A Bayesian Approach for the Analysis of Panel-Count Data with Dependent Termination

Debajyoti Sinha,

Department of Biometry, Medical University of South Carolina, Charleston, South Carolina 29425, U.S.A.

Tapabrata Maiti

Department of Statistics, Iowa State University, Ames, Iowa 50011, U.S.A.

Summary.

We consider modeling and Bayesian analysis for panel-count data when the termination time for each subject may depend on its history of the recurrent events. We propose a fully specified semiparametric model for the joint distribution of the recurrent events and the termination time. For this model, we provide a natural motivation, derive several novel properties, and develop a Bayesian analysis based on a Markov chain Monte Carlo algorithm. Comparisons are made to other existing models and methods for panel-count data. We demonstrate the usefulness of our new models and methodologies through the reanalysis of a data set from a clinical trial.

Résumé

Nous considérons la modélisation et l'analyse bayésienne pour des données de dénombrements groupés quand le temps de terminaison de chaque sujet peut dépendre de l'historique des événements récurrents. Nous proposons un modèle semi-paramétrique complètement spécifié pour la distribution jointe des événements récurrents et le temps de terminaison. Pour ce modèle, nous fournissons une motivation naturelle, dérivons plusieurs propriétés nouvelles et développons une analyse bayésienne basée sur un algorithme de Monte-Carlo par chaînes de Markov. Des comparaisons sont effectuées avec d'autres modèles et méthodes pour des dénombrements d'événements groupés. Nous montrons l'utilité de nos nouveaux modèles et méthodologies par la ré-analyse des données d'un essai clinique.

Keywords

Frailty; Gibbs sampling; Posterior distribution; Recurrent events

1. Introduction

Grouped recurrent event times data, popularly called panel-count data, occur when a subject, inspected over time only at scheduled time points $0 = a_0 < a_1 < \cdots < a_J$, may experience repeated occurrences of an important but possibly nonfatal event. Suppose $N_i(t)$ denotes the

sinhad@musc.edu.

number of nonfatal events experienced by subject i up to time t. The number of new recurrent events $N_{ij} = N_i(a_j) - N_i(a_{j-1})$ to subject i occuring in the inspection interval $I_j = (a_{j-1}, a_j]$ is recorded until the termination time T_i, when the subject is removed from observation either due to an adverse response such as detection of a relapse to malignancy or due to a positive response such as cure. We are assuming that the inspection times are the same for all subjects and T_i is discrete and take values only $a_1, \ldots, a_J$. Let x_i be the explanatory variable for subject i. We use the notation $D = \{N_{ij} : j = 1, \ldots, T_i; x_i; T_i : i = 1, \ldots, n\}$ to denote the observed panel-count data.

See Gail, Santner, and Brown (1980), Oakes (1982), Sinha (1993), Scheike and Zhang (1998), Sun and Wei (2000), and the references therein for different examples of panel-count data from various clinical trials and animal experiments. Earlier works on modeling panel-count data often assumed that given the covariate x of the subject, the risk of termination at any time point t does not depend on the past history of nonfatal events $\mathcal{N}(t)$. But, this assumption will not be valid when, say, the termination is caused by either a significant positive response (say, cure or discharge from the treatment) or a negative response (such as relapse to malignancy). This risk of termination is often associated with the history of nonfatal events. For the bladder cancer data example (Sun and Wei, 2000) from the Veterans Administration Cooperative Urological Research Group (VACURG), the time-independent covariates available for each patient are a binary variable indicating treatment (1 for treatment thiotepa and 0 for placebo), number of tumors at entry, and the size of the largest tumor at entry. After the surgery to remove his/her primary tumor, each patient was monitored every month during scheduled clinic visit for number of new (since the previous clinic visit) superficial bladder tumors. A patient was terminated from further observation during a particular clinic visit probably due to clinically significant improvement of his/her disease symptoms. As the treatment (thiotepa and placebo) had to be instilled in the bladder, we can expect that a patient who is very prone to superficial bladder tumors is also expected to take longer than usual time to termination. A previous analysis of this data ignored the possible association between the risk of termination and the history of superficial tumors. Unlike the previous analysis, we treat both the event of termination (cure) and the recurrent events of superficial tumors as important responses.

Various methods of panel-count data and recurrent event times data, published so far, have put high emphasis on partially specified models. In such work, the model and associated inferential procedure depend on the particular underlying inferential objective, which can be of the following four types: (1) inference about $N(t)$ (e.g., Sun and Wei, 2000), under the assumption that given x, the process $N(t)$ is independent of T; (2) marginal inference about $N(t)$ unconditional on the past history of events (e.g., Wang, Qin, and Chang, 2001); (3) inference about T given the history of $N(\cdot)$; and (4) inference about the termination modulated process $N(t)1_{[T \geq t]}$ (e.g., Ghosh and Lin, 2000), where 1_A is the indicator function.

In many applications, it may be desirable to have a fully specified model which can be used for addressing all four objectives together within a unified statistical analysis and also perform prediction, model diagnostic, simultaneous inferences on multiple parameters, etc. Especially when both the nonfatal events and the termination are deemed as important responses, and the termination depends on the history of nonfatal events, addressing each

objective through separate analysis may be misleading. The results of the analysis of different partially specified models focusing on different objectives may conflict with each other and may be difficult to unify. Our aim is to develop and analyze a fully specified stochastic model of the joint distribution of the nonfatal events and the termination time. This new model also allows us to understand the prognostic values of the history of nonfatal events for evaluating the risk of termination.

A fully parametric model by Lancaster and Intrator (1998), here onward referred to as the LI model, allows *only* positive dependence between a subject's history of nonfatal events and risk of termination. Our model represents a much broader class than the LI model and we can incorporate a negative association between $N(t)$ and T. The data example discussed in this article may require a negative association between $N(t)$ and T. The details of the comparison between our model and the LI model are given later. Recently introduced class of models for panel-count data in animal carcinogenicity experiments (e.g., Dunson, 2000) also do not incorporate the association between T and $N(t)$.

In Section 2, we present a fully specified stochastic model for panel-count data with a dependent termination process. We present novel properties of our model and relationship of our model to other existing models for recurrent events data. Existing Bayesian literature on panel-count data is very limited. See Ibrahim, Chen, and Sinha (2001) for a current literature survey on Bayesian survival analysis. One of our goals is to demonstrate the use and advantages of the Bayesian paradigm for such problems. In Section 3, we present the likelihood of our model and associated Bayesian methods for analyzing such data. In Section 4, our methodology is illustrated with the analysis of panel-count data from the bladder cancer study from Sun and Wei (2000). In Section 5, we conclude our article with some closing remarks.

2.   Model for Panel-Count Data

We assume that each $N_i(t)$ is a realization of a compound nonhomogeneous Poisson process (Karlin and Taylor, 1981). Given the subject-specific unobservable frailty w_i, the cumulative number of nonfatal events to the subject i, under observation at time t (that is, $T_i \geq t$) is given by

$$N_i(t) \mid x_i, w_i \sim \text{Poi}\{\mu_0(t)\exp(\beta_1 x_i)w_i\}, \tag{1}$$

where $\mu_0(t)$ is the baseline cumulative intensity and β_1 is the regression parameter of the covariate effect on the conditional intensity in (1). Here $w_1, \ldots, w_n$ are assumed to be independent $Ga(\eta, \eta)$ with common density $g(w \mid \eta) = \eta^\eta w^{\eta-1} e^{-\eta w}/\Gamma(\eta)$. The mean of w_i is taken to be unity to assure the identifiability of the model parameters and the variance of w_i is $\kappa = \eta^{-1}$. We assume that $\kappa > 0$ to exclude the oversimplified situation when, within the same subject, the increments of $N(t)$ in disjoint intervals are independent. A finite mean frailty distribution is warranted here to assure a finite expected number of events within any finite time interval.

The model further assumes that the hazard of T_i at time t depends on the history of point process $\mathcal{N}_i(t) = \{N_i(s) : s < t\}$ via the unobservable frailty w_i. We use a discretized version of the Cox model (Cox, 1972),

$$P\big[T_i = a_j \mid x_i, w_i, \mathcal{N}_i(a_j), T_i > a_{j-1}\big] = 1 - \pi_j^{\exp(\beta_2 x_i) w_i^\alpha}, \tag{2}$$

where β_2 is the regression parameter and $\pi_j = P\big[T = a_j \mid x = 0, w = 1, \mathcal{N}(t), T > a_{j-1}\big]$ is the discretized baseline conditional hazard rate.

Here, $-\infty < \alpha < +\infty$ quantifies the nature of dependence between $\mathcal{N}_i(t)$ and the hazard of T_i. When $\alpha = 0$, we have $P\{T_i = a_j \mid x_i, T_i > a_{j-1}, \mathcal{N}_i(t)\}$ free of $\mathcal{N}_i(t)$ and we get the model of Sinha (1993), where the hazard of T_i at time t is assumed independent of $\mathcal{N}_i(t)$. When $\alpha > 0$, there is positive association between N_i and T_i, that is, given the same value of x_i, a subject with a higher rate of nonfatal events has a higher risk of termination than another with a lower rate of nonfatal events. Similarly, $\alpha < 0$ indicates a negative relationship between T_i and N_i. When $\alpha > 1$, the variability of termination times are more than the heterogeneity of subjects in terms of their rates of nonfatal events and $0 < \alpha < 1$ implies the opposite. Again, $\alpha = 1$ implies that the heterogeneity among subjects in terms of rates of nonfatal events is the same as the heterogeneity in terms of their termination times. When $\alpha = 1$ and $\pi_j = \exp[-\rho\{\mu_0(a_j) - \mu_0(a_{j-1})\}]$, we get the LI model (Lancaster and Intrator, 1998). This shows that the LI model allows only positive association and has a restrictive assumption of a very precise algebraic relationship between π_j and μ_0. We investigate some further useful and desirable properties of our model.

Theorem 1.

The conditional risk of termination, $P\big[T_i = a_j \mid \mathcal{N}_i(a_j), x_i, T_i > a_{j-1}\big]$, depends on the events history $\mathcal{N}_i(a_j)$ only through the cumulative number of events $N_i(a_j)$. For the special case $\alpha = 1$,

$$P\big[T_i = a_j \mid \mathcal{N}_i(a_j), x_i, T_i > a_{j-1}\big]$$
$$= 1 - \left[1 - \frac{\pi_j^{\exp(\beta_2 x_i)}}{\exp(\beta_1 x_i)\mu_0(a_j) - \exp(\beta_2 x_i)\log(S_j)}\right]^{\eta + N(a_j)}, \tag{3}$$

where $S_j = \prod_{k=1}^{j} \pi_k$.

The proof of Theorem 1 is given in the Appendix. Please note that Theorem 1 is very different from Implications 1–3 in Lancaster and Intrator (1998). When $\alpha \neq 1$, we can numerically evaluate the quantities $P\big[T_i = a_j \mid \mathcal{N}_i(a_j), x_i, T_i > a_{j-1}\big]$ and $P[T_i = a_j \mid x_i, T_i > a_{j-1}]$, but no closed form expressions of these quantities are available. Unlike the model of Wang et al. (2001), $P(T_i = a_j \mid x_i, T_i > a_{j-1})$ in our model does not follow a discretized Cox structure. We present the following theorem about the conditional distribution of nonfatal events in any interval given past history of events.

Theorem 2.

The conditional distribution of $N_{ij} \mid x_i, \mathcal{N}_i(a_{j-1}), T_i > a_{j-1}$ depends on $\mathcal{N}_i(a_{j-1})$ only through the $N_i(a_{j-1})$. For the case $\alpha = 1$, this distribution is a generalized negative binomial with conditional expectation

$$
E[N_{ij} \mid \mathcal{N}_i(a_{j-1}), x_i, T_i > a_{j-1}]
= \{\eta + N_i(a_{j-1})\}\left[\frac{\exp(\beta_1 x_i)\{\mu_0(a_j) - \mu_0(a_{j-1})\}}{\eta + \exp(\beta_1 x_i)\mu_0(a_j)}\right]. \tag{4}
$$

The proof is given in the Appendix. The result in (4) is similar to the property of the model proposed by Oakes (1982). These two theorems also demonstrate that our model is very different from the modulated renewal process of Oakes and Cui (1994). For $\alpha \neq 1$, no such simple expression exists for $E[N_{ij} \mid \mathcal{N}_i(a_{j-1}), x_i, T_i > a_{j-1}]$, but, these can be numerically computed given the parameter values. Following the steps similar to the derivation of (4), we can also numerically evaluate $E[N_{ij} \mid x_i, T_i > a_{j-1}]$. The details are omitted here. These demonstrate that through our fully specified model along with Bayesian analysis, we can estimate different important model quantities of interest using a single unified inference procedure.

3. Likelihood and Posterior

It follows from (1) that, given w_i, N_{ij} for $j = 1, \ldots, J$ are independent Poi$\{w_i \exp(\beta_1 x_i)/\mu_j\}$, where $\mu_j = \mu_0(a_j) - \mu_0(a_{j-1})$ and Poi(γ) denotes the Poisson distribution with mean γ. The likelihood based on the sampling distributions of the N_{ij} is

$$
L_1(\beta_1, \mu_0, \mathbf{w} \mid D) \propto \prod_{i=1}^{n} \prod_{j \in M_i} \left[e^{\beta_1 x_i} w_i \mu_j\right]^{N_{ij}} \exp\left[-e^{\beta_1 x_i} w_i \mu_j\right], \tag{5}
$$

where M_i, is the set of inspection intervals when the subject i is under observation. Using (2), we obtain the likelihood based on the sampling distributions of the $T_1, \ldots, T_n$ as

$$
L_2(\beta_2, \pi, \mathbf{w} \mid D) \propto \prod_{i=1}^{n} \left[1 - \pi_{(i)}^{w_i^\alpha \exp(\beta_2 x_i)}\right] S_{(i)}^{\exp(\beta_2 x_i) w_i^\alpha}, \tag{6}
$$

where $t_i = a_{(i)} \in \{a_1, \ldots, a_J\}$ is the termination time of the subject i and $S_{(i)} = \prod_{k=1}^{(i)-1} \pi_k$. In (2), we use a discretized Weibull baseline hazard rate $\pi_j = \exp\{-\lambda_j(a_j^\nu - a_{j-1}^\nu)\}$, with unknown parameter vector $\phi = (\lambda, \nu)$. This model on π_j allows for a very wide class for the hazard. In principle, it is possible to use a more flexible model with completely unspecified π_j.

The observed data likelihood $L_0(\beta_1, \beta_2, \alpha, \phi, \eta \mid D)$ is obtained via the multiple integration $L_0(\beta_1, \beta_2, \alpha, \phi, \eta \mid D) \propto \int L_1 \times L_2 \times g(\mathbf{w} \mid \eta) \prod_{i=1}^{n} dw_i$, where $g(\mathbf{w} \mid \eta)$ is the joint density of $\mathbf{w} = (w_1, \ldots, w_n)$ given η. We present a semiparametric Bayes survival analysis approach (e.g., Sinha and Dey, 1997; Ibrahim et al., 2001) using a prior process to summarize the prior

information on $\mu_0(\cdot)$ and a prior distribution for the rest of the parameters. We assume that the prior information about $\mu_0(t)$ is independent of the prior information about the rest of the parameters, $\pi(\mu, \beta, \eta, \phi, a) = \pi(\mu) \times \pi(\beta, \eta, \phi, a)$. As (5) involves μ_0 only through $(\mu_1, \dots, \mu_J)$, we only need the joint prior density of the increments $(\mu_1, \dots, \mu_J)$. Using an independent increment gamma process prior (Kalbfleisch, 1978) on $\mu_0(t)$ we have the prior distribution $\mu_j \sim Ga(c\gamma_j, c)$ with independence for j = 1, ... , J. The joint prior density of the increments is given by

$$\pi(\mu) = \prod_{j=1}^{J} \frac{c^{c\gamma_j} \mu_j^{c\gamma_j - 1}}{\Gamma(c\gamma_j)} \exp(-c\mu_j). \tag{7}$$

The hyperparameters, c and γ_j's, assumed to be known here, are determined by the prior expectation (mean) and the precision of the prior expectation of $\mu_0(t)$. But, it is possible to put another level of hierarchy in this model using a hyperprior for (c, γ).

The joint posterior for our model is given by

$$p(\beta, \mathbf{w}, \eta, \phi, \alpha \mid D) \propto L_1(\beta_1, \mu, \mathbf{w} \mid D) \times L_2(\beta_2, \phi, \mathbf{w} \mid D) \times g(\mathbf{w} \mid \eta) \times \pi(\beta, \phi, \eta, \alpha) \\ \times \pi(\mu), \tag{8}$$

where $g(\mathbf{w} \mid \eta) = \prod_i \left[\eta^\eta w_i^{\eta - 1} \exp(-w_i \eta) / \Gamma(\eta) \right]$ is the product of gamma frailty densities and $\pi(\beta, \phi, \eta, a)$ is given below. Instead of dealing with complicated $L_0((\beta_1, \beta_2, a, \phi, \eta \mid D)$, we use popular MCMC tools (e.g., Chen, Ibrahim, and Shao, 2000) to sample from (8) and then use those samples to do the posterior inference. For practical convenience, we further assume that

$$\pi(\beta, \eta, \phi, \alpha) = \pi(\beta_1) \times \pi(\beta_2) \times \pi(\phi) \times \pi(\eta) \times \pi(\alpha). \tag{9}$$

This assumption is reasonable, as β_1 and β_2 are the parameters associated with effects of covariate x, respectively, on $N(t)$ and on T. In many practical applications, the priors on β_1 and β_2 will be very noninformative to balance both the skeptical and the enthusiastic views about the effects of covariates such as assigned treatments. It is also a common practice to assume priors on baseline functions μ and π to be independent of the rest of the parameters. As we can expect to have independent expert opinions about two different types of events $N(\cdot)$ and T, we have the assumption of mutual independence between the set of parameters associated with $N(\cdot)$ and the set associated with T. The prior belief on a, the parameter quantifying the association between the two types of events, is expected to be obtained independently from others. The parameter $\pi(a)$ is assumed to be $N(a_0, \sigma_3)$ with $a_0 = 0$ as we do not know a priori whether the association between the two classes of events should be either positive or negative. For further convenience we use $N(\beta_{10}, \sigma_1)$ for $\pi(\beta_1)$, $N(\beta_{20}, \sigma_2)$ for $\pi(\beta_2)$, and $Ga(\lambda \mid A_1, B_1) \times Ga(\nu \mid A_2, B_2)$ for the density of $\pi(\phi)$. We are choosing the forms of the priors that are flexible, yet facilitate easy-to-sample conditional posteriors at each iteration of the MCMC. We would like to emphasize that if needed, MCMC is very capable of handling more complicated prior structures such as a multivariate $\pi(\beta, \eta, \phi, a)$ inducing prior dependencies among the parameters.

The full conditional distributions for the MCMC steps are given as follows using the notation $[\theta \mid rest]$ to denote the conditional distribution of ϕ given all other parameters:

I. $\quad [\mu_j|rest] \sim Ga\{\tau\Lambda_j + \sum_i N_{ij}V_{ij}, \tau + \sum_i w_i \exp(\beta_1 x_i)V_{ij}\}$

II. $\quad [\lambda \mid rest] \sim Ga\{m + 1 + A_1, B_1 + \sum_i t_i^\nu w_i^\alpha \exp(\beta_2 x_i)\}$

III. $\quad [\nu \mid rest] \propto \prod_i \nu t_i^{\nu-1} \exp\{-\lambda t_i^\nu w_i^\alpha \exp(\beta_2 x_i)\} \times \pi(\nu)$

IV. $\quad [\beta_1 \mid rest] \propto \exp\{\sum_i \beta_1 x_i \sum_j N_{ij}V_{ij} - \sum_i w_i \exp(\beta_1 x_i) \sum_j \mu_j V_{ij}\} \times \pi(\beta_1)$

V. $\quad [\beta_2 \mid rest] \propto \exp\{\sum_i \beta_2 x_i - \lambda t_i^\nu w_i^\alpha \exp(\beta_2 x_i)\} \times \pi(\beta_2)$

VI. $\quad [\eta \mid rest] \propto \prod_i \left[\{\eta^\eta/\Gamma(\eta)\}w_i^{\eta-1}\exp(-\eta\, w_i)\right] \times \pi(\eta)$

VII. $\quad [w_i \mid rest] \propto w_i^{\Sigma_j N_{ij}V_{ij} + \alpha + \eta - 1} \times \exp\left[-w_i\{\exp(\beta_1 x_i)\sum_j \mu_j V_{ij} + \eta\}\right.$
$\left. - \lambda w_i^\alpha t_i^\nu \exp(\beta_2 x_i)\right]$

VIII. $\quad [\alpha \mid rest] \propto \prod_i w_i^\alpha \exp\{-\lambda w_i^\alpha t_i^\nu \exp(\beta_2 x_i)\} \times \pi(\alpha)$.

Here V_{ij} is the indicator variable of whether subject i is examined at time a_j. Except for μ_j and λ, all other conditional densities are not in standard form. Fortunately, the conditional densities for β_1, β_2, η, ν, and α are log concave. So, a readily available software program for the adaptive rejection sampling method (Gilks and Wild, 1992) has been used to sample from these log-concave densities. To avoid lengthening the article, the proofs of their log concavities are omitted. Notice that the conditional density of each w_i is proportional to $f(w)$ $\times h(w)$, where $f(w)$ is proportional to gamma density function and $h(w)$ is a bounded function. To sample w_i's, we have used a Metropolis-Hastings algorithm with acceptance probability $r(w_{\text{old}}, w_{\text{new}}) = \min[\{h(w_{\text{new}})/h(w_{\text{old}})\}, 1]$. Here, w_{new} is the candidate value simulated from $f(w)$ and w_{old} is the current value of w_i at this iteration. In particular we have used 10 Gibbs chains with 20,000 iterations in each chain. To monitor the convergence of MCMC, we have checked that the estimated scale reduction factor $\hat{R}$ of Gelman and Rubin (1992) is less than 1.2 for each of the parameters. This is a very popular and easy method to monitor the convergence of the MCMC algorithm. We have also checked several other convergence diagnostics of the MCMC using the publicly available software package CODA (Best, Cowles, and Vines, 1995).

4. Analysis of the Bladder Cancer Data

We reanalyze the bladder cancer data from the VACURG (Sun and Wei, 2000), allowing a possible association between the history of superficial tumor occurrences and the risk of termination. The data set is available at the website www.blackwellpublishers.co.uk/rss. We use only the binary regression variable treatment in our model. For each patient, some of the monthly clinic visits (before termination) have missing observations as the patients have missed some of their clinic visits. We assume here that patterns of missed clinic visits are noninformative in the sense that the probability of a patient missing a particular clinic visit does not depend on the unobserved number of tumors in that month. This assumption of missing at random (Little and Rubin, 1987) may not be valid for this data example, but the

patterns of missing data and methodologies dealing with them are beyond the scope of this article. To accommodate missing counts in some of the months, the set M_i in (5) should contain only those months when subject i has been observed to have his/her scheduled clinic visits.

Hyperprior specification:

We have used zero mean and high variance (1000) for the regression parameters β_1 and β_2 to reflect our prior opinions of no covariate effects on the rate of nonfatal events as well as on the risk of termination. Because we do not have prior input from any expert, we are using diffuse priors for β_1 and β_2 and similar diffuse priors later for other parameters. We have noticed that our analysis is not sensitive to the choice of the values of the hyperparameters as long as the priors remain sufficiently diffuse.

For hyperparameters $\tau = 1/c$ and $\gamma_1, \dots, \gamma_J$ associated with the gamma process in (7), we take $\tau = 5.0$ and $\gamma_j = \gamma_0(a_j) - \gamma_0(a_{j-1}) = 0.01 \times \left(\sqrt{a_j} - \sqrt{a_{j-1}}\right)$. These choices imply that the prior guess for μ_j is $0.01 \times \left(\sqrt{a_j} - \sqrt{a_{j-1}}\right)$ and the precision of the prior guess is very low, $c = 0.2$ (or variability of the prior guess is 5.0). Our choice of the hyperparameters reflects our prior belief, although a very weak one, that the overall rate of superficial tumors is slightly higher than 0.01 per unit time. The selection of such a diffuse prior process allows the conclusion to be driven by the likelihood.

Table 1 presents the results of our data analysis. The posterior estimate of β_2 is 0.0427, indicating a slight increase in risk of drop-out due to cure (positive response). There is not much evidence for the treatment effect on termination time because the 95% credible interval for β_2 contains zero. There is, however, good evidence that thiotepa reduces the rate of superficial tumors as the 95% credible interval of β_1 is (0.538, 2.77). Please note that the posterior standard deviation (0.507) of β_1 is around 10% larger than the standard error of the corresponding estimate of Sun and Wei (2000). But, unlike the previous analysis, the standard deviation of the marginal posterior does take care of the additional uncertainty in the posterior estimate of a parameter due to the presence of the rest of the parameters of the model.

The posterior mean of a is negative, indicating some evidence of negative association between $\{N(t)\}$ and T. So, the patients who are more prone to superficial tumors are likely to be terminated later (longer follow-up times). But, the evidence for negative a is not strong as the 95% credible interval for a contains zero. There is also strong evidence that the baseline hazard rate for the termination is increasing over time (that is, ν is positive). The results of our Bayesian analysis are in general agreement with the findings of Sun and Wei (2000).

Model validation:

Model validation diagnostics are very important in Bayesian survival analysis (Sinha and Dey, 1997; Ibrahim et al., 2001). We developed a method for model validation for panel-count data via a Bayesian cross-validated residual called the conditional predictive ordinate

(CPO). The cross-validated posterior predictive probability evaluated at the observed data from patient i is given as

$$CPO_{(i)} = P\big[\mathbf{N}_i = \mathbf{N}_{i,\,obs}, T_i = t_i \mid D_{(-i)}\big]$$
$$= \Big[E\big\{P(\mathbf{N}_i = \mathbf{N}_{i,\,obs}, T_i = t_i \mid \theta)^{-1} \mid D\big\}\Big]^{-1}, \qquad (10)$$

where $D_{(-i)}$ is the observed data with the patient i removed, $\theta = (\mu, \beta_1, \beta_2, \eta, a, \lambda, \nu)$ is the set of parameters of the model, and $\mathbf{N}_i = (N_{i1}, \ldots, N_{iT_i})$. The final expression of the CPO in (10) follows from a result of Gelfand, Dey, and Chang (1992). This expression actually enables us to use the samples from the posterior $p(\theta \mid D)$ to compute the CPO using the following Monte Carlo integration algorithm:

Step 1: Simulate from $\{\theta_l\}_{l=1}^{M}$ from $p(\theta \mid D)$ MCMC;

Step 2: For each θ_l, simulate $\{w_{ib}\}_{b=1}^{B}$ from $p(w_i \mid \theta_l, D)$ for $i = 1, \ldots, n$;

Step 3: For $l = 1, \ldots, M$, compute $P(N_i = N_{i,obs}, T_i = t_i \mid \theta_l)$ using $\frac{1}{B}\sum_{b=1}^{B} P(N_i = N_{i,\,obs} \mid \theta, w_{ib}) P(T = t_i \mid \theta_i, x_i, w_i)$;

Step 4: Finally, compute $CPO_{(i)} = \Big[\{1/M\} \sum_{i=1}^{M} \big\{1/P(N_i = N_{i,\,obs}, T_i = t_i \mid \theta_l)\big\}\Big]^{-1}$.

As they are probabilities, the CPOs lie between 0 and 1 and a higher value of the CPO implies a better agreement between the model and the observation. Typically, CPOs are plotted against their covariate values to check whether the CPO values have any relationship with the covariates. Figure 1 presents the plot of the CPO of each subject versus the corresponding termination time. Figure 1 also uses separate symbols for 47 subjects corresponding to the placebo group and the rest corresponding to the treatment group. From this plot, there appears to be neither evidence of any possible outlier nor evidence of a treatment-dependent pattern. The pattern of CPOs does not appear to depend on the termination times. This plot suggests that our model fits the data adequately.

Prediction:

Please note that one of the major advantages of a Bayesian analysis is that one can compute the predictive distributions of nonfatal events and termination time of a future patient given the values of his/her covariate x. The steps for simulating termination time T and number of events until termination $N(T)$ of a future patient are very similar to the computation of CPO in the last paragraph:

Step 1: Simulate $\{\theta_l\}_{l=1}^{M}$ from $p(\theta \mid D)$ using MCMC;

Step 2: For each θ_l, simulate w_l from $Ga(w \mid \eta_l)$;

Step 3: For $l = 1, \ldots, M$, simulate T_l using $P\big[T_l = a_j \mid x, w_l, \theta_l\big] = 1 - \pi_{jl}^{\exp(\beta_{2l}x)w_l}$;

Step 4: Finally, simulate $N(T_l)$ from $\mathrm{Poi}\{\mu_0(T_l)\exp(\beta_{1l}x)\,w_l x\}$.

These $\{T_1, N(T_1)\}, \ldots, \{T_M, N(T_M)\}$ are MCMC samples from the bivariate predictive distribution of $\{T, N(T)\}$ of a future patient with covariate value x. Now these MCMC

samples can be used to compute different quantities (e.g., mean, median, standard deviation, prediction interval, etc.) related to the predictive distributions of T and $N(T)$.

5. Remarks

5.1 Censoring of the Termination

In many data examples encountered in practice, the termination time T_i may be subjected to right-censoring. Typically, the censoring time C_i is assumed independent of T_i and we further assume that risk of C_i does not depend on events history. To allow possible right-censoring, we need a straight-forward modification of L_2 in (6) as

$$L_2(\beta_2, \pi, \mathbf{w} \mid D)$$
$$\propto \prod_{i=1}^{n} \left[1 - \pi_{(i)}^{w_i^{\alpha} \exp(\beta_2 x_i)} \right]^{\delta_i} \pi_{(i)}^{(1-\delta_i) w_i^{\alpha} \exp(\beta_2 x_i)} S_{(i)}^{\exp(\beta_2 x_i) w_i^{\alpha}}, \tag{11}$$

where the observed $t_i = a_{(i)} \in \{a_1, \dots, a_J\}$ is the minimum of termination time T_i and censoring time C_i, and $\delta_i = 1_{[t_i = T_i]}$ is the indicator of censoring of termination for subject i.

Our methodology can accommodate this case with this minor modification of the likelihood and without any major modification of the MCMC tool.

A potentially serious censoring problem may arise when T_i is actually continuous, but only observed as interval censored within the grid intervals $a_1 < a_2 < \cdots < a_J$. Treating T_i as discrete when T_i is in reality only interval censored may give us biased inference, and the level and direction of bias in the inference are, at this point, unknown.

5.2 Comparison to LI Model

The LI model (Lancaster and Intrator, 1998), developed in the context of modeling the joint distribution of hospitalization rate and death in AIDS patients, assumes that the counting process $N(t)$ (hospitalizations) occurs as a compound Poisson process and each hospitalization has a fixed (but unknown) probability of causing the termination (death). Due to the structure of the LI model, only positive association between rate of $N(\cdot)$ and risk of T is allowed. The assumption of discrete T is not very credible when the termination event is death. In Lancaster and Intrator (1998), T was assumed to be equal to one of $a_1, \dots, a_J$. For a patient with $T \in I_j = (a_{j-1}, a_j]$, the increment $N_j = N(a_j) - N(a_{j-1})$ is informative about $T - a_{j-1}$ (length of interval-censored part of T) and ignoring the difference $a_j - T$ in the likelihood may bias the inference. Neither the LI model nor our model can describe accurately a situation with continuous termination time, as we need to use an approximation where the termination time is interval censored and the number of hospitalizations in the last observation interval can be informative about the exact termination time within the interval. For many data examples including the VACURG data, the termination event is not necessarily a random consequence of a nonfatal event, and there are cases when a history of higher rates of nonfatal events implies an expected delay in the favorable termination event such as the cure or discharge from hospital. The LI cannot accommodate this kind of negative relationship between recurrent events history and risk of termination.

Acknowledgements

Appendix

Proof of Theorem 1.

The conditional discrete hazard of the termination event given the nonfatal events history $\mathcal{N}(a_j)$ is given by

$$P\left[T = a_j \mid \mathcal{N}(a_j), x, T > a_{j-1}\right] = \frac{P\left[T = a_j, \mathcal{N}(a_j) \mid x\right]}{P\left[T > a_{j-1}, \mathcal{N}(a_j) \mid x\right]}$$

$$= \frac{\int S_{j-1}^{w^\alpha \exp(\beta_2 x)}\left[1 - \pi_j^{w^\alpha \exp(\beta_2 x)}\right] w^{N(a_j)} \exp\left\{-w\mu_0(a_j)e^{\beta_1 x}\right\} g(w \mid \eta)dw}{\int S_{j-1}^{w^\alpha \exp(\beta_2 x)} w^{N(a_j)} \exp\left\{-we^{\beta_1 x}\mu_0(a_j)\right\} g(w \mid \eta)dw},$$

where $g(w \mid \eta)$ is the $Ga(\eta, \eta)$ density of the frailty w. The above integral expressions are derived from the sampling distribution of $\mathcal{N}(a_j)$ in (1) and the conditional hazard in (2). The rest of the proof follows from here.

Proof of Theorem 2.

Let us define the vector $(N_1, \ldots, N_j)$—the vector of number of nonfatal events to a subject in first j intervals. Also, note that $N(a_{j-1}) = N_1 + \cdots + N_{j-1}$. Then,

$$P\left[N_j = m \mid \mathcal{N}(a_{j-1}), T > a_{j-1}, x\right]$$
$$= \frac{P\left[N_j = m, T > a_{j-1}, \mathcal{N}(a_{j-1}) \mid x\right]}{P\left[T > a_{j-1}, \mathcal{N}(a_{j-1}) \mid x\right]}$$
$$\propto \int_0^\infty P\left[N_j = m, \mathcal{N}(a_{j-1}) = (N_1, \ldots, N_{j-1}) \mid w\right] P\left(T > a_{j-1} \mid w\right) g(w \mid \eta)dw$$
$$\propto \int_0^\infty S_{j-1}^{w^\alpha \exp(\beta_2 x)} \frac{w^{N(a_{j-1})+m}}{m!} \times \exp\left\{-w\mu_0(a_j)e^{\beta_1 x}\right\} g(w \mid \eta)dw.$$

Note that the above probability needs to be evaluated only up to the normalizing constant as a function of m. The rest of the proof follows from here.

References

Best NG, Cowles MK, and Vines SK (1995). CODA Manual Version 0.30, MRC Biostatistics Unit, Cambridge, U.K.

Chen M-H, Ibrahim JG, and Shao Q-M (2000). Monte-Carlo Methods in Bayesian Computation. New York: Springer-Verlag.

Cox DR (1972). Regression models and life tables. Journal of the Royal Statistical Society, Series B 34, 187–220.

Dunson DB (2000). Models for papilloma multiplicity and regression: Application to transgenic mouse studies. Applied Statistics 49, 19–30.

Gail MH, Santner TJ, and Brown CC (1980). An analysis of comparative carcinogenesis experiments with multiple times to tumor. Biometrics 36, 255–266. [PubMed: 7407314]

Gelfand AE, Dey DK, and Chang H (1992). Model determination using predictive distributions with implementation via sampling-based methods In Bayesian Statistics, Volume 4, Bernardo JM, Berger JO, Dawid AP, and Smith AFM (eds), 147–167. Oxford: Oxford University Press.

Gelman A and Rubin DB (1992). Inference from iterative simulation using multiple sequences. Statistical Science 7, 457–741.

Ghosh D and Lin DY (2000). Nonparametric analysis of recurrent events and death. Biometrics 56, 554–562. [PubMed: 10877316]

Gilks WR and Wild P (1992). Adaptive rejection sampling for gibbs sampling. Applied Statistics 41, 337–348.

Ibrahim JG, Chen M-H, and Sinha D (2001). Bayesian Survival Analysis. New York: Springer-Verlag.

Kalbfleisch JD (1978). Nonparametric Bayesian analysis of survival time data. Journal of the Royal Statistical Society, Series B 40, 214–221.

Karlin S and Taylor HM (1981). A First Course in Stochastic Processes. London: Academic Press.

Lancaster T and Intrator O (1998). Panel data with survival: Hospitalization of HIV-positive patients. Journal of the American Statistical Association 93, 46–51.

Little Roderick JA and Rubin DB (1987). Statistical Analysis with Missing Data. New York: Wiley.

Oakes D (1982). Frailty models for multiple event times In Survival Analysis: State of the Art, Klein JP and Goel PK (eds), 371–379. Netherlands: Kluwer Academic Publisher.

Oakes D and Cui L (1994). On semiparametric inference for modulated renewal processes. Biometrika 81, 83–90.

Scheike TH and Zhang M-J (1998). Cumulative regression function tests for regression models for longitudinal data. The Annals of Statistics 26, 1328–1355.

Sinha D (1993). Semiparametric Bayesian analysis of multiple event time data. Journal of the American Statistical Association 88, 979–983.

Sinha D and Dey DK (1997). Semiparametric Bayesian analysis of survival data. Journal of the American Statistical Association 92, 1195–1212.

Sun J and Wei LJ (2000). Regression analysis of panel-count data with covariate-dependent observation and censoring times. Journal of the Royal Statistical Society, Series B 62, 293–302.

Wang M-C, Qin J, and Chang C-T (2001). Analyzing recurrent event data with informative censoring. Journal of the American Statistical Association 96, 1057–1065.
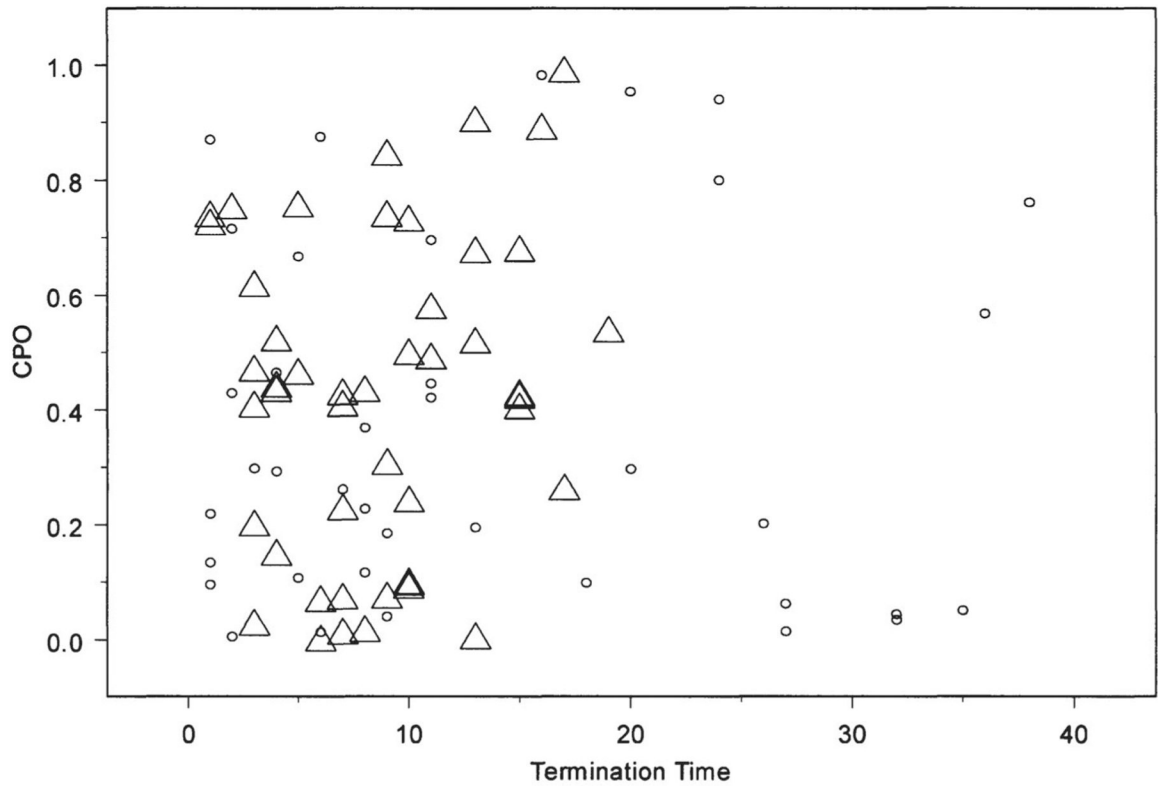
Figure 1.
Plot of termination time versus CPO for each patient: ○ for treatment and △ for placebo.

Table 1

Summary of posterior analysis

Parameter	Mean	SE	95% Credible interval
β_1	−1.491	0.507	(−0.538, −2.77)
β_2	0.0427	0.229	(−0.401, 0.580)
a	−0.0137	0.0295	(−0.081, 0.037)
η	0.1313	0.0298	(0.0817, 0.1968)
ν	2.192	0.1849	(1.778, 2.546)
λ	0.0006	0.0005	(0.0001, 0.0019)