A Bayesian model for longitudinal count data with non-ignorable dropout

Niko A. Kaciroti, Trivellore E. Raghunathan, M. Anthony Schork and Noreen M. Clark

University of Michigan, Ann Arbor, USA

Summary. Asthma is an important chronic disease of childhood. An intervention programme for managing asthma was designed on principles of self-regulation and was evaluated by a randomized longitudinal study. The study focused on several outcomes, and, typically, missing data remained a pervasive problem. We develop a pattern–mixture model to evaluate the outcome of intervention on the number of hospitalizations with non-ignorable dropouts. Pattern–mixture models are not generally identifiable as no data may be available to estimate a number of model parameters. Sensitivity analyses are performed by imposing structures on the unidentified parameters. We propose a parameterization which permits sensitivity analyses on clustered longitudinal count data that have missing values due to non-ignorable missing data mechanisms. This parameterization is expressed as ratios between event rates across missing data patterns and the observed data pattern and thus measures departures from an ignorable missing data mechanism. Sensitivity analyses are performed within a Bayesian framework by averaging over different prior distributions on the event ratios. This model has the advantage of providing an intuitive and flexible framework for incorporating the uncertainty of the missing data mechanism in the final analysis.

Keywords: Gibbs sampling; Longitudinal data; Non-linear mixed effects models; Poisson outcomes; Randomized trials; Transition Markov models

1. Introduction

In longitudinal studies complete follow-up data are often not available for all subjects. Several approaches are available for analysing these incomplete data, e.g. mixed effects models or imputation-based techniques. Using such methods, however, inferences are only valid when the missing data mechanism is ignorable, i.e. we can correctly condition on variables that are necessary to yield a missingness at random (MAR) mechanism. When the missing data mechanism is non-ignorable, inferences based on only the observed data will not be valid. Thus, analysing such data requires more complex models which incorporate the missing data mechanism in the analysis. Two broad approaches are available: selection models and pattern–mixture models (Little and Rubin, 2002). These two approaches arise from different partitions of the observables y and the missing data indicator R. Selection models partition the joint distribution of $\Pr(Y, R)$ as the product of $\Pr(Y)$ and $\Pr(R|Y)$ (Heckman, 1979; Little, 1995; Kenward, 1998). They require explicit modelling of the missing data mechanism where the probability that a subject would drop out depends on the unobserved values. Pattern–mixture models (Little, 1993; Little and Rubin, 2002), in contrast, express the joint distribution as the product of $\Pr(Y|R)$ and $\Pr(R)$. Then they stratify the data by dropout patterns and allow distinct model parame-

Address for correspondence: Niko A. Kaciroti, 300 N. Ingalls Building, 10th Floor, Center for Human Growth and Development, University of Michigan, Ann Arbor, MI 48109, USA.
E-mail: nicola@umich.edu

ters for each stratum. The marginal estimates in pattern–mixture models can be derived as a weighted average across pattern-specific estimates (Little, 1995) or by using multiple imputation (Demirtas and Schafer, 2003). Regardless of which partition is used, additional assumptions or data are needed to identify the parameters in the joint distribution. Pattern–mixture models are commonly used as they do not require specific modelling of the dropout mechanism and the estimates of the identified parameters are not affected by the nature of the dropout mechanism.

Little (1995), Little and Wang (1996), Molenberghs *et al.* (1998), Daniels and Hogan (2000) and Kenward *et al.* (2003) identified parameters in the pattern–mixture model by using constraints. In a model with no constraints, Demirtas (2005) used a Bayesian smoothed pattern–mixture model for normal outcomes. Other approaches for identifying the parameters in pattern–mixture models have been proposed by Wu and Carroll (1988), Little (1994), Hogan and Laird (1997), Albert and Follmann (2000) and Guo *et al.* (2004), who all used latent random effects to relate the response and the missing data indicator. For a detailed literature review see Little (1995), Kenward and Molenberghs (1999) and Thijs *et al.* (2002).

Here we propose a Bayesian pattern–mixture model for analysing clustered longitudinal count data with non-ignorable dropouts. The model is identified by using easy-to-understand parameters, namely, ratios of event rates across missing data patterns with the observed data pattern as the reference group. Each parameter, which we refer to as an ignorability index, provides an intuitive way to capture the effect of a non-ignorable missing data mechanism and is easily used for sensitivity analyses. We have used similar parameters previously in pattern–mixture models for ordinal outcomes (Kaciroti, 2002, Kaciroti *et al.*, 2006). Now we extend the use of such ignorability index parameters for Poisson outcomes. Because the ignorability index cannot be defined by using only the observed data, we introduce an informative prior distribution, whereby the prior distribution reflects the nature of the missing data mechanism. By using a prior distribution we can incorporate in the final inferences any uncertainty, as well as prior knowledge related to the missing data mechanism. Within this framework, models with missing data generated by an ignorable missing data mechanism are a special case, where the ignorability index parameters are set equal to 1. An additional feature of the particular application that is considered in this paper is clustering of subjects. We account for clustering by introducing random-effects parameters. Bayesian inferences are constructed by using Markov chain Monte Carlo simulations.

The model proposed is motivated by, and fitted to, data from an asthma intervention study, which is described in Section 2. In Section 3 the complete-data model is defined and the pattern–mixture model with a potentially non-ignorable missing data mechanism is proposed. A Gibbs sampling algorithm for fitting such models is described in Section 4. The model is then applied to the asthma data in Section 5. A conclusion is given in Section 6.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

```
http://www.blackwellpublishing.com/rss
```

2.   Asthma intervention study

Asthma is the most common chronic disease of childhood; for example in the USA it affects an estimated 9 million children under age 18 years (National Center for Health Statistics, 2002). Thus, managing asthma is important for both reducing the medical costs as well as for improving quality of life. The intervention that is considered here focused on educating physicians about establishing strong partnerships with asthma patients and their families. The intervention took the form of an interactive seminar between general practice paediatricians and their

asthma patients and was based on the theoretical principles of self-regulation (Clark *et al.*, 1998). The efficacy of the intervention was evaluated by using a randomized study with the following outcomes:

(a) treatment practices and communication behaviour of physicians;
(b) health status and medical care use by their asthma patients;
(c) satisfaction of the patients' parents with the medical care.

The intervention programme has already been shown to decrease health care usage over a 2-year period (Clark *et al.*, 2000).

In this paper we explore in more detail how the intervention mechanism accomplished a decrease in health care use over time. We focus on three important questions.

(a) Does the effect of the intervention vary between the first and the second year and, if yes, how?
(b) Does the effect of the intervention vary with the initial severity?
(c) Are the results sensitive to the missing data assumptions?

Answering such questions should result in a better understanding of how the intervention works and could be a starting point to improve future interventions. We address these questions by using a transition Markov model with random effect, as described in Section 3. Our broader methodological aim is to develop statistical models for clustered longitudinal Poisson outcomes with incomplete data.

In this study, physicians were randomized into either an intervention group (38 physicians) or a control group (36 physicians); both groups were compared at two post-intervention follow-ups. The time between each wave was up to 12 months and varied between subjects. Out of the 74 physicians, seven (with a total of 20 patients) decided not to participate (four in the intervention group and three in the control group). Data on hospitalizations were available for 74 physicians and 635 patients at baseline ($t = 1$), 67 physicians and 446 patients at first follow-up ($t = 2$) and 67 physicians and 302 patients at second follow-up ($t = 3$). No information was available about the reasons why patients dropped out during the first follow-up period, other than for 20 patients who dropped out because their physicians withdrew. For the 144 patients who dropped out after first follow-up, 68% was due to disconnected telephones or families moving, 22% no longer had asthma symptoms and about 10% reported other reasons, such as dislike of the research study.

Although the missing data pattern was essentially monotone, the methods that are developed here could be generalized into any pattern of missing data. We classified the subjects into one of three broad patterns: $r = 1$ ($n = 189$) for patients with observed data only on y_1; $r = 2$ ($n = 144$) for patients with observed data only on y_1 and y_2; $r = 3$ ($n = 302$) for complete cases who provided data on y_1, y_2 and y_3.

The study was subject to considerable missing data; therefore, it is important to investigate the nature of the missing data mechanism. Following Ridout (1991) and Diggle *et al.* (2002) we used a logistic regression model to predict the probability of dropout at first and second follow-up. All demographic factors (i.e. parental income, age, education, sex and race), treatment indicator and medication use were initially included in each model. In addition, the number of hospitalizations at baseline and at first follow-up, and the change from baseline to first follow-up, plus their interaction with treatment, were used as independent variables.

For the first follow-up period only parents' income and baseline medication intake predicted the dropout during this period. Patients whose parents had lower incomes or were not taking medication at baseline were more likely to drop out between the baseline and first follow-up.

During the first to second follow-up period patients who were not taking prescription medication at first follow-up were more likely to drop out during this period. The dropout process was different between the two groups. In the control group, patients whose number of hospitalizations increased at first follow-up compared with baseline were more likely to drop out, whereas, in the intervention group, patients whose number of hospitalizations increased at first follow-up compared with baseline were more likely to remain in the study. Such findings show evidence that dropouts differ from the subjects who remained in the study. Therefore, when performing data analysis, it was important to investigate the sensitivity of the results to potentially non-ignorable missing data mechanisms.

3. The model

To evaluate the intervention effect over time on y, the number of hospitalizations, we propose a transition Markov model of first order with random intercept, similar to the model that was proposed by Zeger and Qaqish (1988). The random intercept is used to model the correlation across subjects having the same physician. The within-subject serial correlation is modelled by the transition Markov model, in which the expected response at a given time depends not only on the associated covariates but also on past responses. The analysis addresses the per-protocol question, under an ignorable missing data mechanism where all the randomized subjects comply with the treatment assigned.

3.1. Complete-data model

We model the number of asthma-related hospitalizations at each follow-up time by using Poisson regression. Let y_{ijt} be the number of asthma-related hospitalizations for patient j who is under the care of physician i (cluster) at time $t = 1, 2, \ldots, T$. Let $\mathbf{y}_{ijt} = (y_{ij1}, y_{ij2}, \ldots, y_{ijt})$ be the collection of the responses up to, and including, time t. Let x_{ij} be the set of fixed covariates; then the joint distribution of the follow-up responses $(y_{ij2}, y_{ij3}, \ldots, y_{ijT})$ for subject j in cluster i conditioned on x_{ij} and the baseline measure y_{ij1} can be factorized as

$$\Pr(y_{ij2}, y_{ij3}, \ldots, y_{ijT} | x_{ij}, y_{ij1}, \beta, b_i) = \prod_{t=2}^{T} \Pr(y_{ijt} | x_{ij}, \mathbf{y}_{ijt-1}, \beta_t, b_{it}),$$

where $\beta = (\beta_2, \beta_3, \ldots, \beta_T)$ is the collection of regression coefficients, and $b_i = (b_{i2}, b_{i3}, \ldots, b_{iT})$ is the collection of random effects. We assume that the distribution of y_{ijt} for $t = 2, 3, \ldots, T$, conditional on $u_{ijt} = (x_{ij}, \mathbf{y}_{ijt-1})$, and the random effects, b_{it}, is Poisson with mean μ_{ijt} modelled by

$$\log(\mu_{ijt}) = u_{ijt}\beta_t + b_{it} + o_{ijt} \tag{1}$$

where $\beta_t = (\beta_{t0}, \beta_{t1}, \ldots, \beta_{tp})$ and p is the number of predictors. Since the observation period may not be the same across all the individuals, an offset term o_{ijt} is introduced. Specifically, $o_{ijt} = \log(n_{ijt})$ where n_{ijt} is the number of months (or any time unit) over which y_{ijt} events have been reported. Random effects b_{it} are introduced to account for the correlation due to clustering. We assume that random effects $b_i = (b_{i2}, b_{i3}, \ldots, b_{iT})$ are independently identically multivariate normally distributed with mean 0 and covariance matrix Σ, for $i = 1, \ldots, K$, where K is the number of physicians. Let $b = (b_1, b_2, \ldots, b_K)$; thus the joint posterior distribution is

$$\Pr(\beta, \Sigma, b | y, x) \propto \prod_{i=1}^{K} \prod_{t=2}^{T} \prod_{j=1}^{n_{it}} \Pr(y_{ijt} | u_{ijt}, \beta_t, b_{it}) \, \phi(b_i) \times p(\beta, \Sigma), \tag{2}$$

where $\phi(b_i) = (2\pi)^{-T/2}|\Sigma|^{-1/2}\exp(-b_i\Sigma^{-1}b_i^T/2)$ and n_{it} is the number of patients who are seen by the ith physician at time t. To complete the model specification, a diffuse but proper prior distribution for β and Σ, $p(\beta,\Sigma)$, is assumed, with β having a diffuse normal prior with mean 0 and some large variance. The prior for variance–covariance matrix, Σ, follows an inverse Wishart distribution, $\Sigma \sim \text{IW}(R,\nu)$, where R is a prior guess of the magnitude of Σ, and ν is a number larger than $\dim(\Sigma) + 1$ (Spiegelhalter *et al.*, 2003). The primary parameter of interest is β, but other parameters (Σ, b_i) are also of interest. Given the complexity of the model, inferences are based on simulation techniques. For example, Gibbs sampling or other Markov chain Monte Carlo methods can be used to construct inferences on the basis of values drawn from the posterior distribution (2).

When there are missing values in y, and if the missing data mechanism is ignorable, Gibbs sampling for the complete-data model can be easily modified as described in Section 4.2.

3.2. Pattern–mixture model for non-ignorable missing data mechanisms

When the missing data mechanism is non-ignorable we use pattern–mixture models to derive inferences. Thus, we assume that model (1) applies to each missing data pattern but allow β-parameters to differ across patterns. Let $\beta_t^{(r)}$ denote the parameters of model (1) for missing data pattern r at time t, where r indicates the time of last measurement with $r = T$ corresponding to completers. Because there are no data to estimate all parameters $\beta_t^{(r)}$, for $r < t$, the pattern–mixture model is underidentified. Thus, restrictions or prior information about parameters in the model are required. Let $\beta_t^{(0)}$ be the identified parameters at time t corresponding to the observed data pattern at time t ($r \geqslant t$). Following Little and Rubin (2002), we specify a prior distribution $p(\beta_t^{(r)}|\beta_t^{(0)})$ on the unidentified parameters, $\beta_t^{(r)}$, $r < t$, conditioned on the identified parameters $\beta_t^{(0)}$. We have used a similar approach previously to identify pattern–mixture models with a non-ignorable missing data mechanism for ordinal outcomes (Kaciroti *et al.*, 2006). In that situation the prior distribution was constructed by relating the distribution of the missing data to the distribution of the observed data on the basis of the differences in the cumulative odds. Here we extend the same method for Poisson outcomes with non-ignorably missing data by relating the event rates in the missing data patterns with the event rate in the observed data patterns.

Specifically, let $\mu_t^{(r)} = E(Y_t^{(r)}|u_t, \beta_t^{(r)}, b_t)$ for pattern r at time t. Then, there is some function of u_t, $\tilde{\lambda}_t^{(r)}(u_t)$, such that, for $r = 1, 2, \ldots, t-1$ and $t \geqslant 2$,

$$\mu_t^{(r)} = \tilde{\lambda}_t^{(r)}(u_t)\mu_t^{(0)} \tag{3}$$

where $\mu_t^{(0)}$ is the mean at time t for the observed data at time t. Here $\tilde{\lambda}_t^{(r)}(u_t)$ is the ratio between the event rate in the rth missing data pattern and the event rate in the observed data pattern at time t; and it measures the departure for ignorable dropout. Further, it can be seen as a relative risk of the rth missing data pattern with the observed data pattern as the reference group. We assume that $\tilde{\lambda}_t^{(r)}(u_t)$ has a log-normal distribution with mean $l_t^{(r)}(u_t)$ and variance $c^2 l_t^{2(r)}(u_t)$ where c is the coefficient of variation. In this approach, the uncertainty in the relationship between the distribution of the missing data and the distribution of the observed data is captured by the prior distribution (probabilistic range) that is given to $\tilde{\lambda}_t^{(r)}(u_t)$.

The distribution $p(\beta^{(r)}|\beta^{(0)})$ is derived on the basis of the prior distribution of $\tilde{\lambda}_t^{(r)}$. Let $\beta_t^{(r)} = (\beta_{t0}, \beta_{t1}^{(r)})$, where β_{t0} is the set of parameters that are the same in missing data pattern r and the observed data pattern at time t and $\beta_{t1}^{(r)}$ is the set of parameters in pattern r that are different from the corresponding parameters in the observed data pattern. Let u_{t0} be the set of covariates that are associated with β_{t0} and u_{t1} be the set of covariates that are associated with $\beta_{t1}^{(r)}$. Thus, we have

$$\exp(u_{t0}\beta_{t0} + u_{t1}\beta_{t1}^{(r)} + b_t + o_t) = \tilde{\lambda}_t^{(r)}(u_t)\exp(u_{t0}\beta_{t0} + u_{t1}\beta_{t1}^{(0)} + b_t + o_t), \tag{4}$$

$$\exp(u_{t1}\beta_{t1}^{(r)}) = \tilde{\lambda}_t^{(r)}(u_{t1})\exp(u_{t1}\beta_{t1}^{(0)}).$$

We then consider the case where u_{t1} is a vector of dummy variables, $u_{t1} = (u_{t11}, u_{t21}, \ldots, u_{tp1})$, and where each dummy indicator corresponds to a subgroup. Continuous variables can be categorized into groups, and then we can proceed as in the case of categorical variables. We assume that for each subgroup that is identified by $u_{tk1} = 1$, $k = 1, \ldots, p$, the number of subjects with observed values at time t is non-zero. Let $\tilde{\lambda}_{\beta_{tk}}^{(r)} = \tilde{\lambda}_t^{(r)}(u_{tk1} = 1)$ be the ratio between the event rates of the missing data pattern and the observed data pattern at time t for the subgroup that is identified by $u_{tk1} = 1$ for $k = 1, \ldots, p$. Then

$$\exp(\beta_{tk}^{(r)}) = \tilde{\lambda}_{\beta_{tk}}^{(r)}\exp(\beta_{tk}^{(0)}) \tag{5}$$

or

$$\beta_{tk}^{(r)} = \log(\tilde{\lambda}_{\beta_{tk}}^{(r)}) + \beta_{tk}^{(0)}. \tag{6}$$

From equation (6), the prior distribution $p(\beta_{tk}^{(r)}|\beta_{tk}^{(0)})$ is defined on the basis of the distribution of $\tilde{\lambda}_{\beta_{tk}}^{(r)}$-parameters. Thus, the identifiability of the pattern–mixture model is translated into defining a distribution on $\tilde{\lambda}_{\beta_{tk}}^{(r)}$ for each subgroup, identified by u_{tk1}, $k = 1, 2, \ldots, p$, at time $t = 2, 3, \ldots, T$. Giving a distribution to $\tilde{\lambda}_{\beta_{tk}}^{(r)}$ is easy to understand rather than working directly with β_{tk}. For instance, let u_{t1} be the indicator for intervention, which corresponds to β_{t1}. Then $\tilde{\lambda}_{\beta_{t1}}^{(r)} \sim$ log-normal with mean $l = 0.5$ and $c = 0.1$ indicates that in the intervention group, on average, for a subject who was in missing data pattern r, the adjusted event rate of y_t is half (95% confidence interval CI $= (0.41, 0.61)$) of that for a subject in the observed data pattern. Then inferences derived on the basis of this $\tilde{\lambda}_{\beta_{t1}}^{(r)}$ would be approximately valid even when the missing data mechanism is non-ignorable but is within the range that is identified by $\tilde{\lambda}_{\beta_{t1}}^{(r)}$. The c-parameter captures the uncertainty that is related to the missing data mechanism, i.e. the range of $\tilde{\lambda}_{\beta_{t1}}^{(r)}$. Thus, in the above example, if $c = 0.5$, the 95% CI of $\tilde{\lambda}_{\beta_{t1}}^{(r)}$ would be wider, CI $= (0.16, 1.16)$.

The log-normal distribution family is an attractive choice for $\tilde{\lambda}_{\beta_{tk}}^{(r)}$ as it yields a normal prior distribution for $\beta_{tk}^{(r)}$, although other distributions for $\tilde{\lambda}_{\beta_{tk}}^{(r)}$ are possible. Under the log-normal distribution for $\tilde{\lambda}_{\beta_{tk}}^{(r)}$ the distribution $p(\beta_{tk}^{(r)}|\beta_{tk}^{(0)})$ is $\mathcal{N}\{E(\beta_{tk1}^{(r)}), \text{var}(\beta_{tk1}^{(r)})\}$, where $E(\beta_{tk1}^{(r)})$ and $\text{var}(\beta_{tk1}^{(r)})$ are derived by using equations (5) and (6). On the basis of equation (5) we obtain

$$\exp\{E(\beta_{tk}^{(r)}) + \text{var}(\beta_{tk}^{(r)})/2\} = l_{\beta_{tk}}^{(r)}\exp(\beta_{tk}^{(0)})$$

or

$$E(\beta_{tk}^{(r)}) = \log(l_{\beta_{tk}}^{(r)}) - \text{var}(\beta_{tk}^{(r)})/2 + \beta_{tk}^{(0)}.$$

Alternatively, taking expectations on both sides of equation (6), after using Taylor series expansion for $\log(\tilde{\lambda}_{\beta_{tk}}^{(r)})$, we obtain

$$E(\beta_{tk}^{(r)}) \approx \log(l_{\beta_{tk}}^{(r)}) - c^2/2 + \beta_{tk}^{(0)} \tag{7}$$

from which $\text{var}(\beta_{tk}^{(r)}) \approx c^2$.

For $c = 0$, the model proposed is equivalent to a deterministic constraint. The MAR model under this pattern–mixture framework is a special case with $\tilde{\lambda}_{\beta_{tk}}^{(r)} \equiv 1$ ($l = 1; c = 0$) for all $t \geqslant 2$, $r < t$ and k. Indeed, the Poisson distribution is uniquely defined by its mean structure, so $\tilde{\lambda}_{\beta_{tk}}^{(r)} \equiv 1$ for $t \geqslant 2$, $r < t$, and k is equivalent to $f(y_t|\mathbf{y}_{t-1}, x, r = j) = f(y_t|\mathbf{y}_{t-1}, x, r \geqslant t)$, $\forall t \geqslant 2, \forall j < t$. The latter are the available case missing value restrictions that were defined by Molenberghs *et al.*

(1998), which are equivalent to MAR. Thus, $\tilde{\lambda}_t^{(r)}(u_t)$ can be seen as an ignorability index that is equivalent to 1 if the missing data mechanism is ignorable and different from 1 when the missing data mechanism is non-ignorable.

4. Computations: Gibbs sampling

The posterior distribution of the parameters in model (1) is analytically intractable; hence samples from the posterior are obtained by using Gibbs sampling (Gelfand and Smith, 1990).

4.1. Complete-data inference

With no missing data, draws from $\Pr(\beta, b, \Sigma|y, x)$ are generated via Gibbs sampling based on the following conditional distributions:

(a) $[\beta, b, \Sigma|y, x]$,
    (i) $[\beta|b, \Sigma, y, x]$,
    (ii) $[b|\beta, \Sigma, y, x]$,
    (iii) $[\Sigma|\beta, b, y, x]$.

Distributions (i) and (ii) do not have a closed form and draws from them are based on the Metropolis algorithm (Metropolis *et al.*, 1953) or the Metropolis–Hastings algorithm (Hastings, 1970); the draws for distribution (iii) are obtained from an inverse Wishart distribution.

4.2. Inference under ignorable missing data

Under ignorable missing data mechanisms for y, inferences can be made by drawing values from $\Pr(\beta, b, \Sigma|y_{\text{obs}}, x)$, which is equivalent to drawing from $\Pr(\beta, \Sigma, b, y_{\text{mis}}|y_{\text{obs}}, x)$ fixing $l = 1$ and $c = 0$. These draws are obtained by using Gibbs sampling based on the 'data augmentation' algorithm (Tanner and Wong, 1987) as applied in the following conditional distributions:

(a) $[\beta, b, \Sigma|y, x]$,
    (i) $[\beta|b, \Sigma, y, x]$,
    (ii) $[b|\beta, \Sigma, y, x]$,
    (iii) $[\Sigma|\beta, b, y, x]$;
(b) $[y_{\text{mis}}|\beta, b, \Sigma, y_{\text{obs}}, x]$.

The two blocks (a) and (b) represent an 'outer' Gibbs sampling from which draws from $\Pr(\beta, \Sigma, b, y_{\text{mis}}|y_{\text{obs}}, x)$ are obtained. The first block represents the posterior distribution of the parameters from the outcome model, and the second the posterior distribution of the missing values.

4.3. Inference under non-ignorable missing data

Under non-ignorable missing data mechanisms for y, the Gibbs sampling that was just described is modified to suit pattern–mixture models. The posterior distribution $\Pr(\beta, \Sigma, b, y_{\text{mis}}|y_{\text{obs}}, x, l, c)$ is identified by introducing informative prior distributions on $p(\beta_t^{(r)}|\beta_t^{(0)})$ for $r = 1, 2, \ldots, t - 1$ and $t \geqslant 2$. The draws from the posterior are obtained by using Gibbs sampling in the following conditional distributions:

(a) $[\beta, b, \Sigma|y, x, l, c]$,
    (i) $[\beta|b, \Sigma, y, x, l, c]$,

> (ii) $[b|\beta, \Sigma, y, x, l, c]$,
> (iii) $[\Sigma|\beta, b, y, x, l, c]$;
> (b) $[y_{\text{mis}}|\beta, b, \Sigma, y_{\text{obs}}, x, l, c]$.

For a given prior distribution on $\tilde{\lambda}_{\beta_{tk}}^{(r)}$, the prior distribution $p(\beta_{tk}^{(r)}|\beta_{tk}^{(0)})$ is fully determined (described in Section 3.2) and, hence, so are the corresponding posterior distributions. A log-normal distribution for $\tilde{\lambda}_{\beta_{tk}}^{(r)}$ is used with mean $l_{\beta_{tk}}^{(r)}$ and variance $c^2 l_{\beta_{tk}}^{2(r)}$. Step (i) is then modified to incorporate this important prior distribution by conditioning on l, the vector of all $l_{\beta_{tk}}$, and c. Values of $l_{\beta_{tk}}^{(r)}$ and c are varied to explore the sensitivity of the conclusions across different $l_{\beta_{tk}}^{(r)}$ and c. WinBUGS software is used to implement the draws and to derive inferences on parameters of interest (Spiegelhalter *et al.*, 2003; Gilks *et al.*, 1996).

Pattern–mixture models and selection models are two paths leading to the same joint distribution. Ekholm and Skinner (1998) used a pattern–mixture model and a selection interpretation for their sensitivity analysis. Similarly here, we shall relate the pattern–mixture model to a selection model, which will offer some reassurance that the assumed pattern–mixture models provide sensible implications about the selection mechanism. This can be easily derived in WinBUGS by adding a logistic regression that relates the missing data indicator to the full data (observed and imputed in step (b)). We implement this part in WinBUGS by using the cut(y) function as a valve to control the flow of the information from the pattern–mixture model to the selection model, but not vice versa.

5. Application to the asthma study

We now apply the model that was described in Section 3 to evaluate the effect of the asthma intervention programme on reducing the number of asthma-related hospitalizations over time, as well as the sensitivity of the results to different missing data mechanisms.

5.1. Inferences under an ignorable missing data mechanism

We fit the following transition Markov model of first order for $t = 2, 3$:

$$\log\{E(y_{ijt}|I_i, y_{ijt-1}, b_i)\} = \beta_{t0}(1 - I_i) + \beta_{t1}I_i + \beta_{t2}\log\{(\eta_{ijt-1} + 1)/2\}$$
$$+ \beta_{t3}I_i\log\{(\eta_{ijt-1} + 1)/2\} + b_{it} + o_{ijt}. \tag{8}$$

η_{ijt-1} is the yearly hospitalization rate at time $t - 1$; I_i is the intervention indicator, and $o_{ijt} = \log(n_{ijt})$ is the offset variable for the number of months n_{ijt} over which the hospitalizations are counted. We use $\log\{(\eta_{ijt-1} + 1)/2\}$ as a predictor because it fits the model better and is defined when $\eta_{ijt-1} = 0$. In equation (8), β_{t0} and β_{t1} estimate the adjusted event rate at time t, for both the control and the intervention group. Thus, testing the average intervention effect over time corresponds to estimating $\beta_{11} - \beta_{10}$ and $\beta_{21} - \beta_{20}$, which represent the treatment effect on hospitalizations during the first and second follow-up period respectively when $\eta_{t-1} = 1$.

Inferences under MAR are derived by using a pattern–mixture model (Section 3.2) and assuming that all $\tilde{\lambda}^{(r)}$s are 1. The missing values of the offset variable are set at $o_{ijt} = \log(12 \text{ months}) = 2.48$. Model (8) is fitted by using a Bayesian approach implemented through WinBUGS software. Parameter inferences are derived on the basis of five chains of 20000 iterations, each with different random starting points and following a burn-in of 5000 iterations. The convergence of iterations for each parameter is monitored by using the Gelman and Rubin (1992) univariate scale reduction factor (SRF), with all being less than 1.01. The overall convergence is monitored by using the Brooks and Gelman (1998) multivariate SRF, which equals 1.05,

Table 1. Bayesian estimates of parameters under an MAR mechanism†

Parameter	Posterior mean (posterior standard deviation)
β_{20} (C)	-4.43 (0.31)
β_{30} (C)	-3.60 (0.39)
β_{21} (I)	-4.46 (0.28)
β_{31} (I)	-5.27 (0.70)
$\beta_{21} - \beta_{20}$ (I *versus* C)	-0.03 (0.40)
$\beta_{32} - \beta_{30}$ (I *versus* C)	-1.67 (0.76)
$\beta_{22}:\log(\eta_0)$	2.31 (0.47)
$\beta_{32}:\log(\eta_1)$	3.05 (0.57)
$\beta_{23}: I \times \log(\eta_0)$	-0.10 (0.62)
$\beta_{33}: I \times \log(\eta_1)$	-1.68 (1.15)
σ_1^2	0.38 (0.28)
σ_2^2	0.60 (0.66)
σ_{12}	-0.14 (0.28)

†C stands for control group and I stands for intervention group; Monte Carlo simulation errors range 0.002–0.015.

indicating that no gain will result if the iterations continue. The results under an MAR mechanism are shown in Table 1. For hospitalizations, no intervention effect occurs during the first follow-up period ($\beta_{11} - \beta_{01} = -0.03$; $p = 0.94$), but a significant effect occurs at second follow-up ($\beta_{12} - \beta_{02} = -1.67$; $p = 0.03$). When the intervention group is compared with the control group, the rate of hospitalization is reduced by 71% (95% CI $= (27–96\%)$) between the first and second follow-up. The interaction between intervention and previous hospitalization is non-significant at either time point. The variances of the random intercepts are comparable with their standard errors, which reflect a weak clustering within physician once conditioning on the previous response. As expected, inferences that are derived by using a pattern–mixture model that assumes that all $\tilde{\lambda}^{(r)}$s are 1 are the same as inferences that are derived on the basis of only observed data.

5.2. Inferences under a non-ignorable missing data mechanism

For the pattern–mixture model, different parameters for each missing data pattern are used in model (8) for the control group $\beta_{t0}^{(r)}$ and the intervention group $\beta_{t1}^{(r)}$, at both follow-ups. The effect of the other covariates is held the same across the missing data patterns. Prior distributions on the parameters $\tilde{\lambda}_{\beta_{t0}}^{(r)}$ and $\tilde{\lambda}_{\beta_{t1}}^{(r)}$ are log-normal with means $l_{\beta_{t0}}^{(r)}$ and $l_{\beta_{t1}}^{(r)}$ and coefficient of variation c. The overall β-parameters are derived by averaging across the missing data patterns (Little, 1995), i.e. $\beta_t = \Sigma_r \pi_{tr} \beta_t^{(r)}$, where $\pi_{tr} = \Pr(R_{it} = r)$ is the proportion of subjects in missing data pattern r at time t. The π_{tr} is estimated by m_{tr}/n where m_{tr} is the number of subjects in pattern r at time t and n is the total number of subjects. The uncertainty in π_{tr} is negligible and is ignored.

The missing data mechanism is determined primarily by mean parameters $l_{\beta_{t0}}^{(r)}$ and $l_{\beta_{t1}}^{(r)}$, with different values of $l_{\beta_{t0}}^{(r)}$ and $l_{\beta_{t1}}^{(r)}$ representing different dropout mechanisms. Several values of $l_{\beta_{t0}}^{(r)}$ and $l_{\beta_{t1}}^{(r)}$ are used for sensitivity analyses. Because the treatment effect at $t = 2$ is highly non-significant we focus the sensitivity analyses on the missing data at $t = 3$. We assume that the rates of hospitalizations between the patients in the missing data pattern and those in the observed data pattern at $t = 2$ are the same on average, i.e. $l_{\beta_{20}}^{(2)} = l_{\beta_{21}}^{(2)} = 1$. We also let $l_{\beta_{30}}^{(2)} = l_{\beta_{30}}^{(3)} = l_{\beta_{30}}$ and

$l_{\beta_{31}}^{(2)} = l_{\beta_{31}}^{(3)} = l_{\beta_{31}}$. With this assumption the sensitivity analyses are based on different values of $l_{\beta_{30}}$ and $l_{\beta_{31}}$, which represent the relationship between the missing values across all the missing data patterns at $t = 3$ ($r < 3$) and the observed values at time $t = 3$ ($r \geqslant 3$). We consider several values for $l_{\beta_{30}}$ and $l_{\beta_{31}}$ including a combination that would make the treatment effect at $t = 3$ non-significant. Initially we set $c = 0.1$, which corresponds to a 95% CI for $\tilde{\lambda}_{\beta_t}$ neither too narrow nor too wide. For example, for $\tilde{\lambda}_{\beta_{t1}} = 0.5$ the 95% CI is (0.41, 0.61).

Our primary parameter of interest is $\beta_{t1} - \beta_{t0}$, which estimates the average intervention effect. Inferences on this parameter are affected by the ratio $\tilde{\eta}_{\beta_{t1} - \beta_{t0}} = \tilde{\lambda}_{\beta_{t1}}/\tilde{\lambda}_{\beta_{t0}}$, which is the ratio of the relative risks (intervention *versus* control) between patients in the missing data pattern r and patients in the observed data pattern, i.e.

$$\tilde{\eta}_{\beta_{t1} - \beta_{t0}}^{(r)} = \frac{E(\mu_t^{(r)} | I = 1, Y_t, b_t)/E(\mu_t^{(r)} | I = 0, Y_t, b_t)}{E(\mu_t^{(0)} | I = 1, Y_t, b_t)/E(\mu_t^{(0)} | I = 0, Y_t, b_t)} = \frac{\tilde{\lambda}_{\beta_{t1}}}{\tilde{\lambda}_{\beta_{t0}}}.$$

This follows a log-normal distribution with mean $\log(l_{\beta_{t1}}/l_{\beta_{t0}})$ and variance $2c^2$. With this parameterization the inferences on the treatment effect would be the same across different specifications of $l_{\beta_{t0}}$ and $l_{\beta_{t1}}$, as long as the distribution of their ratio is the same. Without loss of generality, we set $l_{\beta_{30}} = 1$ and we vary $l_{\beta_{31}}$ to perform sensitivity analyses on the $(\beta_{31} - \beta_{30})$-parameter for a range of $l_{\beta_{31}}/l_{\beta_{30}}$.

Choosing $l_{\beta_{31}}/l_{\beta_{30}} = 1.76$ yields an estimate for $\beta_{31} - \beta_{30}$ that is borderline significant (95% CI (0.05–1.00)). Several other combinations of $l_{\beta_{30}}$ and $l_{\beta_{31}}$, including $l_{\beta_{31}}/l_{\beta_{30}} = 1$, are used for sensitivity analyses. In addition, for each pattern–mixture model, a logistic selection model is fitted to relate the missing data indicator (at $t = 3$) to intervention group variable, current and previous ys, and their interaction:

$$\log\left\{\frac{\Pr(R_i = 3)}{1 - \Pr(R_i = 3)}\right\} = \gamma_0 + \gamma_1 I_i + \gamma_2 y_{i2} + \gamma_3 I_i \times y_{i2} + \gamma_4 (1 - I_i) \times y_{i3} + \gamma_5 I_i \times y_{i3} + g_i.$$

The random effect for the physician is g_i, normally distributed with mean 0 and variance σ_g^2. Inferences for each model are derived from five chains with different random starting points of 20000 iterations each following a burn-in of 5000 iterations. The convergence of iterations is monitored by using the SRF; all univariate SRFs are less than 1.02 and all multivariate SRFs are less than 1.09, indicating that the iterations converged. The results for both pattern–mixture model and logistic selection model are shown in Table 2. Dependence of the missing data indicator R_3 on y_3 in the intervention group (γ_5) becomes stronger as the $l_{\beta_{31}}$ is further from 1. The parameter estimates that are derived for the logistic selection model are consistent with assumptions in the pattern–mixture model. Thus, in the pattern–mixture model with $l_{\beta_{30}} = 1$ and $l_{\beta_{31}} = 4$, the selection model yields a positive estimate of $\gamma_5 = 1.26$. Both models indicate that, in the intervention group, subjects with a higher number of hospitalizations are more likely to drop out. For the control group, no difference is assumed on the rate of hospitalizations at $t = 3$ between subjects who are missing and subjects who are observed. This is consistent with the close-to-zero estimate of $\gamma_4 = 0.01$.

Next, other values of c are used for sensitivity analyses. These evaluate how the boundary ratio $l_{\beta_{31}}/l_{\beta_{30}}$, for which the intervention effect at second follow-up is just significant, relates to c. A graphical display of this relationship is shown in Fig. 1. As the coefficient of variation increases, the boundary ratio decreases. This is mainly related to having larger CIs on the estimate of $\beta_{31} - \beta_{30}$ as the prior on $\tilde{\lambda}$ is less informative (high c). We fixed $c = 0.1$ for sensitivity analyses that are shown in Table 2, but other choices of c can be used if deemed appropriate.

Finally, the sensitivity analyses show that the parameter estimates for intervention effect at

Table 2. Sensitivity analyses ($c = 0.1$)†

Parameter	Posterior mean and standard deviation for the following $l_{\beta_{31}}/l_{\beta_{30}}$ ratios:			
	1/1	*1.76/1*	*2.5/1*	*4/1*
Pattern–mixture model				
β_{20} (C)	−4.44 (0.31)	−4.43 (0.31)	−4.44 (0.31)	−4.44 (0.32)
β_{30} (C)	−3.60 (0.39)	−3.58 (0.38)	−3.60 (0.40)	−3.60 (0.40)
β_{21} (I)	−4.46 (0.30)	−4.46 (0.29)	−4.46 (0.29)	−4.46 (0.29)
β_{31} (I)	−5.26 (0.69)	−4.97 (0.70)	−4.79 (0.70)	−4.55 (0.71)
$\beta_{21} - \beta_{30}$ (I *versus* C)	−0.03 (0.40)	−0.03 (0.40)	−0.02 (0.40)	−0.03 (0.40)
$\beta_{31} - \beta_{30}$ (I *versus* C)	−1.66 (0.76)	−1.38 (0.76)	−1.19 (0.77)	−0.95 (0.77)
β_{22}: $\log(\eta_0)$	2.31 (0.47)	2.31 (0.47)	2.31 (0.47)	2.30 (0.47)
β_{32}: $\log(\eta_1)$	3.05 (0.57)	3.05 (0.57)	3.06 (0.57)	3.04 (0.58)
β_{23}: $I \times \log(\eta_0)$	−0.10 (0.62)	−0.10 (0.61)	−0.11 (0.62)	−0.10 (0.62)
β_{33}: $I \times \log(\eta_1)$	−1.67 (1.15)	−1.69 (1.15)	−1.68 (1.15)	−1.68 (1.16)
σ_1^2	0.39 (0.29)	0.39 (0.28)	0.39 (0.27)	0.39 (0.28)
σ_2^2	0.59 (0.62)	0.56 (0.60)	0.61 (0.68)	0.60 (0.66)
σ_{12}	−0.15 (0.30)	−0.15 (0.29)	−0.16 (0.29)	−0.15 (0.30)
Logistic selection model				
γ_0: intercept	0.26 (0.18)	0.25 (0.17)	0.25 (0.17)	0.25 (0.18)
γ_1: I	−0.22 (0.24)	−0.23 (0.24)	−0.27 (0.24)	−0.30 (0.24)
γ_2: y_2	0.05 (0.48)	0.04 (0.49)	0.01 (0.47)	0.04 (0.49)
γ_3: $I \times y_2$	−0.05 (0.56)	−0.08 (0.58)	−0.09 (0.57)	−0.19 (0.60)
γ_4: $(1 - I) \times y_3$	−0.01 (0.46)	−0.01 (0.46)	0.01 (0.46)	0.01 (0.47)
γ_5: $I \times y_3$	−0.23 (1.01)	0.41 (0.83)	0.61 (0.68)	1.26 (0.67)
σ_g^2	0.60 (0.15)	0.59 (0.15)	0.59 (0.15)	0.60 (0.15)

†Monte Carlo simulation errors range 0.002–0.01.
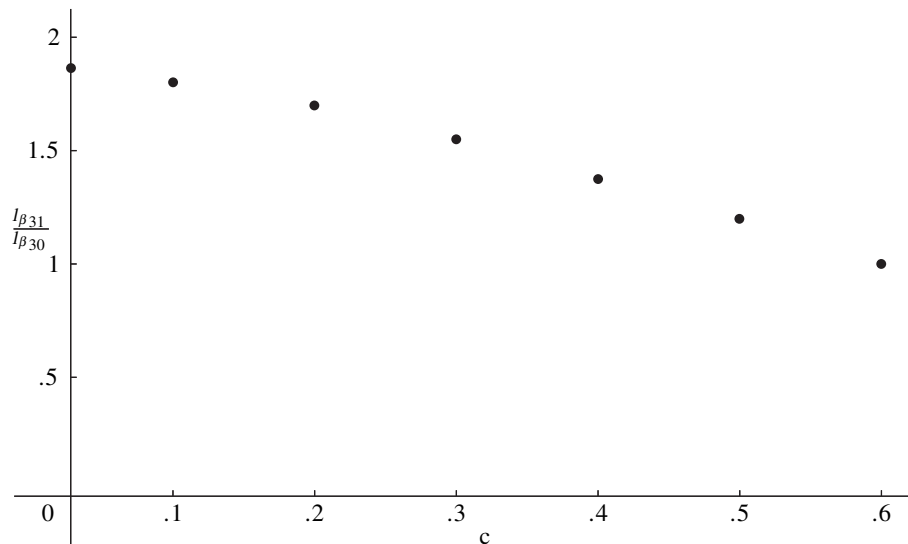
Fig. 1. Illustration of how the value of the ratio $l_{\beta_{31}}/l_{\beta_{30}}$ at which the estimated intervention effect is just significant declines as the coefficient of variation c increases

second follow-up differ across missing data mechanisms. However, the evidence of an intervention effect at second follow-up is robust to a range of missing data mechanisms. It becomes non-significant only when the relative risk of hospitalization for intervention *versus* control at $t = 3$ is on average 1.76 or higher among the missing subjects compared with the observed subjects, assuming that $c = 0.1$. No variation appeared on the other parameters. This is consistent with the fact that in all analyses we assumed that these parameters did not vary by missingness patterns. As a result, the overall estimates will be the same as the estimates by using an MAR mechanism.

6. Conclusions

In this paper we have developed a Bayesian model to fit clustered longitudinal data from Poisson outcomes with potential non-ignorably missing observations. The model for non-ignorably missing values was developed by using a pattern–mixture model that was identified on the basis of easy-to-understand assumptions about missing data. These assumptions used prior distributions on an ignorability index parameter $\tilde{\lambda}^{(r)}_{\beta_{tk}}$, which represents the ratio of event rates between the missing data pattern r and the observed data pattern (conditioned on other covariates and previous responses), for group k at time t. The distributions of $\tilde{\lambda}^{(r)}_{\beta_{tk}}$ are unknown and cannot be derived from the data. However, it is possible for an investigator to give a range for each $\tilde{\lambda}^{(r)}_{\beta_{tk}}$, and then to explore the sensitivity of statistical inferences over that range. Such a parameterization is intuitive and easy to understand. It contains the MAR mechanism as a special case where all $\tilde{\lambda}^{(r)}_{\beta_{tk}}$ are 1. The method was implemented by using WinBUGS1.4 software and applied to the asthma intervention study for evaluating the effect of the interactive seminar on the hospitalization outcome. WinBUGS gives added strength because it can relate the pattern–mixture model and its assumptions to a logistic selection model. Such flexibility provides additional assurance and validity to the sensitivity analyses.

In the asthma study following per-protocol analysis, under an ignorable missing data mechanism, intervention did not show a significant effect in reducing the overall rate of hospitalizations during the first follow-up period. During the second follow-up period the overall rate of hospitalizations in the intervention group compared with the control group was reduced by 71% (95% CI (27–96%)). The intervention showed similar beneficial effects across patients with different rates of hospitalizations in either period. The sensitivity analyses showed that the parameter estimates for the intervention group variable differ across missing data mechanisms. However, the evidence of an intervention effect at second follow-up was fairly robust to possible departures from an ignorable missing data mechanism.

Finally, although the model that was used answered specific questions for the asthma study, it can also be applied more generally, e.g. to impute non-ignorably missing values on count outcomes. Then, once the missing data have been imputed, additional analyses can be performed on the complete-data set by using standard statistical techniques.

Acknowledgements

References

Albert, P. S. and Follmann, D. A. (2000) Modeling repeated count data subject to informative dropout. *Biometrics*, **56**, 667–677.

Brooks, S. P. and Gelman, A. (1998) Alternative methods for monitoring convergence of iterative simulations. *J. Computnl Graph. Statist.*, **17**, 434–455.

Clark, N. M., Gong, M., Schork, M. A., Evans, D., Roloff, D., Hurwitz, M., Maiman, L. A. and Mellins, R. B. (1998) Impact of education for physicians on patients outcomes. *Pediatrics*, **101**, 831–836.

Clark, N. M., Gong, M., Schork, M. A., Kaciroti, N., Evans, D., Roloff, D., Hurwitz, M., Maiman, L. A. and Mellins, R. B. (2000) Long-term effects of asthma education for physicians on patient satisfaction and use of health services. *Eur. Resp. J.*, **16**, 15–21.

Daniels, M. and Hogan, J. (2000) Reparameterizing the pattern mixture model for sensitivity analysis under informative dropout. *Biometrics*, **56**, 1241–1248.

Demirtas, H. (2005) Multiple imputation under Bayesianly smoothed pattern-mixture models for non-ignorable drop-out. *Statist. Med.*, **24**, 2345–2363.

Demirtas, H. and Schafer, J. L. (2003) On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statist. Med.*, **22**, 2553–2575.

Diggle, P. J., Heagerty, P., Liang, K. Y. and Zeger, S. (2002) *Analysis of Longitudinal Data*, 2nd edn. New York: Oxford University Press.

Ekholm, H. and Skinner, C. (1998) The Muscatine children's obesity data reanalysed using pattern mixture models. *Appl. Statist.*, **47**, 251–263.

Gelfand, A. E. and Smith, A. F. M. (1990) Sampling based approaches to calculate marginal densities. *J. Am. Statist. Ass.*, **85**, 398–409.

Gelman, A. and Rubin, D. B. (1992) Inference from iterative simulations using multiple sequences. *Statist. Sci.*, **7**, 457–472.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds) (1996) *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Guo, W., Ratcliffe, S. J. and Ten Have, T. T. (2004) A random pattern-mixture model for longitudinal data with dropouts. *J. Am. Statist. Ass.*, **99**, 929–937.

Hastings, W. K. (1970) Monte Carlo sampling method using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Heckman, J. J. (1979) Sample selection bias as a specification error. *Econometrica*, **47**, 153–162.

Hogan, J. W. and Laird, N. M. (1997) Mixture models for joint distribution of repeated measures and event times. *Statist. Med.*, **16**, 239–257.

Kaciroti, N. (2002) Modeling nonignorable missing data for clustered longitudinal discrete outcomes: a Bayesian approach. *PhD Thesis*. Department of Biostatistics, University of Michigan, Ann Arbor.

Kaciroti, N., Raghunathan, T. E., Schork, M. A., Clark, N. M. and Gong, M. (2006) A Bayesian approach for clustered longitudinal ordinal outcome with nonignorable missing data: evaluation of an asthma education program. *J. Am. Statist. Ass.*, **101**, 435–446.

Kenward, M. G. (1998) Selection models for repeated measures with non-random dropout: an illustration of sensitivity. *Statist. Med.*, **17**, 2723–2732.

Kenward, M. G. and Molenberghs, G. (1999) Parametric models for incomplete continuous and categorical longitudinal data. *Statist. Meth. Med. Res.*, **8**, 51–83.

Kenward, M. G., Molenberghs, G. and Thijs, H. (2003) Pattern-mixture models with proper time dependence. *Biometrika*, **90**, 53–71.

Little, R. J. A. (1993) Pattern-mixture models for multivariate incomplete data. *J. Am. Statist. Ass.*, **88**, 125–134.

Little, R. J. A. (1994) A class of pattern-mixture models for normal missing data. *Biometrika*, **81**, 471–483.

Little, R. J. A. (1995) Modeling the dropout mechanism in repeated measures studies. *J. Am. Statist. Ass.*, **90**, 1113–1121.

Little, R. J. A. and Rubin, D. B. (2002) *Statistical Analysis with Missing Data*, 2nd edn. New York: Wiley.

Little, R. J. A. and Wang, Y. (1996) Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, **52**, 98–111.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. and Teller, A. H. (1953) Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1091.

Molenberghs, G., Michiels, B., Kenward, M. G. and Diggle, P. J. (1998) Monotone missing data and pattern-mixture models. *Statist. Neerland.*, **52**, 153–161.

National Center for Health Statistics (2002) *National Health Interview Survey, 2001-2003*. National Center for Health Statistics, Atlanta.

Ridout, M. S. (1991) Testing for random dropouts in repeated measurement data. *Biometrics*, **47**, 1617–1621.

Spiegelhalter, D. J., Thomas, A., Best, N. G. and Lunn, D. (2003) *WinBUGS User Manual: Version 1.4*. Cambridge: Medical Research Council Biostatistics Unit.

Tanner, M. and Wong, W. H. (1987) The calculation of posterior distribution by data augmentation (with discussion). *J. Am. Statist. Ass.*, **82**, 528–550.

Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G. and Curran, D. (2002) Strategies to fit pattern-mixture models. *Biostatistics*, **3**, 245–264.

Wu, M. and Carroll, R. (1988) Estimation and comparison of change in presence of informative right censoring by modeling the censoring process. *Biometrics*, **44**, 175–188.

Zeger, S. L. and Qaqish, B. (1988) Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, **44**, 1019–1031.