Joint modeling of recurrent event processes and intermittently observed time-varying binary covariate processes

Shanshan Li

Department of Biostatistics, Indiana University Fairbanks School of Public Health

Abstract When conducting recurrent event data analysis, it is common to assume that the covariate processes are observed throughout the follow-up period. In most applications, however, the values of time-varying covariates are only observed periodically rather than continuously. A popular ad-hoc approach is to carry forward the last observed covariate value until it is measured again. This simple approach, however, usually leads to biased estimation. To tackle this problem, we propose to model the covariate effect on the risk of the recurrent events through jointly modeling the recurrent event process and the longitudinal measures. Despite its popularity, estimation of the joint model with binary longitudinal measurements remains a challenge, because the standard linear mixed effects model approach is not appropriate for binary measures. In this paper, we postulate a Markov model for the binary covariate process and a random-effect proportional intensity model for the recurrent event process. We use a Markov chain Monte Carlo algorithm to estimate all the unknown parameters. The performance of the proposed estimator is evaluated via simulations. The methodology is applied to an observational study designed to evaluate the effect of Group A streptococcus (GAS) on pharyngitis among school children in India.

1 Introduction

In many medical studies, both longitudinal covariate data and survival outcomes are collected during follow-up. The Cox model is often employed to study the effects of time-varying covariates on the survival outcome. Estimation of the Cox model requires the covariate processes of a subject to be

completely observed until the end of follow-up. In practice, however, the co-variate values are usually observed periodically rather than continuously. A popular ad-hoc approach is to impute the missing covariate value by using the last observed covariate. However, this simple approach usually leads to biased estimation, especially when the last observed covariate is distant in time (Prentice 1982). As an alternative, the joint modeling approach incorporates information from both processes simultaneously and thus leads to unbiased and efficient inferences.

A typical joint model assumes a linear mixed-effects model for the observed time-varying covariates and a proportional hazards model for the survival out-comes. Association between the longitudinal and survival processes can be modeled in two ways. One is by incorporating the trajectory of the longitu-dinal measurements in the hazard function of the failure event (Faucett and Thomas 1996; Wulfsohn and Tsiatis 1997; Tsiatis and Davidian 2004), and the other is through shared random effects (Ratcliffe et al. 2004; Elashoff et al. 2008; Liu et al. 2008; Liu and Huang 2009). Once the full models are spec-ified, one can estimate the model parameters using either likelihood method or Bayesian method. Likelihood methods obtain the maximum likelihood es-timators by using the expectation-maximization (EM) algorithm (Wulfsohn and Tsiatis 1997; Tsiatis and Davidian 2004; Elashoff et al. 2008). Bayesian methods assume that the model parameters follow some prior distribution and make inference based on the posterior distribution given the observed data (Faucett and Thomas 1996; Faucett et al. 1998; Wang and Jeremy 2001). Though computationally intensive, Bayesian method permits full and exact posterior inference for parameters of interest, and it can borrow additional information from other similar studies in the form of prior distributions.

In many studies, recurrent failure events, rather than a survival outcome, are observed for each subject over the course of follow-up. The joint mod-eling approach has been extended to deal with recurrent event data, and a proportional intensity model is usually assumed for the recurrent event data. Henderson et al. (2000) postulated a joint model for the longitudinal data and the single or recurrent event data, where the association between the two processes is induced by correlated latent processes. Liu and Huang (2009) considered a more complex setting where the observation of the longitudinal measurements and the event process are subject to a dependent terminal event process. They proposed to model the pairwise correlation between the three processes by two shared random effects. Kim et al. (2012) extended the work of Liu and Huang (2009) by imposing semi-parametric transformation mod-els for both the recurrent and terminal events, which allows for violation of the proportionality assumption. Sun et al. (2012) considered a similar setting as Liu and Huang (2009) and proposed a more flexible approach that allows unspecified distributions of the latent variables and unspecified dependence structure between the two latent variables.

Of the existing literature on joint models of longitudinal measurements and recurrent events, most researchers considered continuous measurements and applied mixed-effects model for the longitudinal measures. However, stud-

ies with binary longitudinal covariates and recurrent event outcomes are also frequently encountered in practice. Examples include the effect of life quality (high/low) on the risk of diarrhoea (Borgan et al. 2007), the effect of steroid use on the risk of glomerulonephritis (Clayton et al. 2011), and the effect of Group A Streptococcus on the risk of pharyngitis which will be illustrated in this manuscript. To our knowledge, there have been very few references on the joint modeling of time-varying binary covariate and recurrent event processes. Unlike continuous data, it is not appropriate to model binary data with linear mixed effects models. Also, it is difficult to evaluate the full likelihood of binary longitudinal data, because it usually involves a computationally intractable normalization constant (Humphreys and Titterington 2003). Lastly, it is easy to deal with measurement error in continuous data using mixed effects model, while it is difficult to formalize the measurement error in binary data, unless assumptions such as a latent continuous process are imposed for the observed binary covariates (Palta and Lin 1999).

Intrinsically, a binary process is a two-state process taking one of two possible values, which can be coded as 0 and 1. Various authors have proposed models for characterizing the transition patterns from one state to the other via logistic regression (Zeger and Qaqish 1988; Cox and Snell 1989; Albert 2000). In the context of joint modeling, Faucett et al. (1998) proposed a discrete-time Markov model for the binary covariate process in the survival data setting, where the covariates are assumed constant in each partitioned interval and the conditional transition probability is modeled by a logistic model. By jointly modeling the covariate process and the event time process, the missing covariate value can be interpolated by borrowing information from the entire covariate process. In this paper, we follow Faucett et al. (1998) to assume a discrete-Markov model for the binary covariate process, and investigate the joint modeling approach in the recurrent event setting.

Our analysis plan is motivated by an observational study designed to evaluate the effect of group A streptococcus (GAS) on the risk of developing pharyngitis (sore throat) among school children in India. Pharyngitis is one of the most common conditions encountered by physicians. It is most frequently due to viruses, but several bacteria, including Group A streptococci (GAS) persist as a common cause of pharyngitis. GAS pharyngitis is the cause of 37% of sore throats among children (Shaikh et al. 2010) and 5-15% in adults (Shulman et al. 2012). A study of streptococcus pharyngitis was undertaken in India between March 2002 and March 2004. In this study, cases of pharyngitis were identified weekly, and throat swabs were obtained on those with pharyngitis to identify the presence of GAS. Additionally, monthly throat cultures were obtained on all the children to determine the prevalence of streptococcal infection. Statistically, GAS colonization is a binary time-varying covariate with pharyngitis the recurrent event.

To evaluate the effects of the binary covariate process (GAS) on the recurrent event (pharyngitis), we apply the random-effect proportional intensity model (Lawless 1987). In this study, the regular throat cultures are obtained monthly while the event status are assessed weekly. As a result, the covariate

values at each event time are observed for children who had pharyngitis, but are possibly missing for other children in the corresponding risk set. Hence the partial likelihood method proposed by Lin et al. (2000) cannot be applied for model estimation. In this paper, we adopt the joint modeling approach to study the effect of GAS colonization on pharyngitis. Specifically, we assume a discrete-time Markov model for the covariate process, where the covariates are assumed constant in each partitioned small interval and the transition probability is modeled by a logistic model. We assume a proportional intensity model with frailty for the recurrent events, where the intensity is expressed as a function of the covariate history, both observed and unobserved. We use the Markov chain Monte Carlo technique of Gibbs sampling to estimate the joint posterior distribution of the unknown parameters of the model.

We compare the joint modeling approach with the last covariate carried forward approach, hereafter abbreviated as LCCF. In studies of recurrent events, it is common that covariate values are collected when an event occurs in addition to regular follow-up visits. Thus there are two possible ways to carry forward the last observation. One is to carry forward all covariates regardless of whether it is measured at event time or it is collected at regular follow-up visits. We call it all covariates carried forward (ACCF) method. The other approach is to only carry forward covariates from regular visits, and we call it carriage covariate carried forward (CCCF) method. Using the Indian pharyngitis study as an example, ACCF carries forward the GAS status from both weekly event visits and monthly carriage visits, while CCCF carries forward the GAS status from monthly carriage visits only. Research has shown that the ACCF approach or CCCF approach leads to biased inferential results of the covariate effects in the survival setting (Prentice 1982; Faucett et al. 1998). In this manuscript, we investigate the bias induced by the ACCF method or the CCCF method in recurrent event data analysis, and compare it to the proposed joint modeling approach.

The rest of the manuscript is organized as follows. Section 2 describes the formulation of the joint model. Section 3 gives the details of the Markov chain Monte Carlo estimation procedure. In Section 4, simulations are conducted to compare the bias and efficiency of the proposed joint modeling approach with that of the ACCF and CCCF approach. We apply the methods to the Indian study of GAS and pharyngitis in Section 5. Some concluding remarks are made in Section 6.

2 MODELS

Let subscript i be the index for a subject, $i = 1, 2, \ldots, n$. For subject i, let $Z_i(t)$ denote the covariate process. For the rest of the paper, we shall assume that $Z_i(t)$ is a univariate binary covariate process evolving in the time interval $[0, \tau]$. Extensions to multivariate covariate processes will be discussed towards the end of the manuscript. Following Faucett et al. (1998), we partition the study period into equally spaced small intervals, $I_1, I_2, ..., I_J$, and the binary

covariate process $Z_i(t)$ is assumed to be constant within each interval. Denote by Z_{ij} the value of the binary covariate in I_j. A discrete-time Markov model is assumed for the covariate process: the covariate distribution in the first time interval satisfies

$$\mathrm{pr}(Z_{i1} = 1) = \pi$$

and the transitional probability is modeled by

$$\mathrm{logit}\{\mathrm{pr}(Z_{ij} = 1 \mid Z_{i,j-1})\} = \alpha_0 + \alpha_1 Z_{i,j-1},$$

where π, α_0, and α_1 are unknown parameters.

We now formulate the model for the recurrent event process. For subject i, let $N_i^*(t)$ be the number of recurrent events occurring at or before time $t \in [0, \tau]$, where the recurrent events could potentially be observed beyond a prespecified time point τ. Thus the counting process $N_i^*(t)$ has a jump of size one when an event (such as the sore throat in the Indian pharyngitis study) occurs. Let C_i denote the time to loss to follow-up or end of the study for subject i, $i = 1, \ldots, n$. Define the counting process $N_i(t) = N_i^*(t \wedge C_i)$, where $a \wedge b = \min(a, b)$. Let $Y_i(t) = I(C_i \geq t)$ be the indicator for a subject being under observation at time t. Let $\mathcal{F}(t)$ be the σ-field generated by $\{Y_i(u), N_i(u), Z_i(u), 0 \leq u \leq t\}$, and $\lambda_Z(t)$ be the intensity function of $N_i^*(t)$, that is, $E[dN^*(t) \mid \mathcal{F}_{t-}] = \lambda_Z(t)dt$.

We assume random-effect intensity model for the recurrent events

$$\lambda_{Z_i}(t) = \lambda_0(t) \exp\{\beta Z_i(t) + U_i\}, \tag{1}$$

where β is the regression parameter, $\lambda_0(t)$ is an unspecified baseline intensity function, and U_i is an unobserved subject-specific frailty that characterizes population heterogeneity. Given $Z_i(\cdot)$, a larger value of U_i indicates more frequent occurrences of recurrent events. We assume U_i is a mean zero normal random variable with variance σ_U^2. When $\sigma_U^2 = 0$, the model reduces to the Anderson-Gill proportional intensity model (Anderson and Gill 1982). The baseline intensity function $\lambda_0(t)$ is assumed to be a step function on a predefined set of time intervals. For simplicity, we assume the time intervals of the baseline rate function overlap with the time intervals of the covariate process, that is, $\lambda_0^*(t) = \lambda_j$ for $t \in I_j, j = 1, ..., J$.

Let $d\mu_Z(t)$ be the marginal rate function of the recurrent events, that is, $d\mu_Z(t) = E[dN^*(t) \mid Z(t)]$. Lin et al. (2000) provided a rigorous estimation procedure for the proportional rate model $d\mu_Z(t) = \exp\{\beta Z_i(t)\}d\mu_0(t)$, which includes model (1) as a special case. Specifically, they proposed to estimate the regression parameter β by maximizing the log pseudo-partial likelihood:

$$l(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left(\{\beta Z_i(t)\} - \log \left[\sum_{k=1}^{n} Y_k(t) \exp\{\beta Z_k(t)\} \right] \right) dN_i(t),$$

which is equivalent to solving the (normalized) pseudo-partial score function

$$U(\beta) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} \left[Z_i(t) - \frac{\sum_{k=1}^{n} Z_k(t) \exp\{\beta Z_k(t)\} Y_k(t)}{\sum_{k=1}^{n} \exp\{\beta Z_k(t)\} Y_k(t)} \right] dN_i(t). \tag{2}$$

The construction of the partial score function requires that $Z_i(t)$ is completely observed on $[0, C_i]$, or, minimally, at each event time for all subjects in the corresponding risk set.

In practice, however, covariates are typically measured periodically. An ad-hoc approach is to simply fill in the missing values in function (2) with the last observed covariate values before each event time. However, the last covariate carried forward approach may lead to biased inferential results, especially when the last observation is distant in time. On the contrary, by jointly modeling the covariate process and the recurrent event process, we can essentially "interpolate" the covariate at each event time by borrowing strength of the entire history of the covariate process, thus we can obtain more accurate estimates. For estimation, we adopt a Bayesian approach with weakly informative priors. Markov Chain Monte Carlo (MCMC) method with Gibbs sampling is used to estimate the parameters $\{\beta, \pi, \alpha_0, \alpha_1, (\lambda_1, \ldots, \lambda_J), \sigma_U^2\}$. Details of the Bayesian estimation procedure are given in the next section.

3 Markov Chain Monte Carlo

To introduce more notations, we let m_i denote the number of time intervals up to the end of follow-up for subject i, and let d_{ij}, $j = 1, \ldots, m_i$, denote the indicator of the covariate value being observed in the jth time interval. Finally, let n_i denote the number of recurrent events that occur before the censoring time C_i and t_{ik}, $k = 1, \ldots, n_i$ denote the observed event times for subject i. The development of the MCMC procedure follows similar techniques as in Faucett et al. (1998).

The parameters to be estimated are $\boldsymbol{\Theta} = \{\beta, \pi, \alpha_0, \alpha_1, (\lambda_1, ..., \lambda_J), \sigma_U^2\}$. We specify the prior distribution for each λ_j to be Gamma with mean a/b and variance a/b^2. We choose the hyperparameter values $a = 0.1$ and $b = 0.1$, which gives a very non-informative prior. The prior distributions for all λ_j's are identically distributed, which corresponds to having no prior belief that the intensity function changes over time. Such assumptions have been commonly used in Bayesian analysis of time-to-event data, due to their simplicity and flexibility. Examples include Faucett et al. (1998), Sinha et al. (1999), Brown and Ibrahim (2003), Berry et al. (2004) and Ibrahim et al. (2013). As pointed out by Sinha et al. (1999), the length of each interval may be taken to be sufficiently small to approximate any hazard (or intensity) function for all practical purposes. We also specify weak priors for the other parameters. Specifically, the priors for α_0, α_1, β are normal with mean 0 and standard deviation 100, the prior for π is uniform over the interval $[0, 1]$, and the prior for σ_U^2 is $p(\sigma_U^2) \propto \sigma_U^{-2}$ (Box and Tiao 1973). We aim to make inference about $\boldsymbol{\Theta}$ based on the posterior distribution $p(\boldsymbol{\Theta} \mid \mathbf{OBS})$, where $\mathbf{OBS} = \{(Z_{ij}; d_{ij} = 1, j = 1, ..., m_i), (t_{ik}; k = 1, ..., n_i), C_i; i = 1, ...n\}$ are the observed data. For convenience, we use $p(\cdot \mid \cdot)$ to denote a general conditional probability density function. Due in part to the incompleteness of the whole covariate process, it is difficult to obtain $p(\boldsymbol{\Theta} \mid \mathbf{OBS})$. So we use

Gibbs sampling to draw samples from $p(\mathbf{\Theta}, \mathbf{MIS} \mid \mathbf{OBS})$, where $\mathbf{MIS} = \{(Z_{ij}; d_{ij} = 0, j = 1, ..., m_i), U_i; i = 1, ..., n\}$ are the unobserved data, and summarize the empirical marginal distributions of $\mathbf{\Theta}$ to approximate $p(\mathbf{\Theta} \mid \mathbf{OBS})$.

The validity of drawing samples from $p(\mathbf{\Theta}, \mathbf{MIS} \mid \mathbf{OBS})$ and marginalizing to approximate $p(\mathbf{\Theta} \mid \mathbf{OBS})$ requires some assumptions on the missing-data and censoring mechanisms. Let $\mathbf{Z_i} = \{Z_{ij}; j = 1, ..., m_i\}$ denote the covariate history for subject i, and let $\boldsymbol{\eta}$ denote the parameter for the distribution of C_i. When $\mathbf{Z_i}$ is fully observed (up to censoring), under the assumptions that censoring time C_i is non-informative about the recurrent event process given the covariate information and that the subject-specific frailty U_i is independent of $\mathbf{Z_i}$ and C_i, the likelihood can be expressed as

$$L_i(\mathbf{\Theta}, \boldsymbol{\eta} \mid t_{i1}, ..., t_{i,n_i}, \mathbf{Z_i}, C_i) \propto \int \Big(\Big[\prod_{k=1}^{n_i} \lambda_i\{t_{ik} \mid \mathbf{Z_i}, U_i, \mathbf{\Theta}\} \Big]$$

$$\times \exp \Big[-\int_0^{C_i} \lambda_i\{t \mid Z_i, U_i, \mathbf{\Theta}\} dt \Big] \Big) dp_U(U_i) \times p_C(C_i \mid \mathbf{Z}_i, \boldsymbol{\eta}) \times p_Z(\mathbf{Z}_i \mid \mathbf{\Theta})$$

where p_U, p_C, and p_Z are the density functions of U_i, C_i, and $\mathbf{Z}_i$.

When some values of the covariate $\mathbf{Z}_i$ are missing, $\mathbf{Z}_i$ can be partitioned into observed and missing values $\{\mathbf{Z}_{i,obs}, \mathbf{Z}_{i,mis}\}$. Let $\mathbf{d_i} = \{d_{i1}, ..., d_{im_i}\}$ be the missingness indicator and $\boldsymbol{\zeta}$ be the parameter for the missingness mechanism, then

$$L_i(\mathbf{\Theta}, \boldsymbol{\eta}, \boldsymbol{\zeta} \mid t_{i1}, ..., t_{in_i}, \mathbf{Z}_{i,obs}, C_i, \mathbf{d_i})$$

$$\propto \int p_d(\mathbf{d_i} \mid t_{i1}, ..., t_{in_i}, \mathbf{Z}_i, C_i, \boldsymbol{\zeta}) \times L_i(\mathbf{\Theta}, \boldsymbol{\eta} \mid t_{i1}, ..., t_{in_i}, \mathbf{Z}_i, C_i) d\mathbf{Z}_{i,mis} \quad (3)$$

where p_d is the conditional density function of $\mathbf{d_i}$. Under the assumptions that p_d does not depend on $\mathbf{Z}_{i,mis}$, that is, $\mathbf{Z}_{i,mis}$ are missing at random, and that C_i does not depend on $\mathbf{Z}_{i,mis}$, the likelihood (3) is proportional to the product of the following two terms:

$$p_C(C_i \mid \mathbf{Z}_{i,obs}, \boldsymbol{\eta}) \times p_d(\mathbf{d_i} \mid t_{i1}, ..., t_{in_i}, \mathbf{Z}_{i,obs}, C_i, \boldsymbol{\zeta})$$

and

$$\int \Big\{ \int \Big(\Big[\prod_{j=1}^{n_i} \lambda_i\{t_{ij} \mid \mathbf{Z}_i, U_i, \mathbf{\Theta}\} \Big] \times \exp \Big[-\int_0^{C_i} \lambda_i\{t \mid \mathbf{Z}_i, U_i, \mathbf{\Theta}\} dt \Big] \Big) dp_U(U_i) \Big\}$$

$$\times p_Z(\mathbf{Z}_i \mid \mathbf{\Theta}) d\mathbf{Z}_{i,mis}. \quad (4)$$

Assume $\mathbf{\Theta}$ is distinct from $(\boldsymbol{\eta}, \boldsymbol{\zeta})$ (Rubin 1976) in the sense that they lie in disjoint parameter spaces and any prior distributions for the joint parameters can be factorized into independent factors. By theorem 8.1 in Rubin (1976), the posterior distributions of $\mathbf{\Theta}$ are independent of the posterior distributions

of $(\boldsymbol{\eta}, \boldsymbol{\zeta})$. Therefore, the posterior distribution of $\boldsymbol{\Theta}$ used in the estimation procedure is proportional to the prior of $\boldsymbol{\Theta}$ times (4).

Under the above assumptions, we draw samples from $p(\boldsymbol{\Theta}, \mathbf{MIS} \mid \mathbf{OBS})$ using Gibbs sampling. Details of the full conditional distributions of each parameter and the methods for sampling from those conditional distributions are given in the Appendix. To monitor convergence of the Markov chain, we use the autocorrelation plots and the Geweke's approach (Geweke 1992). The lag k autocorrelation is the correlation between every draw and its kth lag. If autocorrelation is still high with large values of k, this indicates slow mixing and non-convergence. The Geweke's method compares the means of the early part and the late part of the Markov chain (usually the first 10% and the last 50%), using a difference of means test. If the value of the test statistics falls in the extreme tails of a standard normal distribution, it suggests the early part of the chain has not converged.

A sensitivity analysis is performed to determine if posterior inferences depend on how prior distributions are specified. Because the regression parameter β is the main parameter of interest, we focus on the posterior inferences of β. Let $\mathrm{N}(\mu, \sigma)$ denote a normal distribution with mean μ and standard deviation σ. We consider different types of priors for β: one is a diffuse and non-informative prior $\mathrm{N}(0, 100)$, and the other two are tight priors. One tight prior is chosen as $\mathrm{N}(\hat{\beta}_{\mathrm{LCCF}}, 0.5)$, where $\hat{\beta}_{\mathrm{LCCF}}$ is obtained by using the last covariate carried forward approach. This prior indicates a strong belief that the true parameter β does not deviate much from $\hat{\beta}_{\mathrm{LCCF}}$. The other tight prior is specified as $\mathrm{N}(0, 0.5)$, which indicates a strong belief that there is no covariate effect of Z on the recurrent event process. If the true parameter value for β is deviated from zero, the prior $\mathrm{N}(0, 0.5)$ could be a mis-specified prior. The prior distributions for the other parameters are all non-informative priors, as specified at the beginning of this section.

4 SIMULATIONS

We conduct simulation studies to evaluate the finite-sample performances of different estimators with moderate sample size. For each simulation setting, we generate 500 simulated datasets, each with 200 subjects. We compare the proposed estimator with the two last covariate carried forward estimators.

The baseline covariate values are generated from Bernoulli random variables with $\mathrm{pr}(Z_{i1} = 1) = 1/3$ and are observed for all subjects. For subsequent time intervals, the covariate values are generated from the logistic transition model $\mathrm{logit}\{\mathrm{pr}(Z_{ij} = 1 \mid Z_{i,j-1})\} = \alpha_0 + \alpha_1 Z_{i,j-1}$ with $\alpha_0 = -1.5$ and $\alpha_1 = 2$. The subject-specific random effect U_i which characterizes the population heterogeneity is simulated from a normal random variable with mean 0 and variance $\sigma_U^2 = 1/4$. Next, the recurrent events of a subject are generated from a Poisson process with intensity function $\lambda_0^*(t) \exp\{\beta Z_i(t) + U_i\}$, where $\beta = 0.5$ and $\lambda_0^*(t) = 0.1 I(t \leq 10) + 0.5 I(10 < t \leq 20)$. The covariate values at event times are always observed. In addition, each subject has 20 scheduled

visits on $[0, 20]$, with one visit per unit time interval. The time of visit in each interval is uniformly distributed, and the covariate values at regular visits are subject to missingness. To create missing $Z_i(t)$ at regular visits, we generate independent Bernoulli random variables for all the intervals for each person. If the Bernoulli random variable is 1, we set the associated $Z_i(t)$ value to missing. The follow-up time C_i is generated from a mixture of a degenerate random variable at 20 and an uniform random variable on $[10, 20]$ so that 50% of the subjects are expected to be lost to follow-up before the end of the study.

We compare the performances of three estimators of β: (a) JM, the joint modeling approach, (b) ACCF, the rate ratio estimator where all carriage and event covariates are carried forward, (c) CCCF, the rate ratio estimator where carriage covariates are carried forward. We compare the performances of these estimators under different levels of missingness of regular carriage visits, where the probability of missing a pre-scheduled visit is set to be 0%, 20%, 40%, and 60% at any visit.

For the JM approach, initial parameters for the Gibbs sampler are $\pi = 0.5$, $\alpha_0 = 0$, $\alpha_1 = 0$, $\beta = 1$, $\sigma_U^2 = 0.5$, $\lambda_j = 0.2, j = 1, ..., 20$. The initial values for the missing covariates $\{Z_{ij}; d_{ij} = 0, j = 1, ..., m_i, i = 1, ..., n\}$ are generated from a Bernoulli distribution with mean 0.5. The initial values for the frailty $U_i, i = 1, ..., n$ are generated from N(0, 0.5) distribution. The sampling procedures are repeated until 5000 samples are generated. The first 500 samples are discarded as "burn-in". To decrease autocorrelation, the samples are taken at every 5th value of the remaining chain. The posterior means are used as the parameter estimates. Both autocorrelations within chains of β and the Geweke's test are used to assess convergence of the Markov chains. Around 2% of the simulated datasets show significant p-values for the Geweke's diagnostics test, and they are excluded from analyses. The range of the lag-5 autocorrelations across the remaining datasets is -0.09 to 0.09, which confirms convergence of the chains. To examine the sensitivity of the posterior distributions to the choices of priors, we draw samples of β under three different priors: N(0, 100), N($\hat{\beta}_{\text{CCCF}}$, 0.5) and N(0, 0.5), where $\hat{\beta}_{\text{CCCF}}$ is obtained by using the CCCF approach. When the missingness probability is 40%, under three different priors, the empirical mean and empirical standard deviation of the estimates of β are 0.500 (0.073), 0.498 (0.070) and 0.487 (0.072) respectively, which shows that the posterior inferences of β are only slightly affected by the choice of priors. In the following, we report the performance of the JM estimator under the diffuse prior.

Table 1 summarizes the percent bias and the relative efficiency that compares the mean square error of an estimator to that of the rate ratio estimator under the perfect scenario where the covariate process is monitored continuously until loss to follow-up. Note that Monte-Carlo bias is the average difference between a parameter and its estimate, while the standard error is estimated by the empirical standard deviation of the estimates. Under the discrete-time Markov covariate process model, the JM approach shows small empirical bias and possesses high relative efficiency when the probability of missing a scheduled visit is low. Interestingly, because the mean covariate

Table 1 Simulation results for estimates of β by different methods with $n = 200$. Percent bias is the empirical bias as a percentage of the true value; relative efficiency is the ratio of the mean squared error of an estimator to that of the rate ratio estimator under the perfect scenario where the covariate process is monitored continuously until loss to follow-up. Abbreviations: JM, the joint modeling estimator; ACCF, the rate ratio estimator with all carriage and event covariates carried forward; CCCF, the rate ratio estimator with carriage covariates carried forward.

Proportion of missingness	Percent Bias (%)			Relative Efficiency		
	ACCF	CCCF	JM	ACCF	CCCF	JM
0%	-6	-2	0	0.74	0.82	1.00
20%	-10	-2	0	0.59	0.75	0.95
40%	-15	-2	0	0.38	0.54	0.78
60%	-22	-2	1	0.23	0.46	0.55

value is approximately constant under the assumed Markov model, the CCCF approach also has negligible bias, while the ACCF approach yields substantial bias. Some intuition for this behavior can be developed. Because the mean covariate value is approximately constant, at any time t with $\beta > 0$, an estimate of the prevalence rate using all covariate data will tend to be biased too high as events just prior to t will tend to have $Z = 1$. An estimate of the prevalence rate by just carrying forward the carriage data will not be biased and this differential behavior seems to have consequences for estimation of β as well. The JM estimator has higher efficiency than the ACCF and CCCF estimators, because it utilizes information in the longitudinal process and the event data process simultaneously.

5 DATA ANALYSIS

We apply the proposed joint modeling approach to the Indian pharyngitis study to determine the effect of GAS colonization on developing pharyngitis. This study was undertaken in a rural area near Vellore, India between March 2002 and March 2004. A total of 232 school children were enrolled in 2002, and 73 were enrolled in 2003 to replace the children who had left. Virtually all children in the school participated in the study. Cases of pharyngitis were identified weekly, where GAS colonization status were obtained on those with pharyngitis. Additionally, throat cultures were obtained monthly to determine the prevalence of GAS colonization. A two-week rule was applied to determine an episode of pharyngitis, that is, a pharyngitis event occurred within 14 days after a previous episode was considered as the same episode. The time origin for the recurrent event analysis is set to be the first day of the study, that is, March 11, 2002.

We analyze the data in the first year of the study, which included 232 school children who were enrolled in 2002. Figure 1 provides a graphical summary

Table 2 Results of throat swab cultures for the carriage and pharyngitis visits from 232 school-children.

Group A Strep	Carriage Data Monthly Visit	Event Data Pharyngitis Events
Negative	863	258
Positive	153	72
Total	1016	330

of the experience of five selected children volunteers. The blue circles represent carriage visits while red circles represent pharyngitis visits. Note that for children who had at least one event, GAS status can be obtained from both carriage visits and pharyngitis visits. As a comparison, Figure 2 shows the GAS status for the five children from carriage visits only. About 40% of throat cultures were missing at the carriage visits before end of follow-up. Table 2 presents a tally of outcomes based on the carriage and pharyngitis data sets. During the first year of the study, 330 pharyngitis events occurred. With 1016 monthly visits, the carriage data corresponds to a total follow-up of about 194 children years. About 22% of the throat cultures collected at pharyngitis visits were positive for GAS colonization, while 15% of the throat cultures obtained at the carriage visits were positive.

The ACCF method that carries forward all available covariate data yields a rate ratio of 1.28 (95% CI: 1.01–1.58), while the CCCF method that carries forward only carriage covariate yields a rate ratio of 1.33 (95% CI: 1.01–1.76). The joint modeling approach yields an estimated rate ratio of 1.39 (95% CI: 1.15–1.75), where priors and initial parameters are specified the same as in the simulation study. Both the lag-5 autocorrelation and the Geweke's test confirm good convergence of the Markov chains. Confidence intervals for all three estimators are obtained by the percentile method of the nonparametric bootstrap for clustered data with 200 bootstrap samples, where the sampling unit is the child. Since the simulations in the previous section show that the joint modeling approach is unbiased and most efficient, we advocate the joint modeling estimator for inference here. We conclude that there is significant effect of GAS on the risk of developing pharyngitis among school children. The risk of pharyngitis increases by 39% (CI: 15%–75%) for a child who is colonized with GAS.

6 REMARKS

This article proposes a joint modeling approach to estimate the effect of time-varying binary covariates on recurrent events when the covariates are measured periodically. We assume a Markov model on the covariate process and a random-effect proportional intensity model for the recurrent event process. The proposed approach complements the work of Faucett et al. (1998) by

Data from 5 children
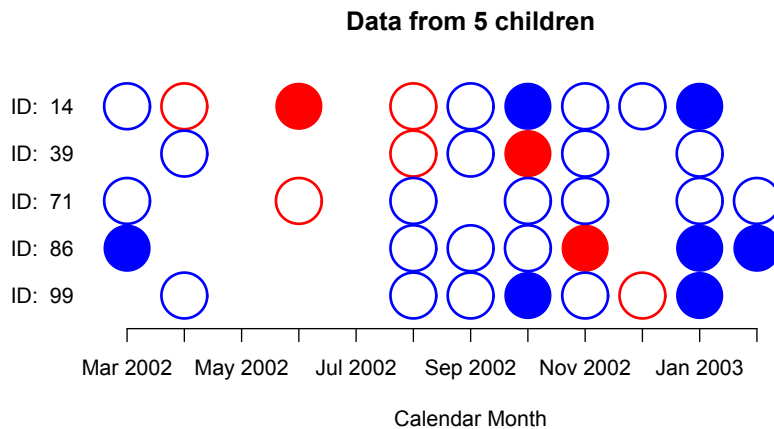
Fig. 1 GAS status from both carriage visits and pharyngitis visits for five selected children in the Vellore study. For each child, the blue circles denote carriage visits while red circles denote pharyngitis visits. Closed (open) circle denotes a throat culture positive (negative) for GAS. Should a pharyngitis event occurs in the same calendar month, we report its result rather than the carriage result

Data from 5 children
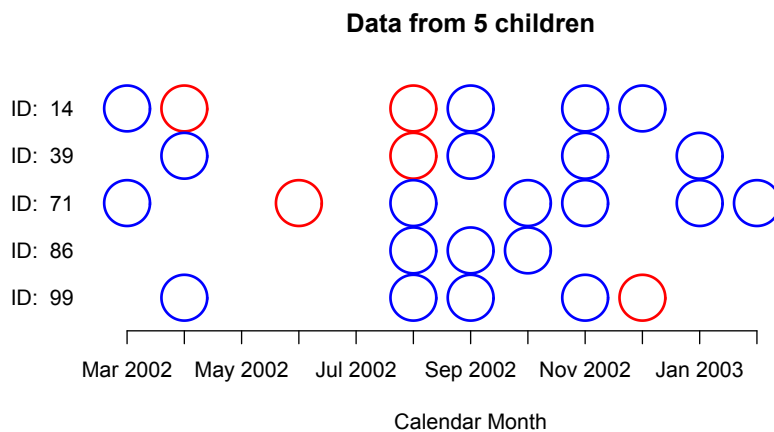
Fig. 2 GAS status from carriage visits for five selected children in the Vellore study. For each child, closed (open) circle denotes a throat culture positive (negative) for GAS

extending single-event time model to recurrent event process model and by allowing subject-specific random effects in the event time process. We compare the proposed approach with the last covariate carried forward approach. Simulation studies show that the joint modeling approach is unbiased and most efficient under the assumed Markov model, while the covariate carried forward approach may lead to substantial bias.

In this paper, we consider univariate binary covariate process. The extensions to multivariate binary process are straightforward, though computationally demanding. One can assume a Markov model for each binary covariate, and the intensity function for the recurrent event process can be written as a joint function of the multiple covariates.

As with all the model-based methods, the validity of the proposed approach depends on the validity of the models used, which could be difficult to verify when there is substantial missingness. To relax the model assumptions, one can consider the semiparametric transformation model for the recurrent event process, which includes proportional intensity model as a special case (Zeng and Lin 2006). For the binary covariate process, one can extend the Markov model to order q, that is, the transition probability can be written as a function of the q previous covariate values (Zeger and Qaqish 1988). Further investigation is warranted.

Appendix

In this Appendix, we describe the details to draw samples from $p(\mathbf{\Theta}, \mathbf{MIS} \mid \mathbf{OBS})$ using Gibbs sampling, following similar steps as in Faucett et al. (1998). The priors for the parameter $\mathbf{\Theta} = \{\beta, \pi, \alpha_0, \alpha_1, (\lambda_1, ..., \lambda_J), \sigma_U^2\}$ take the following forms.

$\pi \sim \text{Unif}(0,1)$
$\alpha_0 \sim N(\mu_{\alpha_0}, \sigma_{\alpha_0})$
$\alpha_1 \sim N(\mu_{\alpha_1}, \sigma_{\alpha_1})$
$\sigma_U^2 \propto \sigma_U^{-2}$
$\beta \sim N(\mu_\beta, \sigma_\beta)$
$\lambda_k \sim \text{Gamma}(a/b, a/b^2)$, for $k = 1, ..., J$.

Furthermore, in Section 2, we impose the following assumptions

$U_i \sim N(0, \sigma_U^2)$
$Z_{i1} \sim \text{Bernoulli}(\pi)$
$\text{logit}\{\text{pr}(Z_{ij} = 1 \mid Z_{i,j-1})\} = \alpha_0 + \alpha_1 Z_{i,j-1}$.

In Section 3, we show that the posterior distribution of $\mathbf{\Theta}$ is proportional to the prior of $\mathbf{\Theta}$ times (4). Therefore, the Gibbs sampler takes the following form:

1. Initialize $\mathbf{\Theta}$ and $\mathbf{MIS}$.
2. Sample π from $p(\pi \mid \{Z_{i1}; i = 1, ..., n\})$.
3. Sample α_0 from $p(\alpha_0 \mid \alpha_1, \{Z_{ij}; i = 1, ..., n, j = 1, ..., m_i\}, \mu_{\alpha_0}, \sigma_{\alpha_0})$; sample α_1 from $p(\alpha_1 \mid \alpha_0, \{Z_{ij}; i = 1, ..., n, j = 1, ..., m_i\}, \mu_{\alpha_1}, \sigma_{\alpha_1})$.
4. For $i = 1, ..., n$ and $k = 1, ..., m_i$, where $d_{ik} = 0$, sample Z_{ik} from $p(Z_{ik} \mid \{Z_{ij}; j \neq k\}, U_i, \pi, \alpha_0, \alpha_1, C_i, \{t_{il}; l = 1, ..., n_i\}, \{\lambda_j; j = 1..., J\}, \beta)$.

5. For $i = 1, ..., n$, sample U_i from $p(U_i \mid \{Z_{ij}; j = 1, ..., m_i\}, C_i, \{t_{il}; l = 1, ..., n_i\}, \{\lambda_j; j = 1..., J\}, \beta, \sigma_U^2)$.
6. Sample σ_U^2 from $p(\sigma_U^2 \mid U_i; i = 1, ...n)$.
7. Sample β from $p(\beta \mid \{Z_{ij}; j = 1, ..., m_i\}, C_i, \{t_{il}; l = 1, ..., n_i\}, U_i, \{\lambda_j; j = 1, ..., J\}, \mu_\beta, \sigma_\beta; i = 1, ..., n)$.
8. For $k = 1, ..., J$, sample λ_k from $p(\lambda_k \mid \{Z_{ij}; j = 1, ..., m_i\}, C_i, \{t_{il}; l = 1, ..., n_i\}, U_i, \beta, \{\lambda_l; l \neq k\}, a, b; i = 1, ...n)$.
9. Go to step 2.

The full conditional distribution and the sampling method for each step are given below.

1. $p(\pi \mid \{Z_{i1}; i = 1, ..., n\}) \propto p(\pi) \times p(\{Z_{i1}; i = 1, ..., n\} \mid \pi)$
$$\propto \pi^{\sum_{i=1}^n Z_{i1}} (1 - \pi)^{n - \sum_{i=1}^n Z_{i1}},$$
which is proportional to a Beta distribution. Therefore, π can be sampled from a Beta distribution with shape parameters $\sum_{i=1}^n Z_{i1}$ and $n - \sum_{i=1}^n Z_{i1}$.

2. $p(\alpha_0 \mid \alpha_1, \{Z_{ij}; i = 1, ..., n, j = 1, ..., m_i\}, \mu_{\alpha_0}, \sigma_{\alpha_0})$
$$\propto p(\alpha_0 \mid \mu_{\alpha_0}, \sigma_{\alpha_0}) \times p(\{Z_{ij}; i = 1, ..., n, j = 1, ..., m_i\} \mid \alpha_0, \alpha_1)$$
$$\propto p(\alpha_0 \mid \mu_{\alpha_0}, \sigma_{\alpha_0}) \times \prod_{i=1}^n \prod_{j=2}^{m_i} p(Z_{ij} \mid Z_{i,j-1}, \alpha_0, \alpha_1)$$
which is proportional to the product of

$$\prod_{i=1}^n \prod_{j=2}^{m_i} \left\{ \frac{\exp(\alpha_0 + \alpha_1 Z_{i,j-1})}{1 + \exp(\alpha_0 + \alpha_1 Z_{i,j-1})} \right\}^{Z_{ij}} \times \left\{ \frac{1}{1 + \exp(\alpha_0 + \alpha_1 Z_{i,j-1})} \right\}^{1 - Z_{ij}}$$

and

$$\exp\left\{ \frac{-(\alpha_0 - \mu_{\alpha_0})^2}{2\sigma_{\alpha_0}^2} \right\}.$$

Because the full conditional does not form a well-known distribution, we use the adaptive rejection sampling (Gilks and Wild 1992) to draw samples of α_0. Adaptive rejection sampling (ARS) is a commonly used sampling method which improves the rejection sampling by constructing proposal distributions that are adaptively refined using past rejected samples. The ARS algorithm is available in many software toolkits, and samples can be drawn using those toolkits as long as we specify the probabilistic models. The parameter α_1 can be sampled in an analogous manner.

3. For $i = 1, ..., n$ and $k = 1, ..., m_i$, where $d_{ik} = 0$, let $\Gamma_i = \{C_i, \{t_{il}; l = 1, ..., n_i\}\}$.
$p(Z_{ik} \mid \{Z_{ij}; j \neq k\}, U_i, \pi, \alpha_0, \alpha_1, C_i, \{t_{il}; l = 1, ..., n_i\}, \{\lambda_j; j = 1..., J\}, \beta)$
$$\propto p(\Gamma_i \mid \text{rest}) \times p(\{Z_{ij}; i = 1, ..., n, j = 1, ..., m_i\} \mid \alpha_0, \alpha_1, \pi)$$
$$\propto p(\Gamma_i \mid \text{rest}) \times \prod_{i=1}^n \prod_{j=2}^{m_i} p(Z_{ij} \mid Z_{i,j-1}, \alpha_0, \alpha_1) \times \prod_{i=1}^n p(Z_{i1} \mid \pi)$$
which is proportional to the product of

$$\exp\left[\sum_{l=1}^{n_i} \beta Z_i(t_{il}) - \int_{I_k} \lambda_k \exp(\beta Z_{ik} + U_i) du \right],$$

$$\left\{ \frac{\exp(\alpha_0 + \alpha_1 Z_{i,k-1})}{1 + \exp(\alpha_0 + \alpha_1 Z_{i,k-1})} \right\}^{Z_{ik}} \times \left\{ \frac{1}{1 + \exp(\alpha_0 + \alpha_1 Z_{i,k-1})} \right\}^{1 - Z_{ik}} \quad \text{for } k > 1,$$

$$\left\{\frac{\exp(\alpha_0 + \alpha_1 Z_{i,k})}{1 + \exp(\alpha_0 + \alpha_1 Z_{i,k})}\right\}^{Z_{i,k+1}} \times \left\{\frac{1}{1 + \exp(\alpha_0 + \alpha_1 Z_{i,k})}\right\}^{1-Z_{i,k+1}} \quad \text{for } k < m_i,$$

and $\pi^{Z_{ik}}(1-\pi)^{1-Z_{ik}}$ if $k = 1$.

Since Z_{ik} is binary, we draw Z_{ik} from a Bernoulli distribution with mean $p_1/(p_1+p_0)$, where p_1 and p_0 are obtained by evaluating the above function at $Z_{ik} = 1$ and $Z_{ik} = 0$ respectively.

4. For $i = 1, ..., n$, let $\Gamma_i = \{C_i, \{t_{il}; l = 1, ..., n_i\}\}$.

   $p(U_i \mid \{Z_{ij}; j = 1, ..., m_i\}, C_i, \{t_{il}; l = 1, ..., n_i\}, \{\lambda_j; j = 1...,J\}, \beta, \sigma_U^2)$
   $\qquad \propto p(\Gamma_i \mid \text{rest}) \times p(U_i \mid \sigma_U^2)$

   which is proportional to the product of

   $$\exp\left\{n_i U_i - \int_{I_j; 1 \le j \le m_i} \lambda_j \exp(\beta Z_{ij} + U_i) du\right\} \quad \text{and} \quad \exp(-U_i^2/2\sigma_U^2).$$

   Similarly as in Step 2, we draw samples of U_i using adaptive rejection sampling.

5. $p(\sigma_U^{-2} \mid \{U_i; i = 1, ...n\}) \propto p(\sigma_U^{-2}) \times p(\{U_i; i = 1, ...n\} \mid \sigma_U^{-2})$
   $\qquad \propto \sigma_U^2 \times \sigma_U^{-2n} \exp\{-\sum_{i=1}^n U_i^2/2\sigma_U^2\}$

   which is proportional to a Chi-Square distribution with parameters $S = \sum_{i=1}^n U_i^2$ and degree of freedom n. Therefore, samples of σ_U^{-2} are drawn from a Chi-Square distribution and σ_U^2 can be obtained by taking the inverse of σ_U^{-2}.

6. Let $\Gamma_i = \{C_i, \{t_{il}; l = 1, ..., n_i\}\}$.

   $p(\beta \mid \{Z_{ij}; j = 1, ..., m_i\}, \Gamma_i, U_i, \{\lambda_j; j = 1, ..., J\}, \mu_\beta, \sigma_\beta; i = 1, ..., n)$
   $\qquad \propto p(\Gamma_i; i = 1, ..., n \mid \text{rest}) \times p(\beta \mid \mu_\beta, \sigma_\beta)$, which is proportional to the product of

   $$\prod_{i=1}^n \exp\left\{\sum_{l=1}^{n_i} \beta Z_i(t_{il}) - \int_{I_j; 1 \le j \le m_i} \lambda_j \exp(\beta Z_{ij} + U_i) du\right\}$$

   and $\exp\{-(\beta - \mu_\beta)^2/(2\sigma_\beta^2)\}$. Similarly as in Step 2, samples from this distribution can be generated using adaptive rejection sampling.

7. Let $\Gamma_i = \{C_i, \{t_{il}; l = 1, ..., n_i\}\}$. For $k = 1, ..., J$,

   $p(\lambda_k \mid \{Z_{ij}; j = 1, ..., m_i\}, \Gamma_i, U_i, \beta, \{\lambda_l; l \ne k\}, a, b; i = 1, ...n)$
   $\qquad \propto p(\Gamma_i \mid \text{rest}) \times p(\lambda_j \mid a, b)$
   $\qquad \propto \prod_{i=1}^n \left[\left\{\prod_{j=1}^{n_i} \lambda_j(t_{ij} \mid \mathbf{Z}_i, U_i, \beta)\right\} \exp\left[-\int_0^{C_i} \lambda_j\{t \mid \mathbf{Z}_i, U_i, \beta\} dt\right]\right]$
   $\qquad \quad \times \lambda_k^{a-1} \exp(-b\lambda_k)$
   $\qquad \propto \lambda_k^{D_k+a-1} \exp\{-(Y_k + b)\lambda_k\}$,

   where $D_k = \sum_{i=1}^n \sum_{j=1}^{n_i} \mathcal{I}_{\{t_{ij} \in I_k\}}$ and $Y_k = \sum_{i=1}^n \mathcal{I}_{\{k \le m_i\}} \int_{I_k} \exp\{\beta Z_{ik} + U_i\} du$. The function $\mathcal{I}_{\{t_{ij} \in I_k\}}$ indicates that the jth event time of subject i occurred during interval I_k, and $\mathcal{I}_{\{k \le m_i\}}$ indicates that subject i was followed during interval I_k. This is proportional to a gamma density. Therefore, λ_k can be sampled from a Gamma distribution with parameters $D_k + a$ and $Y_k + b$.

References

1. Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The annals of statistics*, 1100-1120.
2. Albert, P. S. (2000). A transitional model for longitudinal binary data subject to nonignorable missing data. *Biometrics*, 56(2), 602-608.
3. Berry, S. M., Berry, D. A., Natarajan, K., Lin, C. S., Hennekens, C. H., & Belder, R. (2004). Bayesian survival analysis with nonproportional hazards: metanalysis of combination pravastatin-Aspirin. *Journal of the American Statistical Association*, 99(465), 36-44.
4. Borgan, Ø., Fiaccone, R. L., Henderson, R., & Barreto, M. L. (2007). Dynamic analysis of recurrent event data with missing observations, with application to infant diarrhoea in Brazil. *Scandinavian Journal of Statistics*, 34(1), 53-69.
5. Box, G. E., & Tiao, G. C. (2011). *Bayesian inference in statistical analysis* (Vol. 40). John Wiley & Sons.
6. Brown, E. R., & Ibrahim, J. G. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59(2), 221-228.
7. Clayton, P., McDonald, S., & Chadban, S. (2011). Steroids and recurrent IgA nephropathy after kidney transplantation. *American Journal of Transplantation*, 11(8), 1645-1649.
8. Cox, D. R.,& Snell, E. J. (1989). *Analysis of binary data* (Vol. 32). CRC Press.
9. Faucett, C. L., & Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in medicine*, 15(15), 1663-1685.
10. Faucett, C. L., Schenker, N., & Elashoff, R. M. (1998). Analysis of censored survival data with intermittently observed time-dependent binary covariates. *Journal of the American Statistical Association*, 93(442), 427-437.
11. Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *In Bayesian Statistics*, pp. 169-193, New York: Oxford University Press.
12. Gilks, W. R., & Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Journal of the Royal Statistical Society. Series C*, 41 (2), 337-348.
13. Henderson, R., Diggle, P., & Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4), 465-480.
14. Humphreys, K., & Titterington, D. M. (2003). Variational approximations for categorical causal modeling with latent variables. *Psychometrika*, 68(3), 391-412.
15. Ibrahim, J. G., Chen, M. H., Zhang, D., & Sinha, D. (2013). Bayesian Analysis of the Cox Model. *Handbook of Survival Analysis*, 27.
16. Kim, S., Zeng, D., Chambless, L., & Li, Y. (2012). Joint models of longitudinal data and recurrent events with informative terminal event. *Statistics in biosciences*, 4(2), 262-281.
17. Lawless, J. F. (1987). Regression methods for Poisson process data. *Journal of the American Statistical Association*, 82(399), 808-815.
18. Lin, D. Y., Wei, L. J., Yang, I., & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4), 711-730.
19. Liu, L., Ma, J. Z., & O'Quigley, J. (2008). Joint analysis of multilevel repeated measures data and survival: an application to the end stage renal disease (ESRD) data. *Statistics in medicine*, 27(27), 5679-5691.
20. Liu, L., & Huang, X. (2009). Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1), 65-81.
21. Palta, M., & Lin, C. Y. (1999). Latent variables, measurement error and methods for analysing longitudinal binary and ordinal data. *Statistics in Medicine*, 18(4), 385-396.
22. Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69(2), 331-342.
23. Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, 63 (2), 581-592.
24. Shaikh, N., Leonard, E., & Martin, J. M. (2010). Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis. *Pediatrics*, 126(3), e557-e564.

25. Shulman, S. T., Bisno, A. L., Clegg, H. W., Gerber, M. A., Kaplan, E. L., et al. (2012). Clinical practice guideline for the diagnosis and management of group A streptococcal pharyngitis: 2012 update by the Infectious Diseases Society of America. *Clinical infectious diseases*, 55(10), e86-e102.
26. Sinha, D., Chen, M. H., & Ghosh, S. K. (1999). Bayesian Analysis and Model Selection for Interval Censored Survival Data. *Biometrics*, 55(2), 585-590.
27. Sun, L., Song, X., Zhou, J., & Liu, L. (2012). Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *Journal of the American Statistical Association*, 107(498), 688-700.
28. Tsiatis, A. A., & Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(3), 809-834.
29. Wang, Y., & Taylor, J. M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96(455), 895-905.
30. Wulfsohn, M. S., & Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 330-339.
31. Zeger, S. L., & Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, 1019-1031.
32. Zeng, D., & Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 93(3), 627-640.