

Bayesian Methods for Assessing Ordering in Hazard Functions

Laura H. Gunn^{1,*} and David B. Dunson²

¹ Institute of Statistics and Decision Sciences
Duke University, Durham, NC 27708-0251, U.S.A.

²Biostatistics Branch,
National Institute of Environmental Health Sciences
MD A3-03, P.O. Box 12233
Research Triangle Park, NC 27709, U.S.A.

* laura@stat.duke.edu

SUMMARY. In biomedical studies that collect event time data, it is often appropriate to assume non-decreasing hazards across dose groups, though dose effects may vary with time. Motivated by this application, we propose a Bayesian approach for order restricted inference using an additive hazard model with time-varying coefficients. In order to make inferences on equalities versus increases in hazard functions, a prior is chosen for the time-varying coefficients that assigns positive probability to no dose effect while restricting coefficients to be non-negative. By using a high dimensional piecewise constant model and smoothing the functions by coupling Markov beta and gamma processes, we obtain a flexible and computationally tractable approach for identifying sets of dose and age values at which hazards are increased. This approach can also be used to estimate dose response and survival curves. The methods are illustrated through application to data from a toxicology study.

KEY WORDS: Additive hazards; Bayesian hypothesis testing; Bioassay; Gibbs sampler; Markov prior process; Order restricted inference; Point mass; Survival analysis.

1. Introduction

By identifying regions of age and dose where an effect on the hazard exists in studies involving rodents, researchers obtain valuable information regarding the potential adverse effects of a chemical in humans. For example, Figure 1 presents Kaplan-Meier survival curves from a National Toxicology Program (NTP) study of ethylbenzene (NTP Technical Report, 1999). It is of interest to assess whether survival decreases with dose, while also identifying thresholds, such as the first dose and age at which an increase in hazard occurs. Because we are interested in decreased survival in response to a potentially adverse exposure, it is natural to consider a one-sided analysis that constrains the hazard functions to be non-decreasing in dose. In addition, it would be appealing to obtain smooth estimates of dose group-specific survival curves.

In a frequentist setting, stochastic ordering methods for survival functions have been intro-

duced (Dykstra, 1982; Feltz and Dykstra, 1985; Dykstra and Feltz, 1989; Dykstra, Kochar and Robertson, 1991; 1995). In addition, Arjas and Gasbarra (1996) and Gelfand and Kottas (2001) proposed nonparametric Bayesian methods for accommodating stochastic ordering in two survival functions. Furthermore, Hoff (2003) proposed an innovative Bayesian approach for nonparametric estimation of distributions subject to stochastic ordering. These methods address the issue of estimation subject to strict order constraints, however they are limited in their usefulness for assessing evidence of increasing versus constant hazards. Although several frequentist order-restricted tests are available for proportional hazards models (Sen, 1984; Silvapulle, 1994; Silvapulle and Silvapulle, 1995; Singh and Wright, 1996; 1998), there has been limited consideration of non-proportional hazards cases.

Our primary interest is in identifying those dose levels and ages at which there is an increase in hazard. Equalities in hazard functions for different dose groups effectively correspond to multiple age-specific null hypotheses. Following previous authors (cf., Dunson and Herring, 2003; Gönen, Westfall, and Johnson, 2003 for recent references), we treat null hypotheses as if they could be exactly true (refer to Berger and Delampady, 1987, for a discussion and justification of this viewpoint). Because the null hypotheses correspond to different times, it is natural to assume that hypotheses located “close” to each other are highly correlated *a priori*, resulting in smoothing of the model probabilities. Related ideas have been considered by Gönen et al. (2003) and Neelon and Dunson (2004) in different settings.

In addition to smoothing of the hypothesis probabilities, it is appealing to smooth the baseline hazard function and time-varying coefficients. There is a rich Bayesian literature on defining priors for hazard functions (for a recent review refer to Ibrahim, Chen, and Sinha, 2001). Most Bayesian analyses have relied on a gamma process prior for the cumulative baseline hazard (Kalbfleisch, 1978) or on independent gamma priors for interval-specific hazards (Walker and Mallick, 1997), though several authors have proposed Markov-type priors (Gamerman, 1991; Arjas and Gasbarra, 1994; Gray, 1994; Sinha and Dey, 1998). The Markov gamma and beta processes proposed by Nieto-Barajas and Walker (2002) have

the practical advantage that conjugacy properties are maintained through the use of data augmentation. Such conjugacy properties are particularly important when interest focuses on model or hypothesis selection, and it is necessary to calculate integrating constants in deriving posterior model probabilities.

In order to smooth time-specific hypothesis probabilities and coefficients subject to the non-decreasing hazards constraint, we propose a coupled Markov beta-gamma process prior, which generalizes the structure of Nieto-Barajas and Walker (2002). The beta process component is used to characterize changes with time in the probability that the hazard curves are constant (i.e., the null hypothesis holds), and the gamma process characterizes changes in a time-varying coefficient occurring while the alternative hypothesis holds. The prior has convenient conjugacy properties which facilitate posterior computation and inferences using a data augmentation Markov chain Monte Carlo (MCMC) algorithm. By allowing time-varying coefficients, this approach generalizes the method of Dunson and Herring (2003) to allow non-proportional hazards.

Section 2 considers ordered hazards and introduces the Markov beta-gamma prior. Section 3 describes the likelihood structure and approach to posterior computation. Section 4 applies the methods to simulation and toxicology examples, while Section 5 discusses results.

2. Modeling Ordered Hazard Curves

2.1 Model and Hypothesis Structure

Let $x_i \in \mathcal{X} = \{1, \dots, d\}$ be an ordered categorical predictor or group index for subject i ($i = 1, \dots, n$), and let $\lambda_h(t)$ denote the hazard at time $t \in \mathcal{T}$ for subjects with $x_i = h$. We assume that the hazard curves belong to a restricted functional space, Ω , in which $\lambda_1(t) \leq \lambda_2(t) \leq \dots \leq \lambda_d(t)$ for all $t \in \mathcal{T}$. The hazard for subject i at time t conditional on x_i can be expressed as:

$$\lambda(t; x_i = h) = \lambda_0(t) + \sum_{k=1}^{h-1} \gamma_k(t), \quad \text{for } h = 1, \dots, d, \quad (1)$$

where $\lambda_0(t) \geq 0$ is the baseline hazard function, and $\gamma_k(t) = \lambda_{k+1}(t) - \lambda_k(t)$ ($\gamma_k(t) \geq 0$) is the increase in the hazard function attributable to changing x_i from k to $k + 1$. Equalities in the hazard curves between groups k and $k + 1$ occur at times when $\gamma_k(t) = 0$.

The global null hypothesis of no difference in the hazard curves corresponds to $H_0 : \gamma_1(t) = \dots = \gamma_{d-1}(t) = 0$ for all $t \in \mathcal{T}$, while the global alternative is $H_1 : \gamma_k(t) \geq 0$ for $k = 1, \dots, d-1$ and all $t \in \mathcal{T}$, with $\gamma_k(t) > 0$ for at least one k, t . In addition to assessing overall evidence of any increase in the hazard (H_1), we want to identify specific subregions of \mathcal{X} and \mathcal{T} across which increases occur. Such subregions correspond to locations where $\gamma_k(t) > 0$.

Although we focus on the simple case in which the ordering of each group relative to all other groups is known, partial orderings can be accommodated using trivial modifications to the methodology. For example, an interesting case is that of simple tree ordering in which Ω denotes the space of hazard functions with $\lambda_1(t) \leq \lambda_k(t)$ for $k = 2, \dots, d$ and $t \in \mathcal{T}$. In this case, we can simply modify expression (1) so that $\lambda(t; x_i = h) = \lambda_0(t) + \gamma_{k-1}(t)$, where $\gamma_{k-1}(t) \geq 0$ denotes the difference in hazards between $x_i = 1$ and $x_i = k$. Note that the non-negative constraint on the $\gamma(t)$'s now only implies that groups $2, \dots, k$ (e.g., the treated groups) have hazards at least as big as group 1 (e.g., control group).

2.2 Markov Prior Specification

Under the specification of Section 2.1, a key step is choosing priors for the baseline hazard, $\lambda_0(t)$, and the increments on the hazard function between groups, $\gamma_1(t), \dots, \gamma_{d-1}(t)$, which have the necessary properties. In particular, each of these functions needs to be non-negative, and the $\gamma(t)$ functions must have a prior which allocates positive probability to regions equivalent to 0. In addition, it is appealing to have a prior which favors relatively smooth functions and penalizes large changes in time. To facilitate this specification, we approximate the hazard function using a piecewise constant model, with $\lambda_0(t) = \lambda_{0l}$ and $\gamma_k(t) = \gamma_{kl}$ for $t \in (\tau_{l-1}, \tau_l]$ and $t \leq \tau_L$, where $0 = \tau_0 < \tau_1 < \dots < \tau_L < \tau_{L+1} = \infty$ are pre-specified interval endpoints, which can be chosen to be arbitrarily close to one another. Piecewise constant

models are often chosen as a convenient but flexible specification in Bayesian survival models.

For the baseline hazard function, we use the Markov gamma specification of Nieto-Barajas and Walker (2002). In particular, letting $\boldsymbol{\lambda}_0 = (\lambda_{01}, \dots, \lambda_{0L})'$, we express the prior as follows:

$$\lambda_{00} \sim \mathcal{G}(a_\lambda, b_\lambda) \quad w_l | \lambda_{0l-1} \sim \text{Poisson}(c_\lambda \lambda_{0l-1}) \quad \lambda_{0l} | w_l \sim \mathcal{G}(a_\lambda + w_l, b_\lambda + c_\lambda), \quad (2)$$

where λ_{00} is a latent baseline term, λ_{0l} has mean $(a_\lambda + w_l)/(b_\lambda + c_\lambda)$ with a_λ , b_λ , and c_λ as fixed hyperparameters specified by the investigator, c_λ is a common smoothing parameter in which more smoothing occurs as c_λ becomes larger, and w_l is a latent process introducing dependence into the prior process. Hence, $\boldsymbol{\lambda}_0$ is calculated using the process $\lambda_{00} \rightarrow w_1 \rightarrow \lambda_{01} \rightarrow w_2 \rightarrow \dots$. The prior on $\boldsymbol{\lambda}_0$ extends the independent gamma process prior to incorporate dependence with a Markov relation for the baseline hazard. An advantage of using this prior is that it retains its conditionally conjugate form for convenient posterior computation.

The prior for $\gamma_k(t)$, $k = 1, \dots, d-1$, requires a different approach, since the Markov gamma specification does not allow regions across which $\gamma_k(t) = 0$. For this reason, we propose a novel specification based on coupling Markov gamma and beta processes. First, let $\pi_{kl} = \Pr(\gamma_{kl} = 0)$ denote the probability that there is no difference in the hazard functions for groups k and $k+1$ within the interval $(\tau_{l-1}, \tau_l]$, and let $\gamma_{kl} = (1 - \delta_{kl})\gamma_{kl}^*$ denote the actual difference in hazards within this interval. Following a common technique in the Bayesian variable selection literature (cf., George and McCulloch, 1997), we express the coefficient, γ_{kl} , as the product of an indicator variable, $\delta_{kl} = 1_{(\gamma_{kl}=0)}$ where $\delta_{kl} = \delta_k(t)$ for $t \in (\tau_{l-1}, \tau_l]$, and the potential value if the coefficient is non-zero, γ_{kl}^* .

We induce a prior on γ_{kl} by specifying priors for the point mass probabilities, π_{kl} , and the potential slopes given non-zero values, γ_{kl}^* . In particular, instead of assuming *a priori* independence in the point mass probabilities $\{\pi_{kl}\}$ (as in Hjort, 1990), we use a Markov beta process (Nieto-Barajas and Walker, 2002) to allow autocorrelation:

$$\pi_{k0} \sim \mathcal{B}e(a_{\pi k}, b_{\pi k}) \quad u_{kl} | \pi_{kl-1} \sim \mathcal{B}in(c_{\pi k}, \pi_{kl-1}) \quad \pi_{kl} | u_{kl} \sim \mathcal{B}e(a_{\pi k} + u_{kl}, b_{\pi k} + c_{\pi k} - u_{kl}), \quad (3)$$

where π_{k0} is a latent baseline term included to simplify the analysis, π_{kl} has mean $(a_{\pi k} + u_{kl})/(a_{\pi k} + b_{\pi k} + c_{\pi k})$ with $a_{\pi k}$, $b_{\pi k}$, and $c_{\pi k}$ fixed by the investigator, $c_{\pi k}$ is a smoothing parameter (note $c_{\pi k}$ must be an integer), and u_{kl} is the latent process incorporating dependence into the prior process.

The hyperparameters can be selected so that the probability of the global null hypothesis H_0 is approximately 0.5. This results in a Bayesian correction for multiple comparisons, which is less conservative than the Bonferroni-type correction obtained under *a priori* independence (Westfall, Johnson and Utts, 1997). Of course, setting $\Pr(H_0) = 0.5$ may be overly conservative and may result in too high a degree of shrinkage towards equalities among the groups, particularly when dimensionality is high. Sensitivity of inferences to hyperparameter specification is considered in Section 4 in the context of the data example.

It is common in the Bayesian variable selection literature to specify a beta prior for the probability of including a predictor in a regression model. In models involving time-varying coefficients, it is natural to allow autocorrelation in whether a predictor is included. For example, if there is a treatment effect at time t , then it is very likely that there is a treatment effect at time $t + \epsilon$ for small ϵ . The Markov beta process captures this dependency structure, allowing the correlation in whether a group difference occurs to be decreasing with time. Using the Nieto-Barajas and Walker (2002) structure, which they proposed in a very different context, results in computational simplifications. In particular, closed forms are available for conditional model probabilities, avoiding the need for approximations to the normalizing constants. Details are presented in Section 3.

We complete prior specification by defining a Markov gamma process for $\{\gamma_{kl}^*\}$:

$$\gamma_{k0}^* \sim \mathcal{G}(a_{\gamma^*k}, b_{\gamma^*k}) \quad z_{kl} | \gamma_{kl-1}^* \sim \text{Poisson}(c_{\gamma^*k} \gamma_{kl-1}^*) \quad \gamma_{kl}^* | z_{kl} \sim \mathcal{G}(a_{\gamma^*k} + z_{kl}, b_{\gamma^*k} + c_{\gamma^*k}), \quad (4)$$

where γ_{k0}^* is a latent baseline term, a_{γ^*k} and b_{γ^*k} are specified by the investigator, γ_{kl}^* has mean $(a_{\gamma^*k} + z_{kl})/(b_{\gamma^*k} + c_{\gamma^*k})$, c_{γ^*k} is a smoothing parameter specified by the investigator, and z_{kl} is the latent process introducing dependence into the prior process. By coupling the

Markov beta and gamma components, the prior for γ_{kl} has support on the space of non-decreasing hazard functions, Ω , while still allowing for flat regions over which increases in the level of a predictor have no effect on the hazard. The prior for γ_{kl} is directly induced in terms of δ_{kl} and γ_{kl}^* , as defined earlier in this section. As is the case for related priors used in variable selection, specifying γ_{kl} as a function of δ_{kl} and γ_{kl}^* greatly simplifies computation.

Integrating out γ_{kl}^* and δ_{kl} , the induced prior on γ_{kl} can be formally expressed as follows:

$$\gamma_{kl} | \pi_{kl}, z_{kl}, z_{kl+1} = 1_{\{\gamma_{kl}=0\}} \pi_{kl} + 1_{\{\gamma_{kl}>0\}} (1 - \pi_{kl}) \mathcal{G}(\gamma_{kl}; a_{kl}, b_{kl}), \quad (5)$$

where $a_{kl} = a_{\gamma^*k} + z_{kl} + z_{kl+1}$ and $b_{kl} = b_{\gamma^*k} + 2c_{\gamma^*k}$. We refer to density (5) with shorthand $\delta_0 - \mathcal{G}(\cdot; \pi, a, b)$, consisting of a mixture of a point mass at 0 (with probability π) and a $\mathcal{G}(\cdot; a, b)$ density. This prior has the important practical advantage of being conditionally-conjugate, so that closed forms are available for the conditional model probabilities. This property greatly facilitates posterior computation, and avoids the need for approximations to the normalizing constants.

2.3 Extension to Multiple Predictors

We can generalize the approach to allow for multiple covariates with different constraints on the regression coefficients. For example, we could extend model (1) to incorporate a vector of ordered categorical predictors, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$. Expression (1) would become:

$$\lambda(t; \mathbf{x}_i) = \lambda_0(t) + \sum_{h=1}^p \sum_{k=1}^{x_{ih}-1} \gamma_{hk}(t), \quad (6)$$

where $\gamma_{hk}(t)$ is the change in hazard at time t attributable to increasing x_{ih} from k to $k+1$, $\boldsymbol{\gamma}_h = (\gamma_{h1}, \dots, \gamma_{h,d_h-1})'$ for $h = 1, \dots, p$, d_h is the total number of categories of x_{ih} , and $\lambda(t; \mathbf{x}_i) \geq 0$. We complete a Bayesian model specification by assigning priors to $\boldsymbol{\lambda}_0$ and $\boldsymbol{\gamma}_k$ using the approach described in Section 2.2.

We can extend the approach further to account for continuous covariates. Let $\mathbf{w}_i = (w_{i1}, \dots, w_{iq})'$ be a vector of continuous predictors. As an extension to existing additive hazard models

with continuous predictors (Lin and Ying, 1994; Lin, Oakes, and Ying, 1998), we introduce:

$$\lambda(t; \mathbf{x}_i, \mathbf{w}_i) = \lambda_0(t) + \mathbf{w}'_i \boldsymbol{\beta} + \sum_{h=1}^p \sum_{k=1}^{x_{ih}-1} \gamma_{hk}(t), \quad (7)$$

where $\boldsymbol{\beta}$ are regression coefficients, and the remaining expressions are previously defined. For $\lambda(t; \mathbf{x}_i, \mathbf{w}_i)$ to be non-negative, we apply the constraint $\mathbf{w}'_i \boldsymbol{\beta} \geq 0$ into the prior specification. Following Lin and Ying (1994) and Lin, Oakes, and Ying (1998), we could incorporate a time-dependency on the covariates such that we obtain $\mathbf{w}_i(t)$ as a substitute for \mathbf{w}_i in model (7). We could also potentially include time-varying coefficients, $\boldsymbol{\beta}(t)$, to replace $\boldsymbol{\beta}$.

3. Bayesian Inference

3.1 Counting Process Likelihood and Data Augmentation

Following the counting process notation introduced by Andersen and Gill (1982), let $N_i(t)$, for subject i ($i = 1, \dots, n$), represent the process counting the failures occurring up to time t , while $dN_i(t)$ is a small increment of $N_i(t)$ over the interval $[t, t + dt)$. $N_i(t)$ and $dN_i(t)$ equal 1 if the event occurs in $[0, t]$ or $[t, t + dt)$, respectively, and 0 otherwise. Let $Y_i(t) = 1$ if subject i is at risk at time t , and $Y_i(t) = 0$ otherwise. For subject i in group k at time t , the intensity process for $N_i(t)$ under model (1) is:

$$\lambda_i(t) = Y_i(t) \left(\lambda_0(t) + \sum_{k=1}^{x_i-1} \gamma_k(t) \right). \quad (8)$$

Under noninformative censoring, the observed counting process likelihood is proportional to:

$$\prod_{i=1}^n \left(\prod_{t \geq 0} [Y_i(t) \{ \lambda_0(t) + \sum_{k=1}^{x_i-1} \gamma_k(t) \}]^{dN_i(t)} \right) \exp \left(- \int_{t \geq 0} Y_i(t) \{ \lambda_0(t) + \sum_{k=1}^{x_i-1} \gamma_k(t) \} dt \right), \quad (9)$$

following a Poisson form. The infinitesimal counting process increments, $dN_i(t)$, contribute to the likelihood just as with independent Poisson random variables with means $\lambda_i(t)dt$ over the interval $[t, t + dt)$, though $dN_i(t)$ is at most one for all i, t . Defining the model in this framework allows the intensity to be regarded as constant in that interval (Clayton, 1991).

For efficient posterior computation, we implement a data augmentation approach based on the consideration that the $dN_i(t)$ are independent Poisson random variables under the

likelihood expression (9):

$$dN_i(t) \sim \text{Poisson} \left(Y_i(t) \left\{ \lambda_0(t) + \sum_{k:\delta_k(t)=0}^{x_i-1} \gamma_k(t) \right\} \right) \quad \text{for all } i : Y_i(t) = 1, \quad (10)$$

where $\delta_k(t) = \delta_{kj}$ for $t \in (t_{j-1}, t_j]$. Using the additive form of the Poisson sum prevents us from obtaining an efficient computational approach, since full conditional posterior distributions are non-standard. To remedy this problem, we can augment the $dN_i(t)$ in terms of independent Poisson latent variables corresponding to each term of the Poisson mean in (10). We introduce latent variables $\mathbf{dN}_i(t) = \{dN_{i0}(t), dN_{ik}(t), \text{ for all } t : \delta_k(t) = 0\}$ such that, by integrating out these latent variables, expression (10) is equivalent to:

$$dN_i(t) = dN_{i0}(t) + \sum_{k:\delta_k(t)=0}^{x_i-1} dN_{ik}(t), \quad (11)$$

with $dN_{i0}(t) \sim \text{Poisson}(\lambda_0(t))$ and $dN_{ik}(t) \sim \text{Poisson}(\gamma_k(t))$.

Following previous methods (Clayton, 1994; Dunson and Herring, 2004), we define $\mathbf{t} = (t_1, \dots, t_J)'$ as the union of unique, ordered failure times in the data and the interval endpoints $\boldsymbol{\tau} = (\tau_1, \dots, \tau_L)'$. The observed counting process likelihood (9) can be expressed as a product across the time intervals, $(t_{j-1}, t_j]$, $j = 1, \dots, J$:

$$\prod_{i=1}^n \prod_{j=1}^J \left(\prod_{t \in (t_{j-1}, t_j]} [Y_i(t) \{ \lambda_0(t) + \sum_{k:\delta_{kj}=0}^{x_i-1} \gamma_k(t) \}]^{dN_i(t)} \right) \exp \left(- \int_{t \in (t_{j-1}, t_j]} Y_i(t) \{ \lambda_0(t) + \sum_{k:\delta_{kj}=0}^{x_i-1} \gamma_k(t) \} dt \right). \quad (12)$$

Let $dN_{ij} = 1$ if subject i fails at time t_j , and 0 otherwise. With the assumption that there is a small amount of risk accumulated in $(t_{j-1}, t_j]$,

$$\int_{t_{j-1}}^{t_j} Y_i(t) \{ \lambda_0(t) + \sum_{k:\delta_{kj}=0}^{x_i-1} \gamma_k(t) \} dt \approx 0 \quad \text{for all } i, j, \quad (13)$$

the likelihood contribution for a subject at risk in the j th interval is proportional to:

$$\left(d\Lambda_{0j} + \sum_{k:\delta_{kj}=0}^{x_i-1} d\Gamma_{jk} \right)^{dN_{ij}} \exp \left(- (d\Lambda_{0j} + \sum_{k:\delta_{kj}=0}^{x_i-1} d\Gamma_{jk}) \right),$$

where $d\Lambda_{0j} = \int_{t_{j-1}}^{t_j} \lambda_0(s) ds$ and $d\Gamma_{jk} = \int_{t_{j-1}}^{t_j} \gamma_k(s) ds$. The Poisson approximation to the likelihood is highly accurate under the assumption that a Poisson random variable with mean

(13) has a small chance of being greater than one. Assumption (13) is justified whether the interval $(t_{j-1}, t_j]$ is narrow or wide. Since at most one event will occur in this interval, by definition, then the hazard will be extremely small, even if the interval is large.

Let $Y_{ij} = 1$ if subject i is at risk at time t_j , and 0 otherwise. Using the approach in (11), dN_{ij} is expressed as:

$$dN_{ij} = dN_{ij0} + \sum_{k:\delta_{kj}=0}^{x_i-1} dN_{ijk}, \quad \text{for all } i, j \text{ and } Y_{ij} = 1, \quad (14)$$

with the latent variables having independent Poisson distributions:

$$\Pr(dN_{ij0}) = \text{Poisson}(dN_{ij0}; d\Lambda_{0j}) \quad \text{and} \quad \Pr(dN_{ijk}) = \text{Poisson}(dN_{ijk}; d\Gamma_{jk}). \quad (15)$$

The increment dN_{ij} contributes to the likelihood in a similar manner as $dN_i(t)$, so we replace expression (9) with the augmented data likelihood, which is proportional to:

$$\prod_{i=1}^n \prod_{j:Y_{ij}=1}^J 1 \left(dN_{ij} = dN_{ij0} + \sum_{k:\delta_{kj}=0}^{x_i-1} dN_{ijk} \right) \frac{d\Lambda_{0j}^{dN_{ij0}} \exp(-d\Lambda_{0j})}{dN_{ij0}!} \left\{ \prod_{k:\delta_{kj}=0}^{x_i-1} \frac{d\Gamma_{jk}^{dN_{ijk}} \exp(-d\Gamma_{jk})}{dN_{ijk}!} \right\}. \quad (16)$$

We can now obtain conditionally conjugate posteriors for the MCMC sampling algorithm.

3.2 Conditional Posterior Distributions

The necessary conditional distributions are derived in this Section and in the Appendix. In particular, we derive conditional posterior distributions for λ_0 and γ_k using the augmented data likelihood (16) and priors (2) and (5), respectively. The conditional posterior for λ_0 is:

$$\mathcal{G} \left(\lambda_{0l}; a_\lambda + w_l + w_{l+1} + \sum_{i=1}^n \sum_{j:Y_{ij}=1:t_j \in (\tau_{l-1}, \tau_l]} dN_{ij0}, b_\lambda + 2c_\lambda + \sum_{i=1}^n \sum_{j:t_j \in (\tau_{l-1}, \tau_l]} Y_{ij}(t_j - t_{j-1}) \right). \quad (17)$$

A derivation for expression (17) is found in the Appendix. We also show in the Appendix that the conditional posterior for γ_k is:

$$\Pr(\gamma_{kl} | \gamma_{(-kl)}, \lambda_0, data) = \delta_0 - \mathcal{G} \left(\gamma_{kl}; \tilde{\pi}_{kl}, \tilde{a}_{kl}, \tilde{b}_{kl} \right), \quad (18)$$

where $\boldsymbol{\gamma}_{(-kl)}$ is the vector excluding the kl th element of $\boldsymbol{\gamma}$,

$$\tilde{\pi}_{kl} = \frac{\pi_{kl}}{\pi_{kl} + (1 - \pi_{kl}) \frac{C(\boldsymbol{\gamma}_{kl}; a_{kl}, b_{kl})}{C(\boldsymbol{\gamma}_{kl}; \tilde{a}_{kl}, \tilde{b}_{kl})}} \quad (19)$$

is the conditional posterior probability of $\gamma_{kl} = 0$, $C(\cdot; a, b)$ is the constant in the gamma density ($b^a/\Gamma(a)$),

$$\tilde{a}_{kl} = a_{kl} + \sum_{i,j: Y_{ij}=1, t_j \in (\tau_{l-1}, \tau_l]} 1_{\{x_i > k\}} dN_{ijk} \quad \text{and} \quad \tilde{b}_{kl} = b_{kl} + \sum_{i,j: t_j \in (\tau_{l-1}, \tau_l]} 1_{\{x_i > k\}} Y_{ij} (t_j - t_{j-1}). \quad (20)$$

The conditional posterior density for γ_{kl} follows the same $\delta_0 - \mathcal{G}$ form as prior (5), resulting in a conditionally conjugate structure. It is trivial to sample directly from (18) by setting $\gamma_{kl} = 0$ with probability $\tilde{\pi}_{kl}$, and sampling γ_{kl} from $\mathcal{G}(\boldsymbol{\gamma}_{kl}; \tilde{a}_{kl}, \tilde{b}_{kl})$ otherwise.

Samples of parameter values are easily obtained using a Gibbs sampling algorithm. Our MCMC algorithm proceeds by alternating among the following steps:

Step 1: Sample the latent Poisson variables $\{dN_{ij0}, dN_{ij1}, \dots, dN_{ij,d-1}\}$ from their full conditional posterior distribution in (21) if $dN_{ij} = 1$. If $dN_{ij} = 0$, then set $dN_{ij0} = dN_{ijk} = 0$ for all $k = 1, \dots, d-1$.

Step 2: Update the elements of $\boldsymbol{\lambda}_0$ by alternating between sampling the latent process variables $\{w_l\}$ from their full conditional (22) and $\boldsymbol{\lambda}_0$ from their gamma full conditional distributions (17), for $l = 1, \dots, L$.

Step 3: Update the elements of π_{kl} from their beta full conditional distribution (24) and $\{u_{kl}\}$ from their full conditionals (23), for $l = 1, \dots, L$ and $k = 1, \dots, d-1$.

Step 4: Sample γ_{kl}^* from their gamma full conditional distribution (26) and $\{z_{kl}\}$ from their full conditionals (25), for $l = 1, \dots, L$ and $k = 1, \dots, d-1$.

Step 5: Calculate $\tilde{\pi}_{kl}$ using (19), for $l = 1, \dots, L$ and $k = 1, \dots, d-1$.

Step 6: Sample δ_{kl} from $\mathcal{Bern}(\tilde{\pi}_{kl})$, for $l = 1, \dots, L$ and $k = 1, \dots, d-1$:

If $\delta_{kl} = 1$, then let $\gamma_{kl} = 0$.

If $\delta_{kl} = 0$, then let $\gamma_{kl} = \gamma_{kl}^*$.

Under mild regularity conditions, draws from the above algorithm converge to the joint posterior distribution. We illustrate how samples from the posterior distribution can be used

for inferences on differences between groups through application to simulated and real data in the next Section.

4. Application

4.1 Data Structure and Background

We illustrate the approach of Sections 2 and 3 through application to data from a toxicology study of ethylbenzene (NTP, 1999), a chemical widely used in manufacturing such products as rubber, plastic, gasoline, insecticide sprays, paints, and polyester fibers. Ethylbenzene belongs to a class of chemicals recognized as possible human carcinogens, and determining safe doses for worker exposure is an important concern. Our interest focuses on assessing evidence of increasing mortality with dose, focusing on male rats for illustrative purposes.

Two hundred male rats were exposed to ethylbenzene particles through inhalation at concentrations of 0 (control), 75, 250, and 750 parts per million (ppm) for six hours per day, 5 days per week. Rats were assigned at random to treatment and control groups, resulting in 50 rats per group. For ethical reasons, very ill rats were sacrificed to ease suffering. We group these moribund sacrifices with natural deaths in the analysis, so that the event of interest is defined as the time to critical illness or death. Animals surviving through the end of the study at two years were killed in a terminal sacrifice, which provides an ignorable censoring mechanism. Table 1 provides summaries of the numbers of rats dying from each cause.

4.2 Model and Prior Specification

Letting $x_i \in \{1, 2, 3, 4\}$ be a dose group indicator, ranging from $x_i = 1$ for control to $x_i = 4$ for the 750 ppm group, the hazard of death at age t can be expressed as in equation (2):

$$\lambda(t; x_i = h) = \lambda_0(t) + \sum_{k=1}^{h-1} \gamma_k(t), \quad \text{for } h = 1, \dots, 4,$$

where $\lambda_0(t)$ is the hazard at age t for control rats, and $\gamma_k(t)$ is the increase in hazard attributable to increasing the dose from the level in group k to the level in group $k + 1$.

In order to place more knots at those ages when changes in the hazard of death are more likely and to avoid having many knots in regions across which few events occur, we use unique event time data from an historical study’s control group to define the τ ’s. To remain consistent with the type, gender, and diet of rodent used in the ethylbenzene data, we chose an inhalation historical study on 2-Butoxyethanol (Ethylene Glycol Monobutyl Ether) for male rats under the same diet.

We used this historical control group in specifying hyperparameters for $\lambda_{0,0}$ within the Markov gamma process. Hazards were calculated from survival probabilities of each corresponding time interval in the historical control. Hazards were then averaged over time intervals and used as the mean for the initial element of the baseline hazard, $\lambda_{0,0}$. The variance of the historical control hazards was inflated by a factor of six to allow for study differences in the ethylbenzene example. As a result, $\lambda_{0,0}$ is assigned a gamma density with mean 0.09 and variance 0.50. Furthermore, we fix the smoothing parameter, c_λ , at 50 to allow for correlation in the baseline hazard across time intervals, l . Although we fix c_λ , we later discuss its robustness to different choices of values.

In specifying the Markov beta-gamma density for each γ_{kl} , we assign a prior to initial elements of the beta and gamma processes for π_{kl} and γ_{kl}^* , respectively. The hyperparameters on π_k give the prior guess at no dose effect when increasing dose from group k to $k + 1$. The prior on the initial point mass is specified by choosing hyperparameters such that the prior probability of homogeneity in hazards across levels of dose and time is approximately 0.50. In other words, we want the prior probability of a trend to reflect the fact that we do not have any strong prior belief of an adverse survival effect. Similarly, the initial $\gamma_{k,0}^*$ is assigned a gamma density with mean 0.50 and variance 1.0. Another element to the Markov beta-gamma process prior involves the smoothing parameters for each component process. By including autocorrelation we are smoothing the functions, borrowing information across neighboring intervals. To allow moderate to high autocorrelation in π_k and γ_k , we fix c_{π_k} and c_{γ^*k} , respectively, at 50. We later show robustness to choice of c_{π_k} and c_{γ^*k} through a

sensitivity analysis.

4.3 *Simulation Study*

We first ran the analysis on simulated data similar in structure to historical control data. Using the mean (660 days) and variance (6621.41) of the historical control survival times, we simulated survival times for 200 subjects, with 50 subjects in each of four groups, from a gamma distribution. Values greater than the 2-year mark were given a value of 731 days and were considered censored, as with terminal sacrifices in the data example. One of our goals in the simulation study is to create an example under the null hypothesis.

We implemented the data augmentation sampling algorithm outlined in Section 3.2. A burn-in period of 10,000 iterations was used, with an additional 50,000 iterations obtained to collect every 25th sample for posterior summaries. The burn-in was sufficient for apparent convergence based on diagnostic plots, and the level of autocorrelation in the samples tended to be low to moderate. As a result, a collection interval of 50,000 was judged sufficient to limit Monte Carlo error in estimating posterior summaries.

Posterior means and pointwise 95% credible intervals for the survival curves are plotted in Figure 2 along with the Kaplan-Meier estimates. The close correspondence between the posterior means and the Kaplan-Meier estimates suggests that our estimates provide a good fit to the data, with minimal bias introduced by the prior structure. However, the posterior mean is smoother than the Kaplan-Meier estimate, due to the Markov prior structure, and provides a more realistic estimate. In addition, the 95% credible intervals are substantially narrower than 95% frequentist confidence intervals, with an average reduction of 0.2 in the interval widths. This result possibly reflects improved efficiency attributable to the order restriction and smoothing prior. Note that the estimated survival curves are essentially identical for the different groups even though we are estimating the curves under a non-decreasing constraint. This is certainly due to the prior structure, which allows equalities with positive prior probability.

Table 2 lists prior and posterior probabilities of an overall dose effect as well as Bayes factors. Results confirm little evidence of any dose effect or trend. The top left panel in Figure 3 portrays posterior probabilities of a biologically-important increase in mortality with increasing dose and time relative to the control. Using a reasonable cutoff of 1% to indicate biological significance (as opposed to statistical significance), we estimate the age- and dose- specific posterior probabilities of increases of at least this magnitude. By choosing a 1% threshold, we combine those samples for which the difference in mortality relative to the control is greater than 0% but less than 1% in the same group as samples with a 0% difference. As expected, since we simulated under the null hypothesis, there is no evidence of an increase in mortality in any of the dose groups relative to control. These results parallel those from a frequentist log-rank test ($p=0.6$). However, our approach provides more realistic estimates through the smoothing and borrowing of information across levels of dose and time. The results suggest that our MCMC algorithm is specified appropriately, and that the prior does not lead to incorrect conclusions of differences when they do not occur.

4.4 Toxicology Study Results

We analyzed the real data using the same approach used for the simulated data, and again observed good rates of convergence and mixing. Posterior estimates of survival probabilities across levels of time and dose are presented in Figure 4. Although pointwise 95% credible intervals are not shown, the average reduction in the interval widths from the frequentist 95% confidence intervals is 0.18, similar to the 0.2 reduction under the simulation study. Figure 4 suggests that survival is similar in the 0, 75, and 250 ppm dose groups, but there may be a decrease in survival in the 750 ppm dose group. Our analysis shows clear evidence of an increase in mortality for rats in the 750 ppm dose group, which parallels results showing a significant change in survival with dose based on a frequentist log-rank test ($p=0.001$). Also notice how the curve for dose 3 diverges from the other groups shortly after day 500, corresponding to the age at which the Kaplan-Meier estimates begin to diverge (refer to Figure 1). Our estimates reveal an apparent lower effect for dose 3, with the survival probabilities

for those in the control and doses 1 and 2 pooled close to one another such that their curves are essentially identical.

The top right panel of Figure 3 plots posterior probabilities of a 1% increase in mortality for each dose group relative to control. The posterior probabilities are low for each dose group prior to one year of age when mortality is rare, but then increase dramatically for the high dose group at 521 days of age. This age corresponds to the time at which the Kaplan-Meier curve for dose group 3 starts to diverge from the other curves. By the end of the study at 2 years, there is strong evidence of higher mortality in the high dose group, as evidenced by high posterior probabilities of an overall increase in mortality (shown in Table 2). Table 2 also presents prior probabilities and Bayes factors. Since our prior is chosen to assign equal probability to the global null and alternative hypotheses, posterior samples of the hazard curves tend to congregate together except at locations of time and dose at which there is evidence in the data of a difference.

In addition to our primary analysis, we assessed whether our results are driven by our prior specification by conducting the analysis for lower and higher values of $c_{\pi k}$. Using values of 10 and 100, results did not significantly change. The estimated survival curves produced no change for different values of $c_{\pi k}$. The bottom two panels of Figure 3 show little variation between the posterior probabilities of a 1% increase in mortality across dose and time, compared to that of the primary analysis where $c_{\pi k} = 50$. Notice the difference in results in Table 2 and Figure 3 compared to that of the simulation study under the null hypothesis. Furthermore, we altered values of c_λ and c_{γ^*k} to 10 and 100 and saw no significant changes in posterior estimates of survival curves, posterior probabilities of a change in hazards, or posterior probabilities of a biologically-important increase in mortality. These results indicate that our model and inferences are generally robust to the prior specification.

5. Discussion

This article proposes a useful and easy-to-implement Bayesian approach for inferences on

ordering in hazard functions. Although we assume non-decreasing hazards with increasing dose, we account for flat regions of homogeneity in the hazards with increasing dose levels. In particular, our approach considers an alternative sub-hypothesis of non-decreasing hazards with increasing dose for subjects in consecutive dose groups compared to a local null of equality when subjects in dose group k are compared to group $k + 1$. The Markov beta-gamma structure in our model allows for positive probability to be allocated to such null hypotheses of no effect within specific regions. It is straightforward to make inferences on global hypotheses for hazards across all dose levels. We also include autocorrelation in the model to not only borrow information across neighboring time intervals but to smooth the curve. Furthermore, the conditionally conjugate structure yields efficient posterior computation. It is trivial to find posterior probabilities and Bayes factors of local null and alternative hypotheses.

We consider this approach under the framework of an ordered constraint on the categorical covariate representing the difference in hazards between subjects in dose group k compared to group $k + 1$, however there are numerous settings for which this methodology can be applied. In Section 2.3 we offer a generalization of our approach to the case with multiple ordinal predictors, as well as continuous covariates. In addition, we can extend the methods for inferences on dose response curves with downturns at higher doses where the peak location is unknown, or hazard functions with other shapes, including bathtub orderings.

ACKNOWLEDGMENTS

We thank Beth Brown for useful comments and discussions. We also thank Gregg Dinse and Dalene Stangl for carefully reviewing this article and offering helpful suggestions.

Appendix

Conditional Posterior Distributions & Derivations

Step 1: The full conditional distribution of $(dN_{ij0}, dN_{ij1}, \dots, dN_{ij,d-1})'$, for $i, j : Y_{ij} = 1$, is proportional to:

$$\text{Multinom}(dN_{ij}; P_{ij0}, P_{ijk}, k = 1, \dots, d-1), \quad (21)$$

where $P_{ij0} = \frac{d\Lambda_{0j}}{d\Lambda_{0j} + \sum_{k:\delta_{kj}=0} d\Gamma_{jk}}$ and $P_{ijk} = \frac{d\Gamma_{jk}}{d\Lambda_{0j} + \sum_{k:\delta_{kj}=0} d\Gamma_{jk}}$.

Step 2: The full conditional distribution of $\{w_l\}$, for $l = 1, \dots, L$ and $w_l = 0, 1, 2, \dots$, is proportional to:

$$\begin{aligned} & \frac{(c_\lambda \lambda_{0,l-1})^{w_l} (b_\lambda + c_\lambda)^{w_l}}{w_l! \Gamma(a_\lambda + w_l)} \lambda_{0l}^{w_l} \\ & \propto \frac{\{c_\lambda (b_\lambda + c_\lambda) \lambda_{0,l-1} \lambda_{0l}\}^{w_l}}{\Gamma(w_l + 1) \Gamma(a_\lambda + w_l)}. \end{aligned} \quad (22)$$

Furthermore, the gamma full conditional of λ_{0l} , for $l = 1, \dots, L$, is proportional to:

$$\begin{aligned} & \left\{ \prod_{i,j:Y_{ij}=1} d\Lambda_{0j}^{dN_{ij0}} \exp(-d\Lambda_{0j}) \right\} \lambda_{0l}^{a_\lambda + w_l + w_{l+1} - 1} \exp\{-\lambda_{0l}(b_\lambda + 2c_\lambda)\} \\ & \propto \left[\prod_{i,j:Y_{ij}=1} \left\{ \sum_{l=1}^L \lambda_{0l}(t_j - t_{j-1}) \mathbf{1}_{\{t_j \in (\tau_{l-1}, \tau_l]\}} \right\}^{dN_{ij0}} \exp\left\{ -\sum_{l=1}^L \lambda_{0l}(t_j - t_{j-1}) \mathbf{1}_{\{t_j \in (\tau_{l-1}, \tau_l]\}} \right\} \right] \\ & \quad \times \lambda_{0l}^{a_\lambda + w_l + w_{l+1} - 1} \exp\{-\lambda_{0l}(b_\lambda + 2c_\lambda)\} \\ & \propto \lambda_{0l}^{\sum_{i,j:Y_{ij}=1; t_j \in (\tau_{l-1}, \tau_l]} dN_{ij0}} \exp\left\{ -\lambda_{0l} \sum_{i,j: t_j \in (\tau_{l-1}, \tau_l]} Y_{ij}(t_j - t_{j-1}) \right\} \lambda_{0l}^{a_\lambda + w_l + w_{l+1} - 1} \\ & \quad \times \exp\{-\lambda_{0l}(b_\lambda + 2c_\lambda)\} \end{aligned}$$

The above expression can be expressed as posterior density (17).

Step 3: The full conditional distribution of $\{u_{kl}\}$, $l = 1, \dots, L$, $k = 1, \dots, d-1$, and $u_l = 0, \dots, c_{\pi k}$, is proportional to:

$$\begin{aligned} & \frac{\pi_{kl-1}^{u_{kl}} (1 - \pi_{kl-1})^{-u_{kl}}}{(c_{\pi k} - u_{kl})! u_{kl}!} \frac{\pi_{kl}^{u_{kl}} (1 - \pi_{kl})^{-u_{kl}}}{\Gamma(a_\pi + u_{kl}) \Gamma(b_\pi + c_{\pi k} - u_{kl})} \\ & \propto \frac{(\pi_{kl-1} \pi_{kl})^{u_{kl}} \{(1 - \pi_{kl-1})(1 - \pi_{kl})\}^{-u_{kl}}}{\Gamma(c_{\pi k} - u_{kl} + 1) \Gamma(u_{kl} + 1) \Gamma(a_\pi + u_{kl}) \Gamma(b_\pi + c_{\pi k} - u_{kl})}. \end{aligned} \quad (23)$$

In addition, the π_{kl} 's have a beta full conditional posterior distribution, for $l = 1, \dots, L$ and $k = 1, \dots, d-1$:

$$\pi_{kl}^{u_{kl}+1} (1 - \pi_{kl})^{c_{\pi k} - u_{kl} + 1} \pi_{kl}^{a_\pi + u_{kl} - 1} (1 - \pi_{kl})^{b_\pi + c_{\pi k} - u_{kl} - 1} \{ \mathbf{1}_{\{\gamma_{kl}=0\}} \pi_{kl} + \mathbf{1}_{\{\gamma_{kl}>0\}} (1 - \pi_{kl}) \}$$

$$\propto \mathcal{B}e\left(\pi_{kl}; a_\pi + u_{kl} + u_{kl+1} + 1_{\{\gamma_{kl}=0\}}, b_\pi + 2c_{\pi k} - u_{kl} - u_{kl+1} + 1_{\{\gamma_{kl}>0\}}\right). \quad (24)$$

Steps 4 & 6: The full conditional of $\{z_{kl}\}$, for $l = 1, \dots, L$ and $k = 1, \dots, d-1$, is proportional to:

$$\propto \frac{(c_{\gamma^* k} \gamma_{kl-1}^*)^{z_{kl}} (b_{\gamma^* k} + c_{\gamma^* k})^{z_{kl}} (\gamma_{kl}^*)^{z_{kl}}}{z_{kl}! \Gamma(a_{\gamma^* k} + z_{kl})} \frac{\{c_{\gamma^* k} (b_{\gamma^* k} + c_{\gamma^* k}) \gamma_{kl-1}^* \gamma_{kl}^*\}^{z_{kl}}}{\Gamma(z_{kl} + 1) \Gamma(a_{\gamma^* k} + z_{kl})}, \quad (25)$$

with the gamma full conditional distribution for γ_{kl}^* , for $k = 1, \dots, d-1$ and $l = 1, \dots, L$, as:

$$\begin{aligned} & (\gamma_{kl}^*)^{a_{\gamma^*} + z_{kl} - 1} \exp\{-\gamma_{kl}^* (b_{\gamma^*} + c_{\gamma^*})\} (\gamma_{kl}^*)^{z_{kl} + 1} \exp(-c_{\gamma^*} \gamma_{kl}^*) \\ \propto & \mathcal{G}(\gamma_{kl}^*; a_{\gamma^*} + z_{kl} + z_{kl+1}, b_{\gamma^*} + 2c_{\gamma^*}). \end{aligned} \quad (26)$$

Although the parameters of the above gamma distribution are the same as the parameters of the gamma component in the $\delta_0 - \mathcal{G}$ distribution when $\gamma_{kl} > 0$, we show a formal way of deriving (18) as the full conditional distribution of γ_{kl} :

$$\begin{aligned} & [1_{\{\gamma_{kl}=0\}} \pi_{kl} + 1_{\{\gamma_{kl}>0\}} (1 - \pi_{kl}) \mathcal{G}(\gamma_{kl}; a_{kl}, b_{kl})] \\ \times & \prod_{i,j: Y_{ij}=1} \left[\left\{ \sum_{l=1}^L \gamma_{kl} (t_j - t_{j-1}) 1_{\{t_j \in (\tau_{l-1}, \tau_l]\}} \right\}^{1_{\{x_i > k\}} dN_{ijk}} \exp \left\{ - \sum_{l=1}^L \gamma_{kl} (t_j - t_{j-1}) 1_{\{t_j \in (\tau_{l-1}, \tau_l]\}} \right\} \right], \end{aligned}$$

This expression is simplified to the form:

$$\begin{aligned} & [1_{\{\gamma_{kl}=0\}} \pi_{kl} + 1_{\{\gamma_{kl}>0\}} (1 - \pi_{kl}) C(\gamma_{kl}; a_{kl}, b_{kl})] \\ \times & \gamma_{kl}^{a_{kl} + \sum_{i,j=1, t_j \in (\tau_{l-1}, \tau_l]} 1_{\{x_i > k\}} Y_{ij} dN_{ijk} - 1} \exp \left[-\gamma_{kl} \left\{ b_{kl} + \sum_{i,j: t_j \in (\tau_{l-1}, \tau_l]} 1_{\{x_i > k\}} Y_{ij} (t_j - t_{j-1}) \right\} \right], \end{aligned}$$

where $C(\cdot; a, b)$ is the constant term of the $\mathcal{G}(\cdot; a, b)$ density. Furthermore, the second line in the above expression can also be written as $\mathcal{G}(\gamma_{kl}; \tilde{a}_{kl}, \tilde{b}_{kl}) / C(\tilde{a}_{kl}, \tilde{b}_{kl})$, where \tilde{a}_{kl} and \tilde{b}_{kl} are defined in (20). After dividing by the normalizing constant, the conditional posterior density of γ_{kl} is:

$$1_{\{\gamma_{kl}=0\}} \tilde{\pi}_{kl} + 1_{\{\gamma_{kl}>0\}} (1 - \tilde{\pi}_{kl}) \mathcal{G}(\gamma_{kl}; \tilde{a}_{kl}, \tilde{b}_{kl}), \quad (27)$$

where $\tilde{\pi}_{kl}$ is defined in expression (19).

REFERENCES

- Andersen, P.K. and Gill, R.D. (1982). Cox regression-model for counting-processes - A large sample study. *Annals of Statistics* **10**, 1100-1120.
- Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data using the Gibbs sampler. *Statistica Sinica* **4**, 505-524.
- Arjas, E. and Gasbarra, D. (1996). Bayesian inference of survival probabilities, under stochastic ordering constraints. *Journal of the American Statistical Association* **91**, 1101-1109.
- Berger, J.O. and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science* **2**, 317-352.
- Clayton, D.G. (1991). A Monte-Carlo method for Bayesian-inference in frailty models. *Biometrics* **47**, 467-485.
- Clayton, D.G. (1994). *Bayesian analysis of frailty models*. Technical Report, Medical Research Council Biostatistics Unit, Cambridge, U.K.
- Dunson, D.B. and Herring, A.H. (2003). Bayesian inferences in the Cox model for order restricted hypotheses. *Biometrics* **59**, 916-923.
- Dunson, D.B. and Herring, A.H. (2004). Bayesian model selection and averaging in additive and proportional hazards models. *Institute of Statistics & Decision Sciences discussion paper* (available for download at [http : //ftp.isds.duke.edu/WorkingPapers/04 – 16.pdf](http://ftp.isds.duke.edu/WorkingPapers/04-16.pdf)).
- Dykstra, R.L. (1982). Maximum likelihood estimation of the survival functions of stochastically ordered random variables. *Journal of the American Statistical Association* **77**, 621-628.

- Dykstra, R.L. and Feltz, C.J. (1989). Nonparametric maximum likelihood estimation of survival functions with a general stochastic ordering and its dual. *Biometrika* **76**, 331-341.
- Dykstra, R.L., Kochar, S., and Robertson, T. (1991). Statistical inference for uniform stochastic ordering in several populations. *The Annals of Statistics* **19**, 870-888.
- Dykstra, R.L., Kochar, S., and Robertson, T. (1995). Inference for likelihood ratio ordering in the two-sample problem. *Journal of the American Statistical Association* **90**, 1034-1040.
- Feltz, C.J. and Dykstra, R.L. (1985). Maximum likelihood estimation of the survival functions of N stochastically ordered random variables. *Journal of the American Statistical Association* **80**, 1012-1019.
- Gamerman, D. (1991). Dynamic Bayesian models for survival data. *Applied Statistics* **40**, 63-79.
- Gelfand, A.E. and Kottas, A. (2001). Nonparametric Bayesian modeling for stochastic order. *Annals of the Institute of Statistical Mathematics* **53**, 865-876.
- George, E.I. and McCulloch, R.E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7**, 339-373.
- Gönen, M., Westfall, P.H., and Johnson, W.O. (2003). Bayesian multiple testing for two-sample multivariate endpoints. *Biometrics* **59**, 76-82.
- Gray, R.J. (1994). A Bayesian analysis of institutional effects in a multicenter cancer clinical trial. *Biometrics* **50**, 244-253.
- Hjort, N.L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Annals of Statistics* **18**, 1259-1294.

- Hoff, P.D. (2003). Bayesian methods for partial stochastic orderings. *Biometrika* **90**, 303-317.
- Ibrahim, J.G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer.
- Kalbfleisch, J.D. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society B* **40**, 214-221.
- Lin, D.Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.
- Lin, D.Y., Oakes, D., and Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika* **85**, 289-298.
- National Toxicology Program (1999). Toxicology and carcinogenesis studies of ethylbenzene (case no. 100-41-4) in F344/N rats and B6C3F₁ mice (inhalation studies). Technical Report, U.S. Department of Health and Human Services, Research Triangle Park.
- Neelon, B. and Dunson, D.B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics* **60**, in press.
- Nieto-Barajas, L.E. and Walker, S.G. (2002). Markov beta and gamma processes for modeling hazard rates. *Scandinavian Journal of Statistics* **29**, 413-424.
- Sen, P.K. (1984). Subhypotheses testing against restricted alternatives for the Cox regression model. *Journal of Statistical Planning and Inference* **10**, 31-42.
- Silvapulle, M.J. (1994). On tests against one-sided hypotheses in some generalized linear models. *Biometrics* **50**, 853-858.
- Silvapulle, M.J. and Silvapulle, P. (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association* **90**, 342-349.

Singh, B. and Wright, F.T. (1996). Testing order restricted hypotheses with proportional hazards. *Lifetime Data Analysis* **2**, 363-389.

Singh, B. and Wright, F.T. (1998). Comparing survival times for treatments with those for a control under proportional hazards. *Lifetime Data Analysis* **4**, 265-279.

Sinha, D. and Dey, D.K. (1998). Survival analysis using semiparametric Bayesian methods. In *Practical nonparametric and semiparametric Bayesian statistics*. (eds D. Dey, P. Müller, and D. Sinha). New York: Springer.

Walker, S.G. and Mallick, B.K. (1997). Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing. *Journal of the Royal Statistical Society: Series B* **59**, 845-860.

Westfall, P.H., Johnson, W.O. and Utts, J.M. (1997). A Bayesian perspective on the Bonferroni adjustment. *Biometrika* **84**, 419-427.

Table 1: Survival rate of rats exposed to ethylbenzene

	Control	Dose 1	Dose 2	Dose 3
Natural death	7/50	16/50	11/50	22/50
Moribund sacrifice	28/50	20/50	26/50	26/50
Survived to study's end	15/50	14/50	13/50	2/50

Table 2: Prior probabilities, posterior probabilities, and Bayes factors for overall effects across dose and time by values of $c_{\pi k}$, including the simulation study

$c_{\pi k}$	Prior			Posterior			Bayes factors		
	Dose 1	Dose 2	Dose 3	Dose 1	Dose 2	Dose 3	Dose 1	Dose 2	Dose 3
50, Sim.	0.21	0.36	0.49	0.21	0.41	0.63	1.00	1.24	1.77
10	0.19	0.35	0.48	0.20	0.38	0.99	1.07	1.14	> 100
50	0.21	0.36	0.49	0.25	0.48	0.99	1.25	1.64	> 100
100	0.21	0.36	0.49	0.29	0.65	0.99	1.54	3.30	> 100

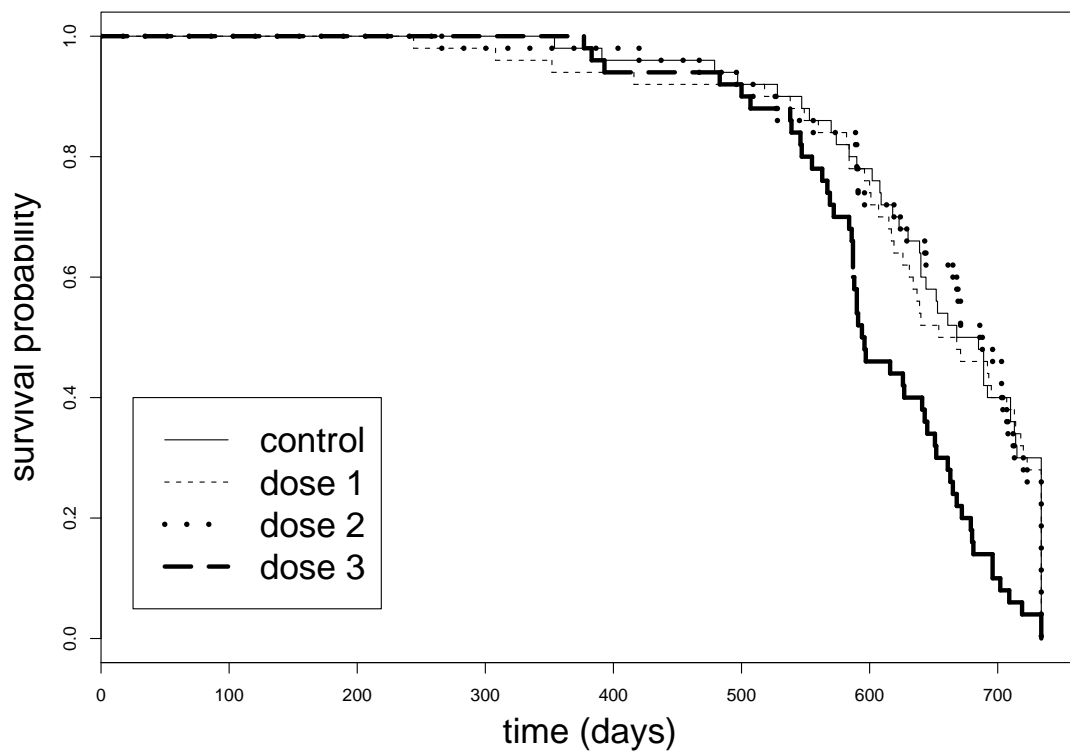


Figure 1: Kaplan-Meier estimates of the survival curves for male rats exposed to ethylbenzene

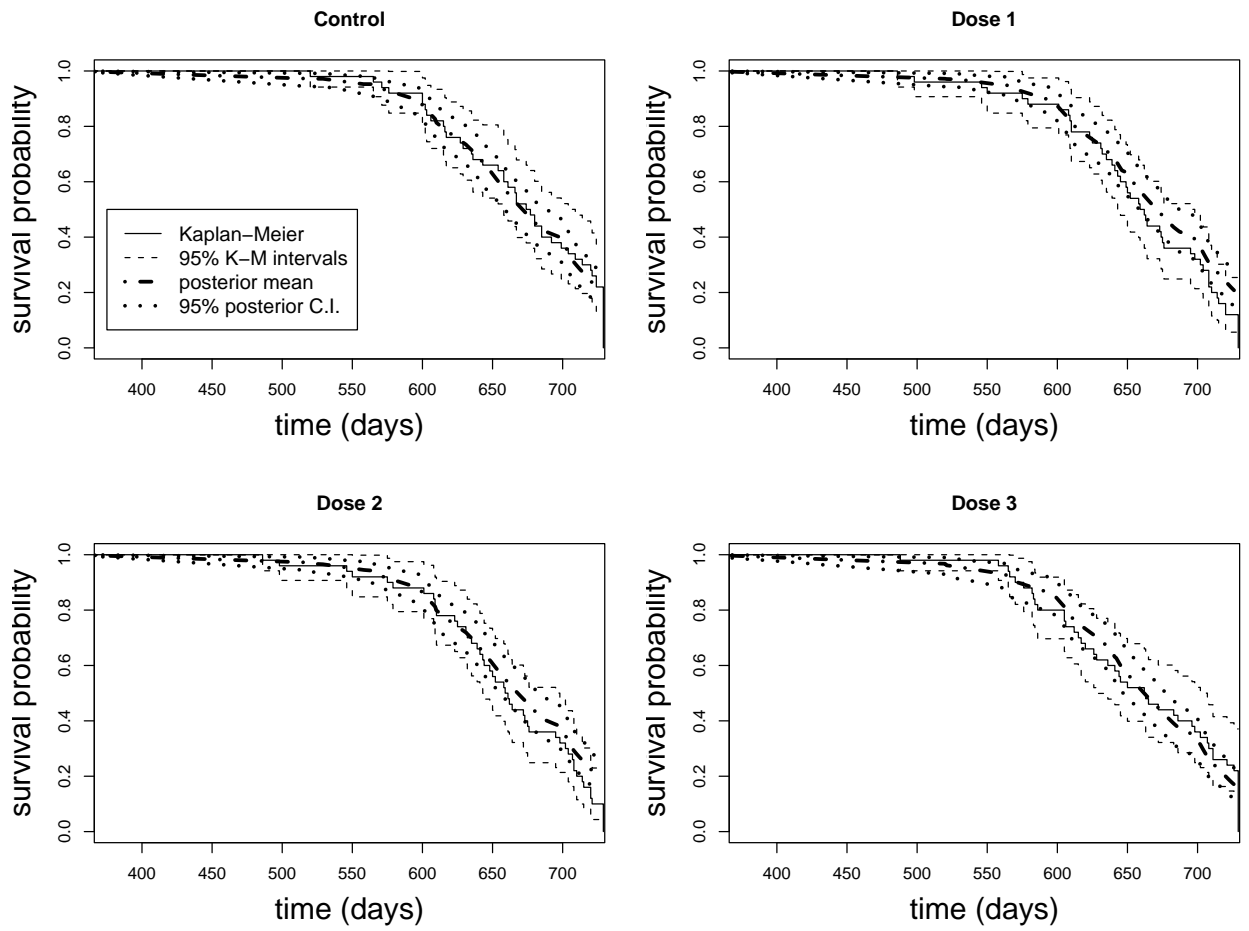


Figure 2: Simulation results: posterior means and 95% pointwise credible intervals of the survival probability, after the first year, overlaid on the Kaplan-Meier curves with their 95% intervals

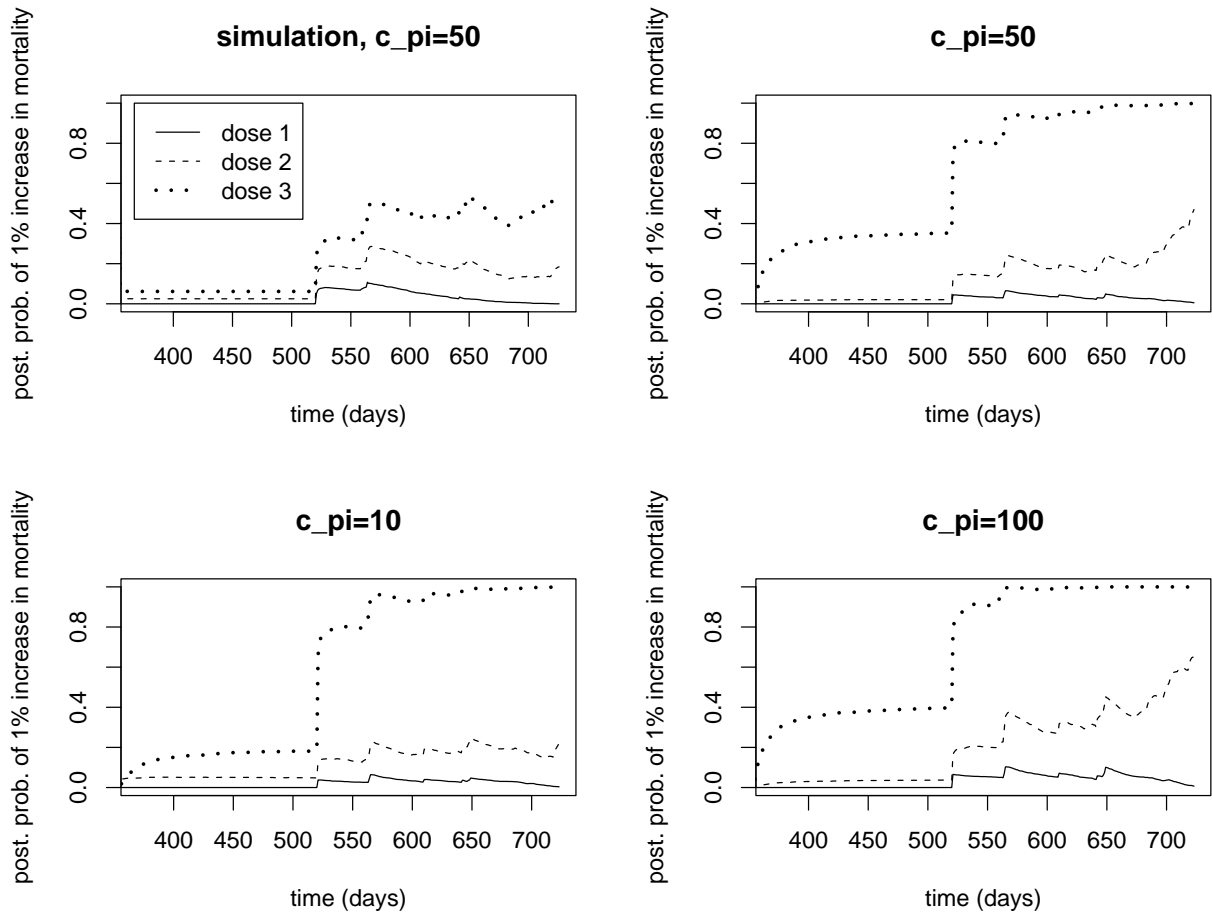


Figure 3: Posterior probabilities of a biologically-significant (i.e., 1%) increase in mortality for each dose x age combination

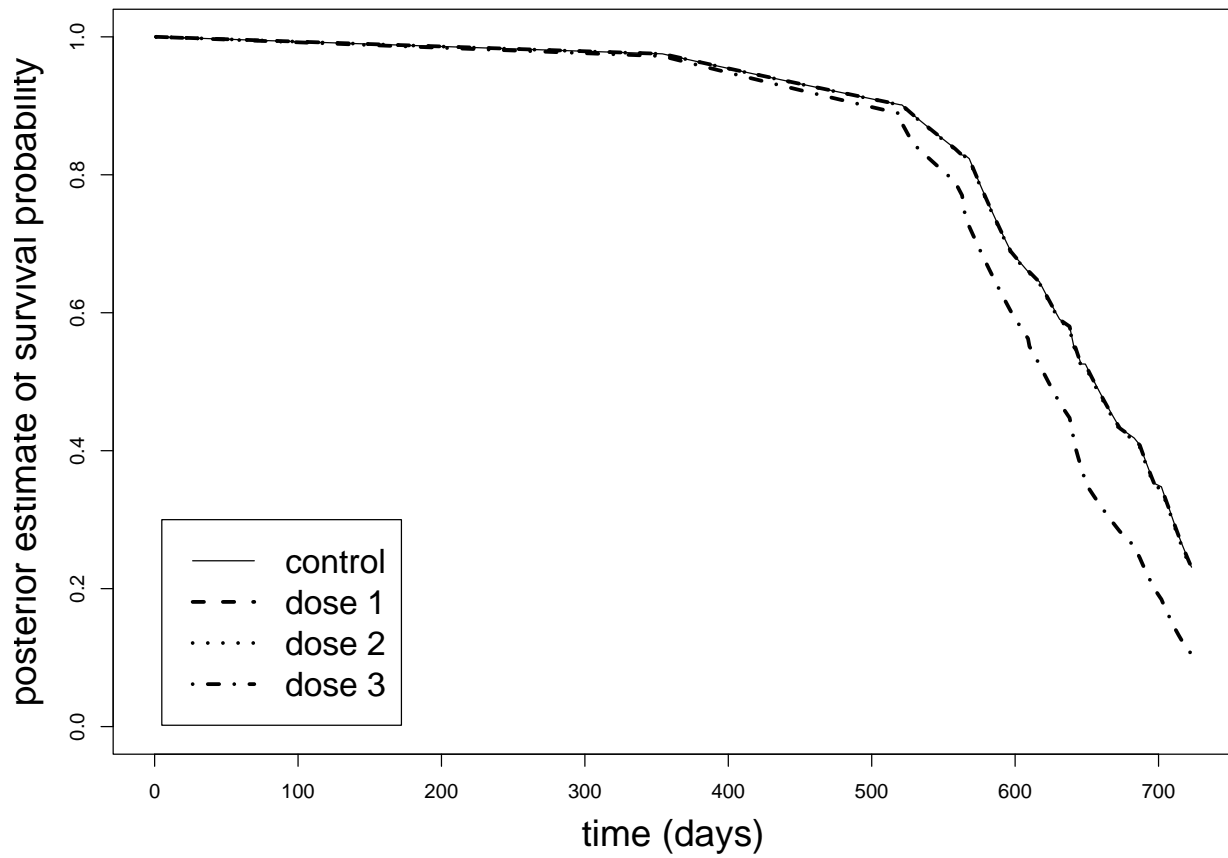


Figure 4: Toxicology results: posterior means of the survival probability for male rats exposed to ethylbenzene