A Bayesian model for time-to-event data with informative censoring

NIKO A. KACIROTI*, TRIVELLORE E. RAGHUNATHAN, JEREMY M. G. TAYLOR

Department of Biostatistics, Center for Human Growth and Development, University of Michigan, Ann Arbor, MI 48109, USA

nicola@umich.edu

STEVO JULIUS

Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA

SUMMARY

Randomized trials with dropouts or censored data and discrete time-to-event type outcomes are frequently analyzed using the Kaplan–Meier or product limit (PL) estimation method. However, the PL method assumes that the censoring mechanism is noninformative and when this assumption is violated, the inferences may not be valid. We propose an expanded PL method using a Bayesian framework to incorporate informative censoring mechanism and perform sensitivity analysis on estimates of the cumulative incidence curves. The expanded method uses a model, which can be viewed as a pattern mixture model, where odds for having an event during the follow-up interval $(t_{k-1}, t_k]$, conditional on being at risk at t_{k-1}, differ across the patterns of missing data. The sensitivity parameters relate the odds of an event, between subjects from a missing-data pattern with the observed subjects for each interval. The large number of the sensitivity parameters is reduced by considering them as random and assumed to follow a log-normal distribution with prespecified mean and variance. Then we vary the mean and variance to explore sensitivity of inferences. The missing at random (MAR) mechanism is a special case of the expanded model, thus allowing exploration of the sensitivity to inferences as departures from the inferences under the MAR assumption. The proposed approach is applied to data from the TRial Of Preventing HYpertension.

Keywords: Clinical trials; Hypertension; Ignorability index; Missing data; Pattern-mixture model; TROPHY trial.

1. INTRODUCTION

Randomized control trials where the primary outcome is time-to-event are subject to censoring. Censoring can occur due either to dropout or violations of the study protocol. The resulting censoring mechanism could be considered noninformative if the time of censoring is independent of the time of event and

---

*To whom correspondence should be addressed.

informative otherwise. To analyze these data, where the censoring mechanism is potentially informative, additional data or assumptions are required to make valid inferences (Zheng and Klein, 1995).

There are 3 general approaches for analyzing data with dropouts or censoring. Rotnitzky *and others* (2001), Scharfstein *and others* (2001), and Scharfstein and Robins (2002) adopt a selection model approach where an explicit censoring mechanism is formulated and related to the unobserved outcome. On the other hand, Little (1995), Little and Rubin (2002), and Birmingham *and others* (2003) used a pattern-mixture (PM) approach where the distribution of the outcome across different patterns formed by censoring times are related by some identifiability constraints. Wu and Carroll (1988) used a shared-parameter model. This model uses common random effects to relate the response to the missing-data indicator (Guo *and others*, 2004). All 3 approaches involve untestable assumptions about the censoring mechanism, therefore must be followed with sensitivity analysis (Molenberghs and Kenward, 2007).

In a PM setting, Daniels and Hogan (2000) and Kenward *and others* (2003) identified parameters using constraints. Roy and Daniels (2008) use latent class PM models with Bayesian model averaging. Previously, we developed a model for binary outcomes with missing data (Kaciroti *and others*, 2009) and used Bayesian estimation methods for sensitivity analysis. Here, we propose a Bayesian model using a parameterization within the PM approach to analyze time-to-event data with informative censoring.

The product limit (PL) estimator is a general purpose nonparametric approach used to analyze data from clinical trials with time-to-event outcomes and noninformative censoring. Using a Bayesian methodology, Susarla and Van Ryzin (1976) propose a Dirichlet process and Hjort (1990) propose a Beta process for analyzing time-to-event data with noninformative censoring. Phadia and Susarla (1983) and Tsai (1986) extend the method of Susarla and Van Ryzin (1976) by introducing a bivariate exponential distribution on the parameters of the Dirichlet process to allow the possibility of dependent censoring. However, choosing a distribution that captures the correct dependence between censoring and event times is not easy to specify as it is not estimable from the observed data. Ibrahim *and others* (2001) give a comprehensive review on Bayesian survival analysis. More recently, Scharfstein *and others* (2003) propose a fully Bayesian methodology that accounts for informative censoring by introducing prior restrictions on the selection mechanism and follow with sensitivity analysis. Alternatively, we propose here a fully Bayesian methodology within the PM setting, using a nonparametric approach for the censoring mechanism. Our goal is to use the appealing methodology of PL estimator but expand it to incorporate informative censoring. We first formulate the PL estimator using a Bayesian framework. When the censoring mechanism is informative, we introduce intuitive parameters that are used for sensitivity analysis. These parameters capture the difference between the censored and the observed data.

To reduce the large number of the sensitivity parameters, we consider them random from a prior distribution. The prior distribution links all these parameters together around an average missing-data mechanism (mean parameter) while simultaneously accommodating possible differences among them by introducing random variation. The parameterization and the prior distribution associated with it are easy-to-use and accommodate modeling of different types of missing data, including those with informative censoring or missing not at random (MNAR). It also contains MAR as a special case enabling a local sensitivity analysis for departures in a neighborhood of MAR (Ma *and others*, 2005). We incorporate a probabilistic range on the sensitivity parameters by introducing a prior distribution on them and then applying Bayesian strategies to derive inferences (Kaciroti *and others*, 2006, 2008, 2009; Daniels and Hogan, 2007).

In Section 2, we develop the Bayesian approach for PL estimate with no censoring and expand it to include informative censoring in Section 3 . In Section 4, we describe the Trial Of Preventing Hypertension (TROPHY) study, which motivated and provided the context for the methods developed here. In Section 5, we apply the proposed models to the TROPHY data and in Section 6, we give results from comparative studies and offer conclusions in Section 7.

2. BAYESIAN FORMULATION OF PL ESTIMATOR

First, we begin with the formulation of a version of a PL method using a Bayesian framework without censoring. Let $Y = (Y_0, Y_1, \ldots, Y_K)'$ be a vector of binary responses, indicating the presence of the event at baseline (Y_0) and K follow-up visits $(Y_1, \ldots, Y_K)'$. Let t_k denote the time from baseline for the follow-up visit $k = 1, 2, \ldots, K$, and $t_0 = 0$. Let T be the measure of time when the event occurred, $T \in \{t_1, t_2, \ldots, t_K\}$; with $T = t_k$ if and only if $Y_{k-1} = 0$ and $Y_k = 1$, and $T > t_K$ if no event occurred by t_K. The focus here is to compare the cumulative incidence, $I(t) = \Pr(T \leqslant t)$, between groups.

Let p_{gk} be the probability of an event in $(t_{k-1}, t_k]$ for a subject who is at risk at t_{k-1} in the treatment group g. The cumulative incidence function up to time t_k is

$$I_g(t_k) = 1 - S_g(t_k) = 1 - \prod_{j=1}^{k} (1 - p_{gj}), \tag{2.1}$$

where $S_g(t_k) = \Pr(T > t_k | Trt = g)$.

Let N_{gk} be the number of individuals at risk at time t_{k-1}; and e_{gk} be the number of events occurring in $(t_{k-1}, t_k]$ among these subjects for $g = 0, 1$ and $k = 1, 2, \ldots, K$. In absence of censoring, $N_{gk} = N_g - \sum_{j=0}^{k-1} e_{gj}$, where N_g is the number of subjects at the beginning of the study and $e_{g0} = 0$. The likelihood function of the parameter $p_g = (p_{g1}, \ldots, p_{gK})$, conditional on $e_g = (e_{g1}, \ldots, e_{gK})$, N_g, and treatment indicator is

$$\mathcal{L}(p_g | e_g, N_g, Trt) = f(e_{g1} | N_g, p_{g1}, Trt) \prod_{k=2}^{K} f(e_{gk} | e_{g1}, \ldots, e_{gk-1}, N_g, p_{gk}, Trt),$$

where $e_{g1} | N_g, p_{g1} \sim \text{Bin}(N_g, p_{g1})$ and for $k = 2, 3, \ldots, K$,

$$e_{gk} | e_{g1}, \ldots, e_{gk-1}, N_g, p_{gk} \sim \text{Bin}(N_{gk} p_{gk}) = \text{Bin}\left(N_g - \sum_{j=1}^{k-1} e_{gj}, p_{gk}\right).$$

Framing the inference problem from a Bayesian perspective, the posterior distribution of p_{gk} assuming a noninformative prior Beta(0.001, 0.001) is

$$p_{gk} | e_{g1}, \ldots, e_{gk}, N_g \sim \text{Beta}\left(e_{gk} + 0.001, N_g - \sum_{j=1}^{k} e_{gj} + 0.001\right). \tag{2.2}$$

The inference on the cumulative incidence $I_g = (I_g(t_1), \ldots, I_g(t_K))$ is derived from (2.1) and using Monte Carlo (MC) simulations, with the posterior mean of $I_g(t_k)$ estimated by

$$\hat{I}_g(t_k) = \frac{1}{S} \sum_{s=1}^{S} I_g^s(t_k) = 1 - \frac{1}{S} \sum_{s=1}^{S} \prod_{j=1}^{k} (1 - p_{gj}^s),$$

where p_{gj}^s is the sth draw from (2.2), and $S = 100,000$ simulations. The posterior standard deviation of $I_g(t_k)$ is also estimated using MC simulations.

Our main hypothesis is that I_0 and I_1 (or equivalently S_0 and S_1) are different over the study period and also at the 4-year mark. We compare these curves between 2 groups by using the average distance on a log scale:

$$\mathcal{Z} = \frac{1}{K} \sum_{k=1}^{K} \{\log(S_0(t_k)) - \log(S_1(t_k))\} = \sum_{j=1}^{K} \frac{K - j + 1}{K} \log\left(\frac{1 - p_{0j}}{1 - p_{1j}}\right). \tag{2.3}$$

If $I_0 = I_1$, i.e., no group differences, then $\mathcal{Z}$ should be 0. In a Bayesian approach, we calculate the posterior distribution of $\mathcal{Z}$ and the Bayesian p value using MC simulations. The p value is equal to $2 * \min\{\Pr(\mathcal{Z} > 0), \Pr(\mathcal{Z} < 0)\}$ and is calculated as the smaller of the proportion of $\mathcal{Z} > 0$ or $\mathcal{Z} < 0$ times 2 for a 2-sided test. Similarly, the incidence rates at the end of the study are compared using $\mathcal{D} = I_1(t_K) - I_0(t_K)$.

3. EXPANDED MODEL

When censored data occurred, the estimates, described in Section 2, are modified to account for censoring mechanism. Let R_i be the random variable indexing the missing-data patterns, which is equal to k if subject i was censored in $(t_{k-1}, t_k]$, $k = 1, \ldots, K$. Let $R_i = K+1$ when subject i had no event and completed the study. If the censoring mechanism is nonignorable, then $\Pr(R_i = k)$ would depend on the unobserved Y_{ik}, and additional assumptions or constraints are needed to identify the model.

In a longitudinal setting, Little (1993) used identifying restrictions by setting the density of missing components, conditional on the observed components, to be equal to the corresponding distribution of the subgroup of completers or some other relevant pattern. For monotone missing data, Molenberghs *and others* (1998) extended Little's approach to available case missing value (ACMV) restriction by setting the density of missing components at visit k, conditioned on the previous components, to be equal to the corresponding distribution of the observed data at visit k. That is, $f(y_k | y_0, y_1, \ldots, y_{k-1}, R = r) = f(y_k | y_0, y_1, \ldots, y_{k-1}, R > k)$ for $\forall k > 0, \forall r \leqslant k$. They showed that ACMV restriction is equivalent to MAR. Here, we extend the ACMV approach to identify PM models when missing data are MNAR: the extension also provides a necessary and sufficient condition for MAR, which is used for local sensitivity analysis around MAR. For MNAR, the identifying constraints relate the distribution of the missing data to that of the observed data. Because there are no data to estimate the identifying constraints, it is useful to have intuitive and easy-to-use parameters, which can then be evaluated using a sensitivity analysis. We propose the following parameterization.

Let $p_{gk}^{(0)} = \Pr(Y_k = 1 | Y_{k-1} = 0, Trt = g, R > k)$ be the probability of an event in $(t_{k-1}, t_k]$ for subjects who are at risk and observed in $(t_{k-1}, t_k]$. Let $N_{gk}^{(0)}$ be the number of subjects at risk and observed in $(t_{k-1}, t_k]$, and $e_{gk}^{(0)}$ be the number of events that occurred in $(t_{k-1}, t_k]$ among these subjects. Next, let $p_{gk}^{(r)} = \Pr(Y_k = 1 | Y_{k-1} = 0, Trt = g, R = r)$ be the probability of an event in $(t_{k-1}, t_k]$ for subjects censored in $(t_{r-1}, t_r]$ who would be at risk at t_{k-1} for $0 < r \leqslant k$. Last, let $N_{gk}^{(r)}$ be the number of subjects censored in $(t_{r-1}, t_r]$ who would have been at risk at t_{k-1}, and $e_{gk}^{(r)}$ be the number of events occurring in $(t_{k-1} t_k]$ among these subjects. Then,

$$e_{gr}^{(r)} | c_{gr}, p_{gr}^{(r)} \sim \text{Bin}(N_{gr}^{(r)}, p_{gr}^{(r)}),$$

and for $k > r > 0$,

$$e_{gk}^{(r)} | e_{gr}^{(r)}, e_{gr+1}^{(r)}, \ldots, e_{gk-1}^{(r)}, c_{gr} p_{gk}^{(r)} \sim \text{Bin}(N_{gk}^{(r)}, p_{gk}^{(r)}), \tag{3.1}$$

where $N_{gr}^{(r)} = c_{gr}$ and $N_{gk}^{(r)} = N_{gk-1}^{(r)} - e_{gk-1}^{(r)}$ for $k > r > 0$. There are no data to estimate the parameters of the distribution (3.1). To overcome this problem, we relate the parameters of the distribution for the missing data to the parameters of the distribution for the observed data, using the following parameterization:

$$\frac{p_{gk}^{(r)}}{1 - p_{gk}^{(r)}} = \tilde{\lambda}_{grk} \frac{p_{gk}^{(0)}}{1 - p_{gk}^{(0)}} \tag{3.2}$$

for any $k > 0$, $0 < r \leqslant k$, and g. Here, $\tilde{\lambda}_{grk}$ is the odds ratio of having an event in $(t_{k-1}, t_k]$, being at risk at t_{k-1}, between those subjects in missing pattern r (censored in $(t_{r-1}, t_r]$), and the observed subjects in $(t_{k-1}, t_k]$ for group g. The $\tilde{\lambda}_{grk}$ parameters measure the departure from MAR and cannot be estimated from the data. We assume that $\tilde{\lambda}_{grk}$ has a log-normal prior distribution with mean l_{grk} and variance $c^2 l_{grk}^2$, where c is the coefficient of variation. The l_{grk} and c parameters are used for sensitivity analysis, where $l_{grk} = 1$ and $c = 0$ correspond to MAR, satisfying the ACMV constraint. Note that for time-to-event outcomes (where $Y_k = 1 \Rightarrow Y_{k'} = 1, \forall k' > k$), the ACMV restriction is equivalent to $p_{gk}^{(r)} = p_{gk}^{(0)}$, $\forall k > 0, \forall r \leqslant k$. Indeed, $f(y_k | y_0, y_1, \dots, y_{k-1}) = f(y_k | y_{k-1})$; and because $f(y_k = 1 | y_{k-1} = 1, R = r) = f(y_k = 1 | y_{k-1} = 1, R > k) = 1$, the only requirement for ACMV is $f(y_k = 1 | y_{k-1} = 0, R = r) = f(y_k = 1 | y_{k-1} = 0, R > k)$ or equivalently $p_{gk}^{(r)} = p_{gk}^{(0)}, \forall k > 0, 0 < r \leqslant k$, and g.

In the longitudinal setting, with K follow-up measures, up to $K + 1$ missing-data patterns can occur, resulting in a maximum of $K \times (K + 1)/2$ constraints ($\tilde{\lambda}_{grk}$ indices) for each group required to identify the model. As K increases, the number of identifying parameters become too large to be feasible for sensitivity analysis. Therefore, it is important to reduce the dimension of the space for the sensitivity parameters. For example, Scharfstein *and others* (2001) introduced a parametric function for their sensitivity parameters; Birmingham *and others* (2003) reduced the dimension of their sensitivity parameters by assuming all to be the same.

Here, we reduce the dimension of the sensitivity space by introducing a prior distribution for the identifying parameters $\tilde{\lambda}_{grk}$ with mean, $E(\tilde{\lambda}_{grk}) = l_g$, yet we introduce random variation that allows each $\tilde{\lambda}_{grk}$ to vary within a range for different r and k. Thus, $\tilde{\lambda}_{grk}$ are consider random variables from a log-normal prior distribution $\log(\tilde{\lambda}_{grk}) \sim \mathcal{N}\left(\log(l_g) - \frac{\log(1+c^2)}{2}, \log(1+c^2)\right)$, (this corresponds to $E(\tilde{\lambda}_{grk}) = l_g$ and $\text{Var}(\tilde{\lambda}_{grk}) = c^2 l_g^2$). Here, l_g and c are second-order sensitivity parameters: l_g indicates the average difference from MAR, and c indicates the degree of the random variation (potential dissimilarities) across different patterns and over time. A coefficient of $c = 0$ corresponds to a deterministic constraint, whereas $c > 0$ corresponds to a stochastic constraint that becomes less informative as c increases. As in complete case analysis, we test the null hypothesis of no difference between 2 curves using the $\mathcal{Z}$ distance (2.3).

3.1 *Fitting the model*

We fit our model following a full Bayesian approach, where the quantities of interest are the cumulative incidence curves for each group as if no subject dropped out: $I_g(t_k) = 1 - \sum_{j=1}^{k}(1 - p_{gj})$. Here, p_{gj} is the probability of having an endpoint in $(t_{j-1}, t_j]$ interval for subjects who would have been at risk at t_{j-1}, this includes both observed and censored subjects. Under PM specification $p_{gj} = \sum_{r=0}^{j} \pi_{gj}^{(r)} p_{gj}^{(r)}$, where $\pi_{gj}^{(r)}$ is the weight (proportion) of the dropout pattern r at time j for each group. Here, $\pi_{gj} = (\pi_{gj}^{(0)}, \pi_{gj}^{(1)}, \dots, \pi_{gj}^{(j)})$ are not observed and thus are unknown parameters.

Under a fully Bayesian approach, the inferences are derived from the joint posterior distribution $[\pi_g^{(R)}, e_g^{(R)}, \tilde{\lambda}_g, p_g^{(0)} | \text{Data}, l_g, c]$. Where Data=$\{e_{g1}^{(0)}, e_{g2}^{(0)}, \dots, e_{gK}^{(0)}, c_{g1}, \dots, c_{gK}, N_g | g = 0, 1\}$; $p_g^{(0)} = (p_{g1}^{(0)}, p_{g2}^{(0)}, \dots, p_{gK}^{(0)})$ is the set of parameters for the observed data; $e_g^{(R)} = (e_{g1}^{(1)}, e_{g2}^{(1)}, \dots, e_{gK}^{(1)}, \dots, e_{gr}^{(r)}, e_{gr+1}^{(r)}, \dots, e_{gK}^{(r)}, \dots, e_{gK}^{(K)})$ are the unobserved data for the censored subjects; $\pi_g^{(R)} = (\pi_{g1}^{(0)}, \pi_{g1}^{(1)}, \pi_{g2}^{(0)}, \pi_{g2}^{(1)}, \pi_{g2}^{(2)}, \dots, \pi_{gK}^{(0)}, \pi_{gK}^{(1)}, \dots, \pi_{gK}^{(K)})$ are the proportions of subjects who would be at risk at a given time for each missing-data pattern; and $\tilde{\lambda}_g = (\tilde{\lambda}_{g11}, \tilde{\lambda}_{g12}, \dots, \tilde{\lambda}_{g1K}, \dots, \tilde{\lambda}_{grr}, \tilde{\lambda}_{grr+1}, \dots, \tilde{\lambda}_{grK}, \dots, \tilde{\lambda}_{gKK})$.

Draws from this posterior distribution are obtained using the following partition:

$$[\pi_g^{(R)}, e_g^{(R)}, \tilde{\lambda}_g, p_g^{(0)}|\text{Data}, l_g, c] = \underbrace{[\tilde{\lambda}_g|\text{Data}, l_g, c]}_{\text{1: Log–Normal}}\underbrace{[p_g^{(0)}|\tilde{\lambda}_g, \text{Data}, l_g, c]}_{\text{2: Beta}}$$

$$\underbrace{[e_g^{(R)}|\tilde{\lambda}_g, p_g^{(0)}, \text{Data}, l_g, c]}_{\text{3: Binomial}} \underbrace{[\pi_g^{(R)}|e_g^{(R)}, \tilde{\lambda}_g, p_g^{(0)}, \text{Data}, l_g, c]}_{\text{4: Dirichlet}}.$$

We sequentially draw values based on this partition as described in the following steps.

1. **Draw from $[\tilde{\lambda}_g|\text{Data}, l_g, c]$.** In a PM setting, the data provide no evidence for sensitivity parameters, $[\tilde{\lambda}_g|\text{Data}, l_g, c] = [\tilde{\lambda}_g|l_g, c]$ (Little, 1995). Thus, $\tilde{\lambda}_{grk}$ are drawn from

$$\log(\tilde{\lambda}_{grk}) \sim \mathcal{N}\left(\log(l_g) - \frac{\log(1+c^2)}{2}, \log(1+c^2)\right).$$

2. **Draw from $[p_g^{(0)}|\tilde{\lambda}_g, \text{Data}, l_g, c]$.** First note that within the PM approach, the fit of the model to the observed data is identical for all choices of sensitivity parameters or $[p_g^{(0)}|\tilde{\lambda}_g, \text{Data}, l_g, c] = [p_g^{(0)}|\text{Data}]$ (Little, 1995). Let N_g be the initial number of subjects at $t = 0$, let $e_{gk}^{(0)}$ and c_{gk} be the number of the observed events and the censored subjects in the $(t_{k-1}, t_k]$ interval ($g = 0, 1$). Then,

$$e_{gk}^{(0)}|e_{g1}^{(0)}, \ldots, e_{gk-1}^{(0)}, c_{g1}, \ldots, c_{gk}, N_g, p_{gk}^{(0)} \sim \text{Bin}(N_{gk}^{(0)}, p_{gk}^{(0)}),$$

where $N_{gk}^{(0)}$ is the number of subjects at risk and observed in the $(t_{k-1}, t_k]$: $N_{gk}^{(0)} = N_g - \sum_{j=0}^{k-1} e_{gj}^{(0)} - \sum_{j=1}^{k} c_{gj}$ for $k = 1, \ldots, K$, and $e_{g0}^{(0)} = 0$. We assume noninformative independent priors $p_{gk}^{(0)} \sim \text{Beta}(0.001, 0.001)$ from which

$$p_{gk}^{(0)}|e_{g1}^{(0)}, \ldots, e_{gk-1}^{(0)}, c_{g1}, \ldots, c_{gk}, N_g \sim \text{Beta}(e_{gk}^{(0)} + 0.001, N_{gk}^{(0)} - e_{gk}^{(0)} + 0.001).$$

3. **Draw from $[e_g^{(R)}|p_g^{(0)}, \tilde{\lambda}_g, \text{Data}, l_g, c]$.** When $k \geqslant 1$ and $0 < r \leqslant k$, let $N_{gk}^{(r)}$ be the number of subjects censored at visit r who would have been at risk for an event at visit k (no event by t_{k-1}). Let $e_{gk}^{(r)}$ be the number among these $N_{gk}^{(r)}$ subjects who would have had the event in the $(t_{k-1}, t_k]$ interval. Then,

$$e_{gr}^{(r)}|c_{gr}, p_{gr}^{(r)} \sim \text{Bin}(N_{gr}^{(r)}, p_{gr}^{(r)})$$

and for $k > r$,

$$e_{gk}^{(r)}|e_{gr}^{(r)}, e_{gr+1}^{(r)}, \ldots, e_{gk-1}^{(r)}, c_{gr}, p_{gk}^{(r)} \sim \text{Bin}(N_{gk}^{(r)}, p_{gk}^{(r)}).$$

where from (3.2),

$$p_{gk}^{(r)} = \frac{p_{gk}^{(0)}\tilde{\lambda}_{grk}}{1 - p_{gk}^{(0)}(1 - \tilde{\lambda}_{grk})}. \tag{3.3}$$

Here, $N_{gr}^{(r)} = c_{gr}$, and $N_{gk}^{(r)} = N_{gk-1}^{(r)} - e_{gk-1}^{(r)}$ for $k > r$. Within this step, the draw of $e_g^{(r)}$ is sequential, specifically, $e_{gr}^{(r)}$ is drawn first, and then $e_{gr+1}^{(r)}$ up to $e_{gK}^{(r)}$ are drawn next, sequentially conditioned on the previous draws. If $N_{gk}^{(r)} = 0$, then $e_{gk}^{(r)} = 0$.

4. **Draw from $[\pi_g^{(R)}|e_g^{(R)}, p_g^{(0)}, \tilde{\lambda}_g, \text{Data}, l_g, c]$.** Here, $\pi_{gk}^{(r)}$, $r = 1, 2, \ldots, k$, is the proportion of subjects who belong to missing-data pattern r at time t_k among subjects who would be at risk at the beginning of $(t_{k-1}, t_k]$ interval; $\pi_{gk}^{(0)}$ corresponds to the proportion of subjects who are at risk and observed in $(t_{k-1}, t_k]$ interval. Then for each k,

$$(N_{gk}^{(0)}, N_{gk}^{(1)}, \ldots, N_{gk}^{(k)}) \sim \text{Multinomial}(\pi_{gk}^{(0)}, \pi_{gk}^{(1)}, \ldots, \pi_{gk}^{(k)}).$$

Assuming a noninformative Dirichlet prior distribution for $\pi_{gk} = (\pi_{gk}^{(0)}, \pi_{gk}^{(1)}, \ldots, \pi_{gk}^{(k)})$ with parameters $a_{gk}^{(r)} = 0.01$, π_{gk} are drawn from the corresponding posterior Dirichlet distribution $D\left(\frac{N_{gk}^{(0)}}{N_{gk}} + a_{gk}^{(0)}, \frac{N_{gk}^{(1)}}{N_{gk}} + a_{gk}^{(1)}, \ldots, \frac{N_{gk}^{(j)}}{N_{gk}} + a_{gk}^{(k)}\right)$.

5. **Derive draws of marginal parameters p_g.** For each draw of $p_{gk}^{(0)}$, $\tilde{\lambda}_{grk}$, and $\pi_{gk}^{(r)}$, the marginal parameters p_{gk} are calculated as $p_{gk} = \sum_{r=0}^{k} \pi_{gk}^{(r)} p_{gk}^{(r)}$, where $p_{gk}^{(r)}$ for $r > 0$ is derived from (3.3).

5A. **Draws of marginal parameters p_g.** Instead of computing p_{gk} in Step (5), we can simulate draws of p_{gk} from the completed data. This can be viewed as a multiple imputation version of Step 5 and has been proposed in Demirtas and Schafer (2003). Thus, under this approach, once $e_{gk}^{(r)}$ are drawn/imputed, the draws for p_{gk} are obtained based on the imputed data set as if there were no censoring. For each p_{gk}, we assume a noninformative prior distribution, Beta(0.001, 0.001), independent of each other, which results in the following posterior distributions:

$$p_{gk}|e_{g1}, \ldots, e_{gk}, N_g \sim \text{Beta}\left(e_{gk} + 0.001, N_g - \sum_{j=1}^{k} e_{gj} + 0.001\right).$$

Here, $e_{gk} = \sum_{r=0}^{k} e_{gk}^{(r)}$, where $e_{gk}^{(0)}$ are observed and $e_{gk}^{(r)}$ are draws from Step 3. Though Step 5 is probably most computationally efficient, but this alternative step is much easier to implement in available software packages such as WinBUGS. Applying either Step 5 or 5A to the smaller data set analyzed in Section 6, we obtained identical results up to the third decimal. For the TROPHY data, which has a larger number of missing-data patterns, we used the Step 5A approach. Note that, for MAR case ($l_g = 1$, $c = 0$), the p_g (based on the full imputed data) was the same as $p_g^{(0)}$ (based on the observed data) as expected.

Sensitivity analysis. We vary l_g and c to perform sensitivity analysis. When the missing data are MAR, then all $\tilde{\lambda}_{grk} = 1$ or $l_g = 1$ and $c = 0$; thus l_g and c can be considered ignorability indices. When the missing data are MNAR, then not all $\tilde{\lambda}_{grk}$ equal to 1. Finding the best combination of sensitivity parameters is usually not possible, but reasonable choices for such parameters can often suffice. For instance, trying different values of sensitivity parameters in a neighborhood of the MAR ($l_g \approx 1$, small c) could often provide reassuring information about whether the inferences are robust to moderately informative censoring mechanisms. In addition, finding values $l_g = \bar{l}_g$ for which group comparisons are just significant is also important. That is, finding how much different the censored data must be from the observed data for treatment differences to be just significant.

Here, $\tilde{\lambda}_{grk}$ are interpretable and easy-to-use for sensitivity analysis. By considering them random, we (1) provide a probabilistic range for differences among censored and observed subjects, (2) allow for random variation on such differences across different missing-data patterns and over time, (3) reduce the dimension of the sensitivity parameters' space by using the shrinkage parameters (l_g and c) for sensitivity analysis, and (4) account for the overall uncertainty related to censoring by using Bayesian methods.

4. MOTIVATION: TROPHY STUDY

The TROPHY (Julius *and others*, 2006) was an investigator-initiated study to examine whether early treatment of prehypertension might prevent or delay the development of hypertension. Our primary objective was to determine whether for patients with prehypertension, 2 years of treatment with candesartan would reduce the incidence of hypertension for up to 2 years after active treatment was discontinued.

The study consisted first of a 2-year double-blind placebo-controlled phase; followed by a 2-year phase in which all study patients received a placebo. Subjects were examined every 3 months as well as 1 month after the beginning of each phase.

4.1 *Primary event*

The development of hypertension was chosen as the primary study event. It was defined as the first occurrence of any one of the following: (i) systolic blood pressure (SBP) $\geqslant$140 and/or diastolic blood pressure (DBP) $\geqslant$ 90 mmHg at any 3 visits during the 4 years of the study; (ii) SBP $\geqslant$ 160 and/or DBP $\geqslant$100 mmHg at any visit during 4 years; (iii) SBP $\geqslant$ 140 and/or DBP $\geqslant$ 90 mmHg at the end of the study; and (iv) patients requiring treatment as decided by the attending physician. After an event occurred, antihypertension treatment with metoprolol, hydrochlorothiazide, or other medications, with the exception of angiotensin-receptor blockers, was offered at no cost.

The "3 times in 4 years" definition of hypertension differs from more contemporary accepted definition of hypertension based on the guidelines published in the Seventh Report of the Joint National Committee on Hypertension (JNC 7). Following the new guidelines, patients with an average clinic reading of systolic 140 mmHg or higher and/or diastolic of 90 mmHg or higher on 2 *consecutive* clinic visits are now considered to need treatment for hypertension. TROPHY results are published for both definitions (Julius *and others*, 2006, 2008). We focus here on the sensitivity analysis for the definition of hypertension that follows the current JNC 7 guidelines.

4.2 *Censored data*

Data in TROPHY are censored for 2 reasons: (1) if subjects dropout before the event or (2) if subjects violated the protocol. The violation is related to a post hoc change in the definition of the hypertension. That is, a subject who satisfied the 3 times in 4 years criteria of hypertension may not have had 2 consecutive readings of SBP $\geqslant$140 and /or DBP $\geqslant$ 90 and so would not be considered as hypertensive based on the new definition. Following the protocol, treatment of blood pressure (BP) was initiated for such subjects an thus the follow-up BP data are affected and cannot be used.

The intention-to-treat population for the study consisted of 772 patients randomly assigned to candesartan (391) or placebo (381). Among the 772 participants, 109 (54 in placebo) dropped out before developing hypertension. When the new definition of hypertension was used, 92 subjects, classified with hypertension based on the original definition, did not satisfy the new definition (JNC 7). Following the protocol, these subjects received antihypertension treatment, therefore, they are censored at the time the treatment was initiated for the analysis using the new event definition. In addition, 2 subjects who had dropped out satisfy the new event definition. Thus, the number of subjects with a missing event per the new definition is increased to 199 (101 in placebo). Figure 1 shows the rate of censoring, which is higher for subjects in the placebo group.

5. APPLICATION

In this section, we apply the method developed in Section 3 to the data from the TROPHY study using WinBUGS1.4 (Spiegelhalter *and others*, 2003). The effect of candesartan compared with a placebo is
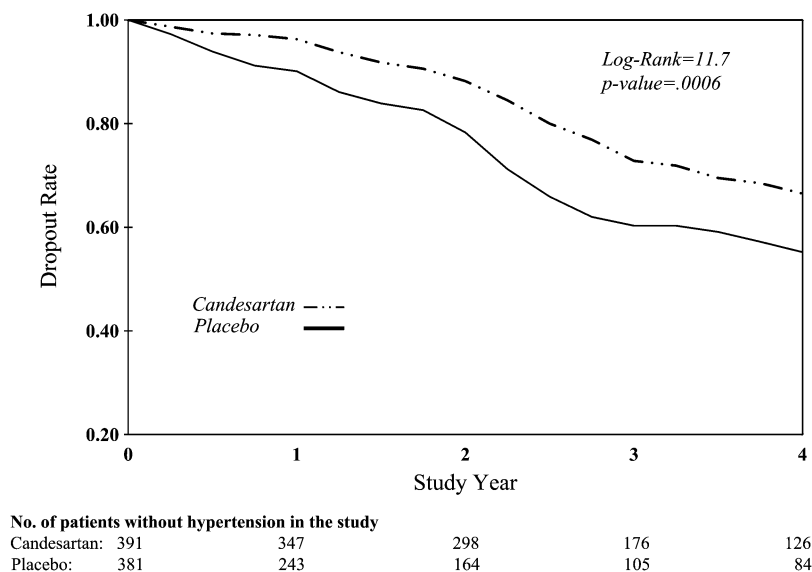
Fig. 1. PL estimate for time-to-dropout in placebo and candesartan group.

estimated over the entire study period as well as its end. Different scenarios about the censoring mechanism are considered to assess the effect of the censored data on the final results. We perform a local sensitivity analysis around the region where missing-data mechanism is MAR ($c = 0$, $l_0 = l_1 = 1$). We consider $c = 0$, $c = 0.5$, and $c = 1$ with l_g varying between 0.8 and 1.2. A value of $l_g > 1$ indicates that, on average, the censored subjects are more likely to develop hypertension than the observed subjects and the opposite would be true for $l_g < 1$. In addition, sensitivity analyses based on assumptions that make the treatment effect just significant, $l_g = \bar{l}_g$, are also included. The results of sensitivity analysis are shown in Table 1. Figure 2 shows the cumulative incidence curves of hypertension for the 2 groups for different values for l_0 and l_1 and $c = 0.5$. The sensitivity analysis shows that the effect of candesartan varies by the assumptions made about the censored data and that the impact of the assumed censoring mechanism on the incidence curves increases over time. That reflects the fact that the cumulative number of the censored subjects that contribute to estimation of the incidence curves increases over time. The bold lines correspond to noninformative censoring, $l_0 = l_1 = 1$ and $c = 0$, and are the same as the PL estimates derived from a frequentist approach using Proc Lifetest in SAS.

The incidence of hypertension at 48 months ranges from 58.6% to 62.1% for the placebo group and from 49.6% to 52.8% for the candesartan group. The overall incidence of hypertension over time is significantly lower in candesartan group for all the assumed scenarios in Table 1. The incidence of hypertension at 48 months is also lower for the candesartan group under all assumed scenarios; but findings become borderline or nonsignificant for censoring mechanisms that increasingly favor the placebo group. For example, in scenario 6b, where the censored subjects compared with the observed subjects (in any interval) are on average 1.32 (95% credible interval [0.47; 2.98]) times more likely to develop hypertension in the candesartan group but are just as likely (95% credible interval [0.36; 2.26]) in the placebo group, then the group difference at 48 months is just significant. In scenario 5b, where the censored subjects compare with the observed subjects are on average 1.2 (95% credible interval [0.43; 2.71]) times more likely to develop hypertension in the candesartan group, but 0.8 times as likely (95% credible interval [0.28; 1.81]) in the placebo group, the effect of candesartan is nonsignificant at 48 months, $p = 0.087$. Under noninformative

Table 1. *Sensitivity analysis under different censoring mechanisms*

		Placebo I_{48} (95% CR)	Candesartan I_{48} (95% CR)	p value (48 months)	Z mean (SD)	Z p value
$c = 0$						
1a	$l_0 = 1, l_1 = 1$	60.8 (55.1–66.4)	50.9 (45.3–56.6)	0.015	0.296 (0.049)	< 0.0001
2a	$l_0 = 1.2, l_1 = 1$	62.1 (56.3–67.7)	50.9 (45.4–56.6)	0.007	0.309 (0.050)	< 0.0001
3a	$l_0 = 1, l_1 = 1.2$	60.8 (55.1–64.4)	52.2 (46.5–57.8)	0.034	0.288 (0.049)	< 0.0001
4a	$l_0 = 1.2, l_1 = 0.8$	62.0 (56.3–67.7)	49.6 (44.1–55.2)	0.002	0.317 (0.049)	< 0.0001
5a	$l_0 = 0.8, l_1 = 1.2$	59.4 (53.8–65.0)	52.2 (46.5–57.8)	0.074	0.273 (0.048)	< 0.0001
6a	$l_0 = 1, \bar{l}_1 = 1.32$	60.8 (55.2–66.4)	52.8 (47.2–58.4)	0.05	0.283 (0.049)	< 0.0001
7a	$\bar{l}_0 = 0.72, l_1 = 1$	58.8 (53.3–64.3)	50.9 (45.4–56.6)	0.05	0.275 (0.048)	< 0.0001
$c = 0.5$						
1b	$l_0 = 1, l_1 = 1$	60.7 (55.0–66.3)	50.8 (45.2–56.5)	0.016	0.295 (0.049)	< 0.0001
2b	$l_0 = 1.2, l_1 = 1$	61.9 (56.1–67.6)	50.8 (45.2–56.5)	0.007	0.308 (0.050)	< 0.0001
3b	$l_0 = 1, l_1 = 1.2$	60.7 (55.0–66.4)	52.0 (46.3–57.7)	0.034	0.288 (0.049)	< 0.0001
4b	$l_0 = 1.2, l_1 = 0.8$	61.9 (56.2–67.6)	49.5 (44.0–55.1)	0.002	0.316 (0.050)	< 0.0001
5b	$l_0 = 0.8, l_1 = 1.2$	59.3 (53.7–64.9)	52.0 (46.3–57.7)	0.017	0.273 (0.049)	< 0.0001
6b	$l_0 = 1, \bar{l}_1 = 1.32$	60.7 (55.0–66.3)	52.7 (46.9–58.3)	0.05	0.283 (0.050)	< 0.0001
7b	$\bar{l}_0 = 0.73, l_1 = 1$	58.8 (53.2–64.4)	50.8 (45.2–56.6)	0.05	0.275 (0.048)	< 0.0001
$c = 1.0$						
1c	$l_0 = 1, l_1 = 1$	60.5 (54.8–66.2)	50.6 (44.9–56.3)	0.016	0.294 (0.049)	< 0.0001
2c	$l_0 = 1.2, l_1 = 1$	61.7 (55.9–67.4)	50.6 (45.0–56.3)	0.008	0.307 (0.050)	< 0.0001
3c	$l_0 = 1, l_1 = 1.2$	60.5 (54.8–66.2)	51.7 (46.0–57.4)	0.033	0.287 (0.050)	< 0.0001
4c	$l_0 = 1.2, l_1 = 0.8$	61.7 (55.9–67.4)	49.6 (44.1–55.2)	0.002	0.317(0.049)	< 0.0001
5c	$l_0 = 0.8, l_1 = 1.2$	59.2 (53.5–64.8)	51.7 (46.0–57.4)	0.068	0.274 (0.048)	< 0.0001
6c	$l_0 = 1, \bar{l}_1 = 1.32$	60.5 (54.7–66.2)	52.3 (46.6–58.1)	0.05	0.283 (0.050)	< 0.0001
7c	$\bar{l}_0 = 0.72, l_1 = 1$	58.6 (53.0–64.2)	50.6 (44.9–56.3)	0.05	0.274 (0.048)	< 0.0001

censoring mechanism the incidence of hypertension at 48 months was lower for the candesrtan group, at 50.9%, compared with the placebo group, 60.8%. Finally, when comparing the overall incidence curves, the significance is strong regardless of the censoring mechanism. This is because large differences between curves occur early when the number of the censored subjects is small, and thus such differences are fairly robust to the assumptions on the censoring mechanism.

In summary, the differences in the incidence of hypertension between 2 groups over the entire study are robust for different censoring mechanisms. However, the group differences at 48 months are more sensitive to the censoring mechanism, varying from 12.4% under scenario 4a to 7.2% under scenario 5a. When the censoring mechanism is noninformative, a 2-year treatment with candesartan reduces the incidence of hypertension at 4 years by 9.9% (95% credible interval [1.9; 17.8]). In addition, the effect candesartan treatment is robust around a local region of the MAR mechanism. It would take differences on the odds of developing hypertension among dropouts favoring placebo (scenarios 6–7) for the effect of treatment to become borderline significant.

6. COMPARATIVE STUDIES

The proposed model can be directly applied to other studies with similar design. For illustration as well as comparison with other alternative approaches, we apply our method to the data analyzed by Scharfstein *and others* (2001) (SRER), which are relevant for this setting. SRER analyzed data from a randomized
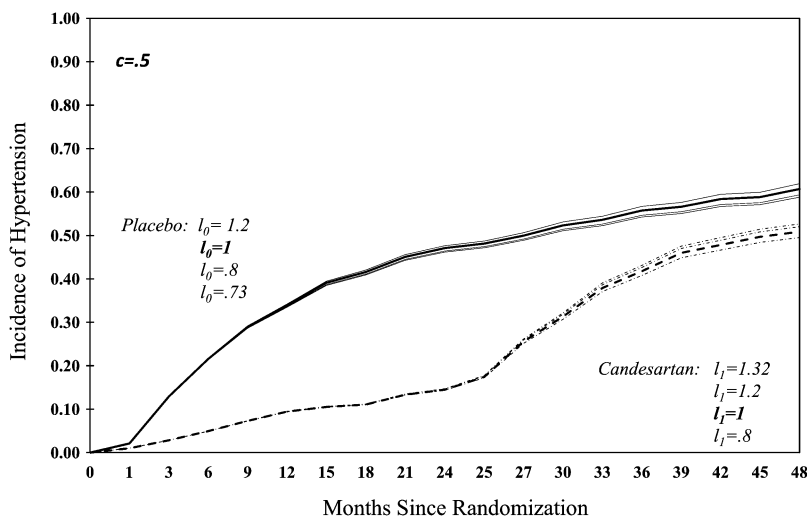
Fig. 2. Sensitivity analysis for cumulative incidence of hypertension between placebo and candesartan groups over a 48-month period. Bold lines correspond to noninformative censoring, $l_0 = l_1 = 1$ and $c = 0$.

clinical trial focusing on the effect of 2 mg risperidone in treating schizophrenia. The endpoint outcome was defined as a 20% or more reduction in the Positive and Negative Symptom Scale relative to the baseline value. The subjects were followed up at 1, 2, 4, 6, and 8 weeks; data are shown in their Table 1 (for full details of the study refer to their paper). Following a selection model approach, SRER propose a model to perform sensitivity analysis under different censoring mechanisms by introducing censoring bias functions q_t, for $t = 1, 2, \ldots, K$. They used intuitive parametric functions $q_t(T; \alpha, Trt = g) = \alpha_g\{T - (t + 1)\}$ to reduce the number of the sensitivity parameters but, as the authors pointed out in their discussion, one can consider an infinite number of forms for q_t, and therefore the results of the sensitivity analysis would be specific to the parametric functions q_t. Specifically, they chose a parameterization where α is the log-odds ratio of dropping out between t and $t + 1$ per unit increase in T for subjects who were at risk at t. As an alternative, we propose a full Bayesian approach within a PM setting using intuitive identifying parameters $\tilde{\lambda}_{rt}$, for sensitivity analysis, which are assumed random following a log-normal distribution: $E(\tilde{\lambda}_{rt}) = l$, $\text{Var}(\tilde{\lambda}_{rt}) = c^2 l^2$. There is no an exact relationship between our identifying parameters, l and c, and SRER sensitivity parameter α. However, our sensitivity parameters and theirs have an overall connection in terms of measuring the degree and the nature of the departure from a noninformative censoring mechanism.

Under noninformative censoring $l = 1$ and $c = 0$ that corresponds to $\alpha = 0$. When censoring is informative, $l > 1(< 1)$ corresponds to $\alpha < 0(> 0)$. That is, an $l > 1$ indicates that compared with the observed subjects, the censored subjects are more likely to have an endpoint in t to $t + 1$ and so are more likely to have a shorter event time T. Or equivalently, subjects with smaller T are more likely to be censored, which corresponds to a negative α. For small departures from a noninformative censoring mechanism, l is close to 1, which corresponds to α close to 0.

Next, we apply our method to the data analyzed by SRER. Our results of the sensitivity analysis are summarized in Figure 3, which similar to Figure 2(b) in SRER, showing a contour plot for $p$value $= 0.05$. The region favoring risperidone is the upper left corner region of l_0, l_1. That is consistent with the corresponding region shown in SRER Figure 2(b) (the lower right corner region of α_0, α_1). Under either approach, no significant difference appears at $p \leqslant 0.05$ between treatment and placebo—assuming
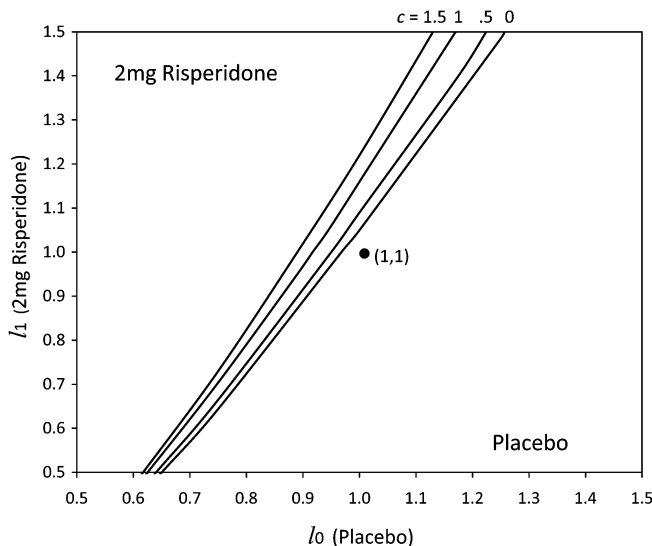
Fig. 3. Sensitivity analysis on the data from SRER.

noninformative censoring. However, it requires only small departures from a noninformative censoring mechanism for the difference to become significant. For instance, treatment is better when $l_0 = 1$, $l_1 = 1.07$, and $c = 0.5$. That is, the censored subjects compared with the observed subjects (in any interval) are on average 1.07 times more likely to have an endpoint in the treatment group but are just as likely in the placebo group. Both methods show that small departures from noninformative censoring mechanism produce significant results.

The strength of our model lies in using a fully Bayesian approach with random sensitivity parameters, which are easy-to-use, and at the same time does not assume a parametric function on the censoring mechanism. Alternatively, one could expect that if a good parametric function for the censoring mechanism is known, the sensitivity analysis based on the correct specification of such function might well be more efficient.

7. CONCLUDING REMARKS

In this paper, we presented a Bayesian model for time-to-event data with informative censoring. We approached the missing data problem by using a PM model, which is identified by introducing an intuitive parameterization. The identifying parameters, $\tilde{\lambda}_{grk}$, are defined for each group g as: the odds ratio of having an endpoint in $(t_{k-1}, t_k]$, conditioned on being at risk at t_{k-1}, between subjects from missing-data pattern r with the observed subjects in that interval ($k = 1, \ldots, K$). Sensitivity analyses are performed using different prior distributions of $\tilde{\lambda}_{grk}$. Even though the distribution of $\tilde{\lambda}_{grk}$ is unknown, it is possible for a subject matter expert to give it a range and then explore the sensitivity of statistical inferences over such a range. The new parameterization is a nonparametric approach for modeling the censoring mechanism; and therefore is robust to model misspecification that may result from parametric modeling. However, when the number of follow-up visits increases, so does the number of the sensitivity parameters under a nonparametric approach; and soon a fully unrestricted nonparametric approach becomes impractical. The robustness of modeling the correct dropout mechanism and reducing the number of the sensitivity parameters are competing requirements requiring compromise and adjustment according to the

specific problem. We reduced the dimensionality of the sensitivity space by using a stochastic constraint to incorporate different levels of uncertainty. That is, we let $\tilde{\lambda}_{grk}$ follow a log-normal distribution with mean l_g and variance $c^2 l_g^2$. Then the set of the sensitivity parameters is reduced to l_g and c. Our proposed parameterization is flexible for sensitivity analysis and contains noninformative censoring as a special case ($l_g = 1$, $c = 0$). We assumed no knowledge on l_g, and sensitivity analysis could be performed for l_g values close to 1.

REFERENCES

BIRMINGHAM, J., ROTNITZKY, A. AND FITZMAURICE, G. M. (2003). Pattern-mixture and selection models for analyzing longitudinal data with monotone missing patterns. *Journal of the Royal Statistical Society, Series B* **65**, 275–297.

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

DANIELS, M. AND HOGAN, J. (2007). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis.* Boca Raton, FL: Chapman & Hall.

DEMIRTAS, H. AND SCHAFER, J. L. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine* **22**, 2553–2575.

GUO, W., RATCLIFFE, S. J. AND TEN HAVE, T. T. (2004). A random pattern-mixture model for longitudinal data with dropouts. *Journal of the American Statistical Association* **99**, 929–937.

HJORT, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Annals of Statistists* **18**, 1259–1294.

IBRAHIM, J. G., CHEN, M. AND SINHA, D. (2001). *Bayesian Survival Analysis.* New York: Springer.

JULIUS, S., KACIROTI, N., NESBITT, S., EGAN, B. M. AND MICHELSON, E. L. FOR THE TRIAL OF PREVENTING HYPERTENSION (TROPHY) INVESTIGATORS (2008). TROPHY study: outcomes based on the JNC 7 definition of hypertension. *Journal of the American Society of Hypertension* **2**, 39–43.

JULIUS, S., NESBITT, S., EGAN, B., WEBER, M. A., MICHELSON, E. L., KACIROTI, N., BLACK, H. R., GRIMM, R. H., MESSERLI, F. H., OPARIL, S. AND SCHORK, M. A. FOR THE TRIAL OF PREVENTION HYPERTENSION (TROPHY) STUDY INVESTIGATORS (2006). Feasibility of treating prehypertension with an angiotensin-receptor blocker. *New England Journal of Medicine* **354**, 1685–1697.

KACIROTI, N., RAGHUNATHAN, T. E., SCHORK, M. A. AND CLARK, N. M. (2008). A Bayesian model for longitudinal count data with non-ignorable dropout. *Applied Statistics* **57**, 521–534.

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. *Journal of the American Statistical Association* **101**, 435–446.

KACIROTI, N., SCHORK, M. A., RAGHUNATHAN, T. E. AND JULIUS, S. (2009). A Bayesian sensitivity model for intention-to-treat analysis on binary outcomes with dropouts. *Statistics in Medicine* **28**, 572–585.

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. *Journal of the American Statistical Association* **88**, 125–134.

LITTLE, R. J. A. (1995). Modeling the dropout mechanism in repeated measures studies. *Journal of the American Statistical Association* **90**, 1113–1121.

LITTLE, R. J. A. AND RUBIN, D. B. (2002). *Statistical Analysis With Missing Data*, 2nd edition. New York: John Wiley.

MA, G., TROXEL, A. B. AND HEITJAN, D. (2005). An index of local sensitivity to nonignorable drop-out in longitudinal modeling. *Statistics in Medicine* **24**, 2129–2150.

MOLENBERGHS, G. AND KENWARD, M. G. (2007). *Incomplete Data in Clinical Studies*. Chichester, UK: John Wiley.

MOLENBERGHS, G., MICHIELS, B., KENWARD, M. G. AND DIGGLE, P. J. (1998). Monotone missing data and pattern-mixture models. *Statistica Neerlandica* **52**, 153–161.

PHADIA, E. G. AND SUSARLA, V. (1983). Nonparametric Bayesian estimation of a survival curve with dependent censoring mechanism. *Annals of the Institute of Statistical Mathematics* **35**, 389–400.

ROTNITZKY, A., SCHARFSTEIN, D. O., SU, T. AND ROBINS, J. M. (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics* **57**, 103–113.

ROY, J. AND DANIELS, M. (2008). A general class of pattern-mixture models for nonignorable dropout with many possible dropout times. *Biometrics* **64**, 538–545.

SCHARFSTEIN, D. O., DANIELS, M. J. AND ROBINS, J. M. (2003). Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Biostatistics* **4**, 495–512.

SCHARFSTEIN, D. O. AND ROBINS, J. M. (2002). Estimation of the failure time distribution in the presence of informative censoring. *Biometrika* **89**, 617–634.

SCHARFSTEIN, D. O., ROBINS, J. M., EDDINGS, W. AND ROTNITZKY, A. (2001). Inference in randomized studies with informative censoring and discrete time-to-event endpoints. *Biometrics* **57**, 404–413.

SPIEGELHALTER, D. J., THOMAS, A., BEST, N. G. AND LUNN, D. (2003). *Sampling Based Approaches to Calculate Marginal Densities*. Cambridge, UK: MRC Biostatistics Unit. WinBUGS User Manual: Version 1.4.

SUSARLA, V. AND VAN RYZIN, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *Journal of the American Statistical Association* **71**, 897–902.

TSAI, W. (1986). Estimation of survival curves from dependent censorship models via a generalized self-consistent property with nonparametric Bayesian estimation application. *The Annals of Statistics* **14**, 238–249.

WU, M. C. AND CARROLL, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188.

ZHENG, M. AND KLEIN, J. P. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* **82**, 127–138.