem 3 shows that such an estimator exists.

THEOREM 3 (*Existence*). *Suppose that* $\Lambda_0(t)$ *is a step function with steps taken at each failure time, the observed covariates* W_{ij} *and* Z_{ij} *are bounded in* R^{p+1}, σ_x^2 *is bounded away from* 0, $\Sigma_{x\mu}$ *is positive definite and its norm is bounded away from* 0, *and* β_x, β_z *and* μ *are bounded. If* $\sum_{i=1}^m \sum_{j=1}^n N_{ij}(\infty) \geqslant 1$, *the maximiser* $(\hat{\Omega}_m, \hat{\Lambda}_{0,m})$ *of* $l_m(N, Y, W; \Omega, \Lambda_0)$ *exists and is finite.*

5. THE EM ALGORITHM FOR CALCULATING THE NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATE

5·1. *The* EM *algorithm when X follows the fixed effects model*

When the covariate X is a cluster-level covariate, it is often appropriate to assume that it follows the fixed effects model (6). Let $\tilde{X}_i = (X_i, Z_i)$ and $\beta = (\beta_x, \beta_z')$. Define $T = (T_1, \ldots, T_m)'$, and Δ, W, X, Z and b similarly. The complete data are (T, Δ, W, Z, X, b), while the observed data are (T, Δ, W, Z). When the unobserved covariate X follows the fixed effects model (6), the complete data loglikelihood is, apart from an additive constant,

$$
\begin{aligned}
& l(T, \Delta, W, Z, X, b; \Theta) \\
& \quad = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n_i} [\delta_{ij} \log\{\Delta\Lambda_{ij}(t_{ij})\} - \Lambda_{ij}(t_{ij})] - \frac{n_i}{2} \log \sigma_u^2 - \frac{1}{2\sigma_u^2} (W_i - X_i)'(W_i - X_i) \right. \\
& \qquad \left. - \frac{n_i}{2} \log \sigma_x^2 - \frac{1}{2\sigma_x^2} (X_i - \tilde{Z}_i \mu)'(X_i - \tilde{Z}_i \mu) - \frac{1}{2} \log |D| - b_i' D^{-1} b_i \right\},
\end{aligned}
$$

where $\Lambda_{ij}(t) = \Lambda_0(t) \exp(\tilde{X}_{ij}'\beta + B_{ij}' b_i)$, $\tilde{Z}_i = (1_i, Z_i)$ and $\Theta = \{\Omega, \Delta\Lambda_0(t)\}$.

Let $\hat{\Theta}^{(k)}$ be the estimate of Θ at the kth iteration. The M-step updates Θ by solving $E\{\partial l(T, \Delta, W, Z, X, b; \Theta)/\partial\Theta \mid T, \Delta, W, Z; \hat{\Theta}^{(k)}\} = 0$. For simplicity, we here consider the case where the covariance matrix D of the frailty b_i is unstructured. Then Θ is updated

using

$$\hat{\mu}^{(k+1)} = \left(\sum_{i=1}^{m} \tilde{Z}_i' \tilde{Z}_i \right)^{-1} \sum_{i=1}^{m} \tilde{Z}_i' E(X_i | T, \Delta, W, Z; \hat{\Theta}^{(k)}),$$

$$\hat{\sigma}_x^{2(k+1)} = \frac{1}{\sum n_i} \sum_{i=1}^{m} E\{(X_i - \tilde{Z}_i'\mu)'(X_i - \tilde{Z}_i'\mu) | T, \Delta, W, Z; \hat{\Theta}^{(k)}\},$$

$$\hat{D}(\theta)^{(k+1)} = \frac{1}{m} \sum_{i=1}^{m} E(b_i b_i' | T, \Delta, W, Z; \hat{\Theta}^{(k)}),$$

$$\Delta\hat{\Lambda}_0(t)^{(k+1)} = \frac{\sum_{i,j} \delta_{ij} I(t_{ij} = t)}{\sum_{i,j} E(e^{\tilde{X}_{ij}'\beta + B_{ij}'b_i} | T, \Delta, W, Z; \hat{\Theta}^{(k)}) I(t_{ij} \geqslant t)}, \tag{17}$$

$$U(\beta) = \sum_{i,j} E[\{\delta_{ij} - \hat{\Lambda}_0(t_{ij}) e^{\tilde{X}_{ij}'\beta + B_{ij}'b_i}\} \tilde{X}_{ij} | T, \Delta, W, Z; \hat{\Theta}^{(k)}] = 0, \tag{18}$$

where $U(\beta)$ is the conditional score equation for β. Equation (17) resembles the Breslow baseline hazard estimator. If we substitute (17) into (18), some calculations give

$$U(\beta) = \sum_{i,j} \delta_{ij} \left\{ E(\tilde{X}_{ij} | T, \Delta, W, Z; \hat{\Theta}^{(k)}) - \frac{\sum_{s,t} E(\tilde{X}_{st} e^{\tilde{X}_{st}'\beta + B_{st}'b_s} | T, \Delta, W, Z; \hat{\Theta}^{(k)}) I(t_{st} \geqslant t_{ij})}{\sum_{s,t} E(e^{\tilde{X}_{st}'\beta + B_{st}'b_s} | T, \Delta, W, Z; \hat{\Theta}^{(k)}) I(t_{st} \geqslant t_{ij})} \right\},$$

which takes the same form as the partial likelihood score equation in the conventional Cox proportional hazards model with the Breslow approximation for ties. The Newton–Raphson method can be used to solve (18) iteratively.

The above conditional expectations are calculated at the E-step using Monte Carlo simulations by generating (X, b) under the conditional density of $(X, b | T, \Delta, W, Z)$, which takes the form

$$f(X, b | T, \Delta, W, Z; \Theta) = \frac{f(T, \Delta | X, Z, b; \Theta) f(W | X; \Theta) f(X | Z; \Theta) f(b; \Theta)}{\int f(T, \Delta | X, Z, b; \Theta) f(W | X; \Theta) f(X | Z; \Theta) f(b; \Theta)\, dX\, db}.$$

5·2. *The* EM *algorithm when X follows the random effects model*

When X is a subject-level covariate, it is often more proper to assume that X follows the random effects model (5). The complete data now are $(T, \Delta, W, Z, X, b, a)$, where $a = (a_1, \ldots, a_m)'$, and we add $\Sigma_{x\mu}$ to the parameter vector Θ. The equations used to solve for $\{\beta, D(\theta), \Delta\Lambda_0(t)\}$ at the M-step remain the same as those in § 5·1, while

$$\hat{\mu}^{(k+1)} = \left(\sum_{i=1}^{m} \tilde{Z}_i' \tilde{Z}_i \right)^{-1} \sum_{i=1}^{m} \tilde{Z}_i' E\{(X_i - A_i a_i) | T, \Delta, W, Z; \hat{\Theta}^{(k)}\},$$

$$\hat{\sigma}_x^{2(k+1)} = \left(\sum_{i=1}^{m} n_i \right)^{-1} \sum_{i=1}^{m} E\{(X_i - \tilde{Z}_i\mu - A_i a_i)'(X_i - \tilde{Z}_i\mu - A_i a_i) | T, \Delta, W, Z; \hat{\Theta}^{(k)}\},$$

$$\hat{\Sigma}_{x\mu}^{(k+1)} = m^{-1} \sum_{i=1}^{m} E(a_i a_i' | T, \Delta, W, Z; \hat{\Theta}^{(k)}).$$

Alternatively, one can use the ECME approach of Liu & Rubin (1994), which has a faster

convergence rate, to estimate μ as

$$\hat{\mu}^{(k+1)} = \left(\sum_{i=1}^{m} \tilde{Z}_i' V_i^{-1} \tilde{Z}_i \right)^{-1} \sum_{i=1}^{m} \tilde{Z}_i' V_i^{-1} E(X_i | T, \Delta, W, Z; \hat{\Theta}^{(k)}),$$

where $V_i = \hat{\text{cov}}\{X_i | \Theta^{(k)}\} = A_i \hat{\Sigma}_{x\mu}^{(k)} A_i' + \hat{\sigma}_x^{2(k)} I_i$.

The conditional expectations are calculated at the E-step using Monte Carlo simulations by generating (X, a, b) under the conditional density of $(X, a, b | T, \Delta, W)$, namely

$$f(X, a, b | T, \Delta, W, Z; \Theta) = \frac{f(T, \Delta | X, Z, b; \Theta) f(W | X; \Theta) f(X | Z, a; \Theta) f(a; \Theta) f(b; \Theta)}{\int f(T, \Delta | X, Z, b; \Theta) f(W | X; \Theta) f(X | Z, a; \Theta) f(a; \Theta) f(b; \Theta)\, dX\, da\, db}.$$

5·3. *Standard error calculations using profile likelihood*

The standard errors of the maximum likelihood estimates are conventionally calculated by inverting the observed information matrix. However, in the case of the nonparametric maximum likelihood estimates, the number of parameters is large and this approach is not feasible. Hence we calculate the variance of $\hat{\Omega}$ using the profile likelihood method (Hu et al., 1998; Murphy, 1995). The profile loglikelihood of Ω is defined as $l_p(\Omega) = \sup_{\Lambda_0} l(\Omega, \Lambda_0)$, where the baseline cumulative hazard Λ_0 is the nuisance parameter. Then the covariance matrix of the maximum likelihood estimate $\hat{\Omega}$ is estimated by $\hat{\text{cov}}(\hat{\Omega}) = \{-\partial^2 l_p(\hat{\Omega})/\partial\Omega\, \partial\Omega'\}^{-1}$. Numerical differentiation is used to calculate this derivative.

6. SIMULATION STUDY

In our simulation studies of the finite sample performance of our methods, survival times v_{ij} were generated within each cluster under the conditional hazard

$$\lambda_{ij}(t) = \lambda_0(t) \exp(\beta_x X_{ij} + \beta_z Z_{ij} + b_i) \quad (j = 1, \ldots, n;\ i = 1, \ldots, m),$$

where the Z_{ij} were generated as independent $N(0, 1)$. Censoring times c_{ij} were generated as independent uniform random variables on $[0, c]$.

We chose true parameter values as follows: the baseline hazard was $\lambda_0(t) = t$; we took $n = 3$ and $m = 40$; X followed the random intercept model (10) with $\mu = 1$, $\sigma_{xu}^2 = 1$ and $\sigma_x^2 = 1$; the frailty variance was $\theta = 0{\cdot}5$; we took $\beta_x = 2$ and $\beta_z = 1$; the values of c were chosen to yield four different censoring proportions, 0%, 30%, 50% and 80%; and the measurement error variance σ_u^2 was 0·5. We calculated the nonparametric maximum likelihood estimates by the EM algorithms in § 5 using SAS/IML. We ran 500 simulations under each combination of parameter configurations.

For each simulated dataset, we calculated the naive estimates which ignore measurement error and the nonparametric maximum likelihood estimates which account for measurement error. The averages of the estimates, the empirical standard errors, the estimated profile likelihood standard errors and the mean squared errors are displayed in Table 1. This shows that the naive estimates of β_x and β_z are attenuated and the naive estimate of θ is inflated. Higher censorship results in more biased naive estimates of β_x and β_z and a less biased naive estimate of θ. These results are consistent with our theoretical bias calculations in § 3.

The results in Table 1 show that the nonparametric maximum likelihood estimates calculated using the EM algorithm perform very well and that the average estimates are

Table 1. *Simulation results for the frailty measurement error model based on* 500 *replicates. The true values are* $\beta_x = 2$, $\beta_z = 1$, $\theta = 0{\cdot}5$, $\mu = 1$, $\sigma_x^2 = 1$ *and* $\sigma_{xu}^2 = 1$. *The measurement error variance is* $\sigma_u^2 = 0{\cdot}5$.

Censoring level	Parameter	Full likelihood Estimate	SE_e	SE_p	MSE	Naive Estimate	SE_e	SE_p	MSE
0%	θ	0·47	0·37	0·40	0·137	0·69	0·28	0·31	0·115
	β_x	1·97	0·31	0·34	0·098	1·48	0·14	0·15	0·290
	β_z	0·97	0·17	0·18	0·029	0·88	0·13	0·13	0·031
	μ	1·00	0·19	0·18	0·036	—	—	—	—
	σ_x^2	1·02	0·30	0·29	0·090	—	—	—	—
	σ_{xu}^2	1·02	0·34	0·35	0·116	—	—	—	—
30%	θ	0·48	0·42	0·47	0·177	0·68	0·35	0·30	0·148
	β_x	2·04	0·36	0·32	0·131	1·39	0·17	0·20	0·392
	β_z	1·02	0·21	0·22	0·044	0·82	0·14	0·16	0·052
	μ	1·01	0·21	0·20	0·044	—	—	—	—
	σ_x^2	1·01	0·25	0·26	0·063	—	—	—	—
	σ_{xu}^2	0·98	0·41	0·40	0·168	—	—	—	—
50%	θ	0·48	0·43	0·48	0·185	0·64	0·39	0·34	0·171
	β_x	1·95	0·36	0·32	0·132	1·33	0·17	0·14	0·442
	β_z	0·96	0·22	0·21	0·049	0·72	0·16	0·15	0·104
	μ	1·00	0·22	0·21	0·048	—	—	—	—
	σ_x^2	1·01	0·26	0·25	0·067	—	—	—	—
	σ_{zu}^2	1·01	0·36	0·38	0·130	—	—	—	—
80%	θ	0·51	0·97	1·01	0·941	0·59	0·90	0·93	0·818
	β_x	2·03	0·43	0·39	0·185	1·21	0·19	0·16	0·660
	β_z	1·03	0·28	0·27	0·079	0·60	0·19	0·18	0·196
	μ	0·99	0·19	0·19	0·036	—	—	—	—
	σ_x^2	1·02	0·30	0·32	0·090	—	—	—	—
	σ_{xu}^2	1·01	0·35	0·37	0·123	—	—	—	—

SE_e, empirical standard error; SE_p, average of the standard errors obtained by the profile likelihood; MSE's are mean squared errors using SE_e.

close to the true values. One can easily see the trade-off between bias and variance. The nonparametric maximum likelihood estimates effectively correct the biases in the naive estimates, but their variances are larger. The mean squared errors of the nonparametric maximum likelihood estimates are smaller than those of the naive estimates, especially for the estimates of the regression coefficients. As the percentage of censoring increases, the variances of the estimates become larger.

We also examined the performance of the standard error estimates calculated using the profile likelihood method. As shown in Table 1, the estimated standard errors using the profile likelihood method agree well with the empirical standard errors.

7. Application to the Kenya parasitaemia data

A major interest of the western Kenya parasitaemia study (McElroy et al., 1997) was to investigate the effect of daily mean dose of infective mosquito bite exposure on the risk of parasitaemia, which is an indicator for potential malaria. The main challenges in this study are as follows: the exposure variable, daily mean dose of infective bites, was measured with substantial measurement error; multiple children from the same household were

involved in the study and their outcome variables, times until parasitaemia, might be correlated.

A total of 607 children aged from six months to six years were enrolled into the study between February 1986 and July 1987. Parasitaemia is highly prevalent among African children, and 94% of children in the study were affected upon enrolment. At entry into the study, each child, regardless of parasitaemia status, was treated with drugs sulfadoxine and pyrimethamine to eliminate the parasitaemia infection in blood. Their blood films were examined two weeks after enrolment. Children with positive blood films were excluded from the study to minimise the chance that a recurrent parasitaemia was caused by sulfadoxine/pyrimethamine drug resistance. This left 542 children from 309 households. These children were then followed for the first recurrence of parasitaemia up to 22 months.

In the first two weeks after enrolment, two field workers visited each household one night each week and took turns to collect mosquitoes from each other's legs every 30 minutes through the night. The numbers of infective mosquitoes were identified in a laboratory and the daily mean dose of infective mosquito bites in the first two weeks was calculated using the average of the two night measures. It was hence measured with substantial error: first, only one night measure was available per week; secondly, the mosquitoes were collected from the legs of the field workers rather than from the children directly. The other covariates included gender, coded as 1 for female and 0 for male, age and baseline parasitaemia density. The average follow-up time was 9 months, and about 90% of children experienced recurrent parasitaemia during the follow-up. The daily mean dose of infective mosquito bites was 0·89 on average. The baseline parasitaemia density was log-transformed to give 'logbase', and the daily mean dose of infective bites in the first two weeks was quartic-root transformed, creating a variable 'qbite', to make the normality assumption more plausible.

We fitted a random intercept frailty measurement error model with $X =$ true 'qbite' and $Z =$ ('age', 'gender', 'logbase'). Since different children within the same household might enter into the study at different times, their 'qbite' might be different, and we therefore assumed that X follows the random intercept model (10). The observed 'qbite', W, also follows a linear random intercept model with the variance of the random intercept equal to $\sigma^2_{x\mu}$ and the residual variance equal to $\sigma^2_x + \sigma^2_u$. Since there were no validation data available and there was no replication, the measurement error variance σ^2_u could not be estimated from the available data. We fitted a linear random intercept model to W. This allowed us to estimate the sum of σ^2_x and σ^2_u, which was 0·20. Then we conducted a sensitivity analysis by varying σ^2_u from 0, that is naive analysis, to moderate measurement error, $\sigma^2_u = 0{\cdot}08$, to severe measurement error, $\sigma^2_u = 0{\cdot}20$. In our illustration, we treated σ^2_u as fixed and known. We fitted the model using the EM algorithm discussed in § 5·2 and calculated the standard errors of the parameter estimates using the profile likelihood method discussed in § 5·3. The results are presented in Table 2.

All analyses showed that a higher daily mean dose of infective mosquito bites significantly increased the risk of parasitaemia. Older children had a significantly higher risk of parasitaemia and higher baseline parasitaemia density also significantly increased the risk of recurrent parasitaemia. Ignoring measurement error attenuated the regression coefficient estimates. As σ^2_u increased, the estimates of the regression coefficients became larger. For example, the estimate of the coefficient of 'qbite', with estimated standard error in brackets, increased from 0·33 (0·11) when $\sigma^2_u = 0$ to 0·99 (0·23) when $\sigma^2_u = 0{\cdot}20$. This indicates that accounting for measurement error increases the magnitude of the estimated

Table 2. *Parameter estimates for the Kenya parasitaemia data. Moderate error corresponds to the measurement error variance* $\sigma_u^2 = 0{\cdot}08$. *Severe error corresponds to* $\sigma_u^2 = 0{\cdot}20$. *Naive estimates assume that* $\sigma_u^2 = 0$ *by ignoring measurement error*

	Naive estimate		Moderate error		Severe error	
Covariate	Estimate	SE	Estimate	SE	Estimate	SE
qbite	0·333	0·108	0·491	0·149	0·990	0·230
age	0·041	0·016	0·039	0·020	0·039	0·021
gender	0·056	0·033	0·057	0·035	0·053	0·036
logbase	0·092	0·016	0·092	0·024	0·094	0·024
θ	0·161	0·114	0·146	0·111	0·114	0·118

SE, standard error.

effects of 'qbite'. As σ_u^2 increased from 0 to 0·20, the estimate of the frailty variance $\hat{\theta}$ decreased from 0·161 to 0·114. These results coincide with our theoretical findings in § 3.

8. Discussion

Our simulation study shows that the EM algorithm performs well. The algorithm can however be computationally intensive, especially when the number of clusters and the cluster sizes are large, since a large number of clusters often implies estimation of more jumps of the cumulative baseline hazard and a large cluster size implies evaluating a higher dimensional integral at the E-step. The convergence of the EM algorithm can be very slow when numerous parameters are estimated.

The proof of consistency and asymptotic normality of the nonparametric maximum likelihood estimator is still an open problem, but our simulation results seem to point to the asymptotic validity of the proposed method.

The measurement error literature distinguishes structural modelling and functional modelling by whether or not a distribution is assumed for the unobserved covariate X. In this paper we have considered structural modelling using the approach, which requires correct specification of the distribution of X. In a subsequent paper, we study the robustness of maximum likelihood estimation with respect to misspecification of the distribution of the unobserved covariate. Our results show that the biases in the misspecified maximum likelihood estimators of the regression coefficients are small, but the biases in the variance components are often high. An alternative more robust approach is to explore functional modelling using the SIMEX approach (Carroll et al., 1995), which does not make a distributional assumption about X. The SIMEX method could yield more robust estimators of the model parameters, but the estimators could be less efficient compared to those under structural modelling when the X model is correctly specified. We will report on this elsewhere.

Acknowledgement

This work was supported in part by a grant from U.S. National Cancer Institute. The authors thank Susan Murphy for helpful discussions and Peter McElroy for providing the Kenya parasitaemia data. The authors also gratefully acknowledge the editor and a

referee for their insightful comments and suggestions that substantially improved the paper.

APPENDIX

Proofs

Proof of Theorem 1. For simplicity, we only prove model identifiability when X follows the fixed effects model (6) and the frailty b_i is a scalar. The proof when X follows the random effects model (5) and b_i is a vector is similar.

Suppose that the true parameters are

$$\Theta_0 = \{\Omega_0, \Lambda_0^0(t)\} = \{\beta_{x0}, \beta_{z0}, \mu_{x0}, \mu_{z0}, \theta_0, \sigma_{x0}^2, \Lambda_0^0(t)\}.$$

Let pr_Θ denote the probability measure on the σ-algebra

$$\mathscr{F}_t = \sigma\{W, Z, N(u), Y(u+) : 0 \leqslant u \leqslant t\},$$

with respect to Θ. It is well known that the Kullback–Leibler information is nonnegative (Bickel et al., 1993), and is equal to 0 only when $\mathrm{pr}_\Theta = \mathrm{pr}_{\Theta_0}$.

The identifiability can be proved by showing that the joint distribution of (N, Y, W) is uniquely determined by the parameters Θ_0. That is, we need to show that, if the joint distributions of (N, Y, W) are equal under two sets of parameters, Θ_0 and $\Theta_1 = \{\beta_{x1}, \beta_{z1}, \mu_{x1}, \mu_{z1}, \theta_1, \sigma_{x1}^2, \Lambda_0^1(t)\}$, then $\Theta_0 = \Theta_1$. Since the equivalence of the joint distributions implies the equivalence of the marginal distributions, we first prove the identifiability of $(\mu_x, \mu_z, \sigma_x^2)$ by considering the marginal distribution of W. Marginally, W_{ij} follows normal distributions $N(\mu_{x0} + Z_{ij}'\mu_{z0}, \sigma_{x0}^2 + \sigma_u^2)$ and $N(\mu_{x1} + Z_{ij}'\mu_{z1}, \sigma_{x1}^2 + \sigma_u^2)$ under Θ_0 and Θ_1 respectively. As a result of the non-collinearity of Z, we immediately have that $\mu_{x0} = \mu_{x1}$ and $\mu_{z0} = \mu_{z1}$. Since σ_u^2 is known, we have that $\sigma_{x0}^2 = \sigma_{x1}^2$.

Suppressing the index i, we next consider the marginal intensities $N_j(.)$ with respect to the filtration $\sigma\{N_j(s), Y_j(s), Z_j, W_j | 0 \leqslant s \leqslant u\}$ before u reaches a failure, that is $u \leqslant \inf\{s | N_j(s) > 0\}$. Under the two sets of parameters Θ_0 and Θ_1, we have that $\lambda(u|Z_j, W_j; \Theta_0) = \lambda(u|Z_j, W_j; \Theta_1)$, that is

$$E_{\Theta_0}\{\lambda_0^0(u) e^{\beta_{x0}X_j + \beta_{z0}'Z_j + b} | T_j \geqslant u, Z_j, W_j\} = E_{\Theta_1}\{\lambda_0^1(u) e^{\beta_{x1}X_j + \beta_{z1}'Z_j + b} | T_j \geqslant u, Z_j, W_j\}, \tag{A1}$$

where T_j is the true survival time for the jth observation.

Under Regularity Condition 2 in § 4, we can consider the joint counting process for two individuals within the same cluster. The joint intensity of $N_1(.)$ and $N_2(.)$ with respect to the filtration of

$$\sigma\{N_j(s_j), Y_j(s_j), W_j, j = 1, 2 | 0 \leqslant s_j \leqslant u_j\},$$

for u_1 and u_2 such that $\max(u_1, u_2) \leqslant \inf\{s | N_1(s) + N_2(s) > 0\}$, is defined as

$$\lim_{\Delta u_1 \to 0, \Delta u_2 \to 0} \frac{E[N_1(u_1 + \Delta u_1) N_2(u_2 + \Delta u_2) | \sigma\{N_j(s_j), Y_j(s_j), Z_j, W_j, 0 \leqslant s_j \leqslant u_j, j = 1, 2\}]}{\Delta u_1 \Delta u_2}$$

Then, under Θ_0 and Θ_1, we have that

$$\begin{aligned} &E_{\Theta_0}\{\lambda_0^0(u_1)\lambda_0^0(u_2) e^{\beta_{x0}(X_1 + X_2) + \beta_{z0}'(Z_1 + Z_2) + 2b} | T_j \geqslant u_j, Z_j, W_j, j = 1, 2\} \\ &\qquad = E_{\Theta_1}\{\lambda_0^1(u_1)\lambda_0^1(u_2) e^{\beta_{x1}(X_1 + X_2) + \beta_{z1}'(Z_1 + Z_2) + 2b} | T_j \geqslant u_j, Z_j, W_j, j = 1, 2\}. \end{aligned} \tag{A2}$$

Let $u, u_1, u_2 \to 0$ in (A1) and (A2). We have that

$$\lambda_0^0(0) e^{\beta_{z0}'Z_j} E_{\Omega_0}(e^{\beta_{x0}X_j + b} | W_j) = \lambda_0^1(0) e^{\beta_{z1}'Z_j} E_{\Omega_1}(e^{\beta_{x1}X_j + b} | W_j), \tag{A3}$$

$$\{\lambda_0^0(0)\}^2 e^{\beta_{z0}'(Z_1 + Z_2)} E_{\Omega_0}\{e^{\beta_{x0}(X_1 + X_2) + 2b} | W_1, W_2\} = \{\lambda_0^1(0)\}^2 e^{\beta_{z1}'(Z_1 + Z_2)} E_{\Omega_1}\{e^{\beta_{x1}(X_1 + X_2) + 2b} | W_1, W_2\}. \tag{A4}$$

Comparing the coefficients of Z_j in (A3), we have that $\beta_{z0} = \beta_{z1}$. Since b and W are independent and X_j and $X_{j'}$ ($j \neq j'$) are independent under the fixed effects model (6), we have that, for $k = 0, 1$,

$$E_{\Omega_k}(e^{\beta_{xk}X_j + b} \mid W_j) = E_{\Omega_k}(e^b) E_{\Omega_k}(e^{\beta_{xk}X_j} \mid W_j) = e^{\theta_k/2} E_{\Omega_k}(e^{\beta_{xk}X_j} \mid W_j),$$

$$E_{\Omega_k}\{e^{\beta_{xk}(X_1 + X_2) + 2b} \mid W_1, W_2\} = e^{2\theta_k} \prod_{j=1}^{2} E_{\Omega_k}(e^{\beta_{xk}X_j} \mid W_j).$$

If we substitute these two identities into (A3) and (A4), simple calculations give that $\theta_0 = \theta_1$. For $k = 0, 1$, $E_{\Omega_k}(e^{\beta_{xk}X_j} \mid W_j) = c_k e^{\beta_{xk}\alpha W_j}$, where the c_k are some constants and $\alpha = \sigma_x^2/(\sigma_x^2 + \sigma_u^2)$. Comparing the coefficients of W_j in (A3), we have that $\beta_{x0} = \beta_{x1}$.

We now have finished proving the identifiability of the finite-dimensional parameter vector Ω. We next show that $\Lambda_0^0(.) = \Lambda_0^1(.)$. Consider the survival function under Θ_k ($k = 0, 1$),

$$S(t \mid W_j, Z_j; \Theta_k) = E_{\Omega_0}\{e^{-\Lambda_0^k(t)\exp(\beta_{xk}X_j + \beta_{zk}'Z_j + b)} \mid W_j, Z_j\} = \mathrm{LP}\{\Lambda_0^k(t)\},$$

where LP(.) is the Laplace transformation of the random variable $\exp(\beta_{x0}X_j + \beta_{z0}'Z_j + b)$ conditional on W_j, Z_j. Then, by the invertibility of the Laplace transformation, $\Lambda_0^0(t) = \Lambda_0^1(t)$. This completes the proof of identifiability. □

Proof of Theorem 2. Denote by $\tau_1 < \ldots < \tau_K$ the distinct failure times. Consider the function

$$\Lambda_M(t) = \frac{M}{K} \sum_{k=1}^{K} \left\{ \left(t - \tau_k + \frac{1}{2M} \right)_+ - \left(t - \tau_k - \frac{1}{2M} \right)_+ \right\},$$

where $M > \max\{1/(\tau_k - \tau_{k-1})\}$ and $(t - a)_+ = 0$ if $t < a$ and $t - a$ if $t \geqslant a$. Then $\Lambda_M(t) \in C^+[0, \infty)$ and, for fixed Ω, $l_m(N, Y, W; \Omega, \Lambda_M) \to \infty$ as $M \to \infty$. This implies the unboundedness of $l_m(N, Y, W; \Omega, \Lambda_M)$ in the Euclidean space $\times\, C^+[0, \infty)$. □

Proof of Theorem 3. Denote by $\tau_1 < \ldots < \tau_K$ the distinct failure times. Replace $\lambda_0(t)$ by $\Delta\Lambda_0(t)$ in the likelihood function (16). Introduce the failure set for the ith cluster, $D_i = \{(i, j) : \delta_{ij} = 1\}$, with corresponding failure times $\tau_{k_{ij}} = t_{ij}$. Then $l(N_i, Y_i, W_i; \Omega, \Lambda_0)$ in (16) can be rewritten as

$$l(N_i, Y_i, W_i; \Omega, \Lambda_0) = \log \int \left\{ \prod_{(i,j) \in D_i} f_{ij}(\tau_{k_{ij}} \mid X_i, b_i) \prod_{(i,j) \notin D_i} S_{ij}(t_{ij} \mid X_i, b_i) \right\} L(W_i \mid X_i) L(X_i)\, dX_i\, d\Phi(b_i; \theta),$$

where

$$f_{ij}(\tau_{k_{ij}} \mid X_i, b_i) = \Delta\Lambda_0(\tau_{k_{ij}}) \exp\{\beta_x X_{ij} + \beta_z' Z_{ij} + b_i - \Sigma_{l:\tau_l < \tau_{k_{ij}}} \Delta\Lambda_0(\tau_l) \exp(\beta_x X_{ij} + \beta_z' Z_{ij} + b_i)\},$$

$$S_{ij}(t_{ij} \mid X_i, b_i) = \exp\{-\Sigma_{l:\tau_l < t_{ij}} \Delta\Lambda_0(\tau_l) \exp(\beta_x X_{ij} + \beta_z' Z_{ij} + b_i)\}.$$

Now we want to show that, if $\{\Delta\Lambda_0(\tau_k)\}|_1^K$ does not remain bounded, the loglikelihood (16) goes to minus infinity. Without loss of generality, we assume that, for some k_0, $\Delta\Lambda_0(\tau_{k_0}) \to \infty$, with the other $\{\Delta\Lambda_0(\tau_k)\}$ ($k \neq k_0$) bounded. Suppose that subject (i_0, j_0) fails at time τ_{k_0}. We can write (16) as

$$l_m(N, Y, W; \Omega, \Lambda_0) = \sum_{i \neq i_0}^{m} l(N_i, Y_i, W_i; \Omega, \Lambda_0) + l(N_{i0}, Y_{i0}, W_{i0}; \Omega, \Lambda_0), \tag{A5}$$

where the first term is bounded and the second term can be written

$$\log \int \left\{ f_{i_0 j_0}(\tau_{k_0}) \prod_{(i_0, j) \in D_{i_0}, j \neq j_0} f_{i_0 j}(\tau_{i_0 j}) \prod_{(i_0, j) \notin D_{i_0}} S_{i_0 j}(t_{i_0 j}) \right\} L(W_{i_0} \mid X_{i_0}) L(X_{i_0})\, dX_{i_0}\, d\Phi(b_{i_0}; \theta).$$

It can be easily seen that $f_{i_0 j_0}(\tau_{k_0}) \to 0$ uniformly as $\Delta\Lambda_0(\tau_{k_0}) \to \infty$ in any bounded set containing β_x, β_z, θ, $X_{i_0 j_0}$ and b_{i_0}. Hence the second term in (A5) tends to $-\infty$. It follows that, if $\{\Delta\Lambda_0(\tau_k)\}|_1^K$ does not remain bounded, the likelihood (16) goes to minus infinity. Further, using the boundedness of Ω, these results show that we only need to consider maximising $l_m(N, Y, W; \Omega, \Lambda_0)$ in a compact set. Since $l_m(N, Y, W; \Omega, \Lambda_0)$ is a continuous function of Ω and $\{\Delta\Lambda_0(\tau_k)\}|_1^K$, its supremum exists and can be attained in any compact set. □

REFERENCES

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. & WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: Johns Hopkins University Press.

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

CLAYTON, D. G. & CUZICK, J. (1985). Multivariate generalizations of the proportional hazards model (with Discussion). *J. R. Statist. Soc.* A **148**, 82–117.

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

HUGHES, M. D. (1993). Regression dilution in the proportional hazards model. *Biometrics* **49**, 1056–66.

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

LIU, C. & RUBIN, D. B. (1994). The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* **81**, 633–48.

MCELROY, P. D. BEIER, J. C., OSTER, C. N., ONYANGO, F. K., OLOO, A. J., LIN, X. BEEDLE, C. & HOFFMAN, S. L. (1997). Dose- and time-dependent relations between infective anopheles inoculation and outcomes of plasmodium falciparum parasitaemia among children in western Kenya. *Am. J. Epidemiol.* **145**, 945–56.

MURPHY, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712–31.

MURPHY, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182–98.

NIELSEN, G. G., GILL, R. D., ANDERSEN, P. K. & SORENSEN, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, 25–43.

OAKES, D. (1989). Bivariate survival models induced by frailties. *J. Am. Statist. Assoc.* **84**, 487–93.

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

PEPE, M. S., SELF, S. G. & PRENTICE, R. L. (1989). Further results on covariate measurement errors in cohort studies with time to response data. *Statist. Med.* **8**, 1167–78.

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

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

WANG, N., LIN, X., GUTIERREZ, R. & CARROLL, R. J. (1998). Bias analysis and SIMEX approach in generalized linear mixed error models. *J. Am. Statist. Assoc.* **93**, 249–61.

[*Received September* 1999. *Revised May* 2000]