

# Bayesian Model Selection and Averaging in Additive and Proportional Hazards Models

David B. Dunson<sup>1,\*</sup> and Amy H. Herring<sup>2</sup>

<sup>1</sup>Biostatistics Branch,

National Institute of Environmental Health Sciences

MD A3-03, P.O. Box 12233

Research Triangle Park, NC 27709, U.S.A.

<sup>2</sup> Department of Biostatistics

The University of North Carolina

Chapel Hill, NC

\* [dunson1@niehs.nih.gov](mailto:dunson1@niehs.nih.gov)

**SUMMARY.** Although Cox proportional hazards regression is the default analysis for time to event data, there is typically uncertainty about whether the effects of a predictor are more appropriately characterized by a multiplicative or additive model. To accommodate this uncertainty, we place a model selection prior on the coefficients in an additive-multiplicative hazards model. This prior assigns positive probability, not only to the model that has both additive and multiplicative effects for each predictor, but also to sub-models corresponding to no association, to only additive effects, and to only proportional effects. The additive component of the model is constrained to ensure non-negative hazards, a condition often violated by current methods. After augmenting the data with Poisson latent variables, the prior is conditionally conjugate, and posterior computation can proceed via an efficient Gibbs sampling algorithm. Simulation study results are presented, and the methodology is illustrated using data from the Framingham heart study.

**KEY WORDS:** Additive hazards; Cox model; Gibbs sampler; Order restricted inference; Posterior probability; Proportional hazards; Survival analysis; Variable selection

## 1. Introduction

Cox proportional hazards regression (Cox, 1972) is by far the most widely used approach for the analysis of time to event data. Unfortunately, the proportional hazards assumption is often violated, and there is a need for alternative and more flexible models. Many possibilities have been proposed, including accelerated life models, additive hazards models, and additive-multiplicative models, which have both additive and proportional hazards components. When the form of the model and the important predictors are known, a variety of methods are available for parameter estimation. Unfortunately, the regression structure is typically not known *a priori*, and one must account for uncertainty in the model to avoid biased inferences (e.g., about the effect of an important predictor).

This article focuses on the problem of variable selection and inference, first in the additive hazards model (Aalen, 1980; Cox and Oakes, 1984; Lin and Ying, 1994) and then in the more general additive-multiplicative model (Lin and Ying, 1995; Martinussen and Scheike, 2002; Scheike and Zhang, 2002). Most research in this area has focused on frequentist estimation for a given model, and not on model selection or averaging. In addition, there has been essentially no consideration of additive or additive-multiplicative models in the Bayesian literature. A notable exception is the recent article by Beamonte and Bermúdez (2003), which proposed an approach for Bayesian inference in an additive gamma-polygonal hazard model. They did not consider variable selection or additive-multiplicative hazard models.

One drawback of previous methods for estimation of the additive hazards model is that negative hazard estimates can be produced. Although parameter restrictions can be incorporated to solve this problem (e.g., using constrained least squares or the post-processing approach of Lin and Ying, 1994), current inferential procedures and asymptotic theory pertain to the unconstrained case, and it is very difficult to develop appropriate theory for the constrained estimators. Motivated by this consideration and by difficulty in formally accounting for uncertainty in model selection within the frequentist paradigm, this article

proposes a Bayesian approach.

We first develop an approach for Bayesian inference in the additive hazards model, and then generalize the methods to allow additive-multiplicative hazards. A model selection prior is proposed that allows negative, zero, or positive coefficients in the additive and multiplicative components of the model for each predictor. Related priors have been proposed previously for variable selection and model averaging in linear regression (refer to Chipman, George, and McCulloch, 2001; George and McCulloch, 1997; Geweke, 1996) and in the Cox model (Dunson and Herring, 2003). To avoid negative hazards, we adaptively redefine predictors with negative coefficients. Due to the structure of the prior and the likelihood, posterior computation can proceed via a highly efficient Gibbs sampling algorithm.

Our approach differs from previous methods for Bayesian survival analysis, which primarily focus on inferences under an assumed model (see the references in Ibrahim, Chen and Sinha, 2001) or on selecting an optimal model under some criterion (Ibrahim, Chen and MacEachern, 1999; Sinha, Chen and Ghosh, 1999; Faraggi and Simon, 1998). In contrast, our focus is on conducting inferences on covariate effects while accounting for model uncertainty, though the methodology can also be used for variable selection, estimation, and prediction. In this sense, our approach is more closely related to methods for Bayesian model averaging (Volinsky et al., 1997; Volinsky and Raftery, 2000). These earlier methods relied on a large sample approximation to the posterior model probabilities, while we estimate exact posterior probabilities using MCMC. Although previous methods can be used to construct frequentist tests for additive versus proportional hazards, we are the first to consider the problem of model averaging in this context.

Our Bayesian approach has several advantages over frequentist tests. First, the method accounts for uncertainty in the model structure in performing inferences on the effect of a given predictor. If interest focuses on testing for an association between a predictor and the survival time, adjusting for covariates, the approach of using the p-value from the preferred

model (additive or multiplicative) can result in an inflated type I error rate. Accounting for uncertainty in model selection is a well known advantage of Bayesian model averaging procedures. Second, allowing each predictor to have proportional, additive, additive and proportional, or no effect results in a huge number of models to consider. The frequentist approach would require fitting of each of these models, while our approach adaptively moves among the different models based on the posterior model probabilities.

Section 2 proposes an approach for Bayesian inference in the additive hazards model. Section 3 generalizes this approach to accommodate additive-multiplicative models. Section 4 presents results for simulated data. Section 5 applies the methods to data from the Framingham heart study, and Section 6 discusses the results. Details on posterior computation are provided in an appendix.

## 2. Bayesian Inference in Additive Hazards Models

### 2.1 *The Model and a Standard Bayesian Approach*

Consider the additive hazards model (Lin and Ying, 1994; McKeague and Sasini, 1994):

$$\lambda(t; \mathbf{x}_i) = \lambda_0^*(t) + \mathbf{x}_i' \boldsymbol{\alpha}^*, \quad (1)$$

where  $\lambda_0^*(t)$  is the baseline hazard function,  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$  is a  $p \times 1$  vector of predictors,  $\boldsymbol{\alpha}^* = (\alpha_1^*, \dots, \alpha_p^*)'$  are unknown regression coefficients, and  $\lambda(t; \mathbf{x}_i) \geq 0$  for all  $i, t$  to maintain the interpretation of  $\lambda(t; \mathbf{x}_i)$  as a hazard rate. In principle, this restriction can be incorporated in a Bayesian analysis by choosing a prior distribution for  $\lambda_0^*(t)$  and  $\boldsymbol{\alpha}^*$  with constrained support. For example, the prior for  $\alpha_k^*$  conditional on  $\boldsymbol{\alpha}_{(-k)}^*$ ,  $\lambda_0^*(t)$ , and  $\mathbf{X}$  could be chosen as truncated normal subject to the restriction

$$\alpha_k^* \geq \max_{i,t} \left[ -\frac{1}{x_{ik}} \{ \lambda_0^*(t) + \mathbf{x}_{i(-k)}' \boldsymbol{\alpha}_{(-k)}^* \} \right].$$

After specifying priors for  $\lambda_0^*(t)$  and the remaining elements of  $\boldsymbol{\alpha}^*$  subject to similar restrictions, posterior computation can proceed via an MCMC algorithm.

There are three main drawbacks of this approach. First, it only guarantees that the restriction holds for the values of  $\mathbf{x}$  in the study sample. Because a common goal is estimation and prediction for individuals having specific values of the predictors that may not be represented, it would be better to have the restriction apply for all possible values of  $\mathbf{x}$  in the range of the data. Appropriate bounds can be time consuming to calculate, particularly because this calculation needs to be repeated for each parameter at each iteration of the MCMC algorithm. A second problem is that the MCMC algorithm can be inefficient due to autocorrelation induced by the constraints on a particular parameter being highly dependent on the values of the other parameters. Finally, the above type of prior is not structured to allow efficient variable selection.

## 2.2 Model Formulation and Parameter Interpretation

Motivated by the drawbacks of standard approaches, we propose an alternative strategy, focusing initially on the restructured additive hazards model:

$$\lambda(t; \mathbf{x}_i) = \lambda_0(t) + \sum_{k=1}^p \left\{ 1_{(M_k=-1)}(1 - x_{ik}) + 1_{(M_k=1)}x_{ik} \right\} \alpha_k = \lambda_0(t) + \mathbf{x}_i^{*'} \boldsymbol{\alpha}, \quad (2)$$

where  $\lambda_0(t)$  is the redefined baseline hazard function,  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$  is a  $p \times 1$  vector of candidate predictors standardized so that  $x_{ik} \in [0, 1]$  for  $k = 1, \dots, p$ ,  $\mathbf{x}_i^* = (x_{i1}^*, \dots, x_{ip}^*)'$ ,  $x_{ik}^* = 1_{(M_k=-1)}(1 - x_{ik}) + 1_{(M_k=1)}x_{ik}$ ,  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)'$  are unknown positive parameters, and  $M_k$  is a model indicator that takes values  $M_k = -1$  if the  $k$ th predictor is associated with a reduction in the hazard,  $M_k = 0$  if there is no change, and  $M_k = 1$  if there is an increase.

Bounded predictors can be standardized to the unit interval by subtracting the minimum possible value and dividing by the range. For unbounded predictors, the sample minimum and range can be used, potentially expanding the range to include interesting values of the predictors outside of the sample range in order to use the model for extrapolation and more flexible predictions. Although standardization is data-dependent for unbounded predictors, this does not cause problems in interpretation, since one can transform back to the original

scale in interpreting the coefficients, as we describe below.

Expression (2) ensures that the hazard function is non-negative,  $\lambda(t; \mathbf{x}_i) \geq 0$ , by reversing the direction of predictors having a negative impact on the hazard ( $M_k = -1$ ), and restricting the coefficients  $\{\alpha_k\}$  to be positive. The resulting model can be equivalently expressed in the form shown in (1) by letting  $\lambda_0^*(t) = \lambda_0(t) + \sum_{k=1}^p 1_{(M_k=-1)}\alpha_k$  and  $\alpha_k^* = M_k\alpha_k$ . The process of reversing the direction of predictors with  $M_k = -1$  effectively redefines the baseline hazard depending on the model indicators. The regression coefficient  $\alpha_k^*$  depends on both the model indicator,  $M_k$ , and the strictly-positive coefficient,  $\alpha_k$ .

To illustrate that this formulation does not adversely affect model interpretation, consider a simple example. In particular, suppose that  $\tilde{x}_{ik}$  denotes the unstandardized body mass index (bmi) for subject  $i$ , let  $A_k$  denote the minimum value of bmi (among the sample and other values of interest), let  $B_k$  denote the maximum value of bmi, and let  $x_{ik} = (\tilde{x}_{ik} - A_k)/(B_k - A_k)$  denote the standardized bmi. If increasing bmi predicts a reduction in the hazard ( $M_k = -1$ ), then expression (2) replaces  $x_{ik}$  by  $1 - x_{ik}$ . Suppose that our interest focuses on estimating the change in hazard attributable to increasing bmi from 25 to 30, holding other predictors constant. Under expression (2), this change is simply

$$\left(\frac{30 - 25}{B_k - A_k}\right)M_k \alpha_k,$$

which is an unbounded negative number if  $M_k = -1$ , is 0 if  $M_k = 0$ , and is an unbounded positive number if  $M_k = 1$ . In practice, a posterior distribution can be estimated for the change in hazard between any two values of the predictor, with  $[A_k, B_k]$  chosen to be wide enough to include all values of interest. Samples from this posterior can include negative, zero, and positive values, representing averaging across the model indicator  $M_k$ .

It is important to note that  $\alpha_k$  is not interpretable by itself as a regression coefficient, and inferences about predictor effects should be based on combining  $\alpha_k$  and  $M_k$ . The practice of expressing a regression coefficient in terms of a model indicator and a potential slope in order

to facilitate computation has a long history in the Bayesian variable selection literature.

### 2.3 Counting Process Likelihood Formulation

We focus on the case in which data consist of continuous right censored survival times, with censoring assumed to be non-informative. For subject  $i$  ( $i = 1, \dots, n$ ), let  $N_i(t) = 1$  if failure occurs in interval  $[0, t]$  and  $N_i(t) = 0$  otherwise, and let  $Y_i(t) = 1$  if the subject is at risk at time  $t$  and  $Y_i(t) = 0$  otherwise. Letting  $dN_i(t)$  denote the increment on  $N_i(t)$  over the small interval  $[t, t + dt)$ , the likelihood of the data is proportional to

$$\prod_{i=1}^n \left( \prod_{t \geq 0} [Y_i(t) \{\lambda_0(t) + \mathbf{x}_i^{*'} \boldsymbol{\alpha}\}]^{dN_i(t)} \right) \exp \left( - \int_{t \geq 0} Y_i(t) \{\lambda_0(t) + \mathbf{x}_i^{*'} \boldsymbol{\alpha}\} dt \right). \quad (3)$$

Since the increments are infinitesimal, the  $dN_i(t)$  contribute to the likelihood in the same manner as independent Poisson random variables even though  $dN_i(t) \leq 1$  for all  $i$  and  $t$ .

Let  $\mathbf{t} = (t_1, \dots, t_J)'$  denote the union of the unique failure times observed in a data set, pre-specified interval endpoints  $\mathbf{s} = (s_1, \dots, s_H)'$  with  $s_0 < s_1 < \dots < s_H < s_{H+1} = \infty$ , and the maximum follow-up time. Note that likelihood (3) can be reexpressed as

$$\prod_{i=1}^n \prod_{j=1}^J \left( \prod_{t \in (t_{j-1}, t_j]} [Y_i(t) \{\lambda_0(t) + \mathbf{x}_i^{*'} \boldsymbol{\alpha}\}]^{dN_i(t)} \right) \exp \left( - \int_{t \in (t_{j-1}, t_j]} Y_i(t) \{\lambda_0(t) + \mathbf{x}_i^{*'} \boldsymbol{\alpha}\} dt \right). \quad (4)$$

Under the assumption that the risk accrued in the interval  $(t_{j-1}, t_j]$  is small,

$$\int_{t_{j-1}}^{t_j} Y_i(t) \{\lambda_0(t) + \mathbf{x}_i^{*'} \boldsymbol{\alpha}\} dt \approx 0 \quad \text{for all } i, j, \quad (5)$$

the likelihood contribution across this interval for individuals at risk is approximately

$$\{d\Lambda_{0j} + \mathbf{x}_i^{*'} \boldsymbol{\alpha}(t_j - t_{j-1})\}^{dN_{ij}} \exp \left( - \{d\Lambda_{0j} + \mathbf{x}_i^{*'} \boldsymbol{\alpha}(t_j - t_{j-1})\} \right),$$

where  $d\Lambda_{0j} = \int_{t_{j-1}}^{t_j} \lambda_0(t) dt$ , and  $dN_{ij} = 1$  if subject  $i$  fails at time  $t_j$  and  $dN_{ij} = 0$  otherwise.

Assumption (5) is reasonable even when  $t_j - t_{j-1}$  is large, since by definition at most one event can occur in  $(t_{j-1}, t_j]$  in the data set, and hence the hazard must be very small if the interval is wide. The likelihood approximation relies on the assumption that a Poisson

random variable with mean (5) has a low probability of being greater than one, which is a very weak assumption, particularly for moderate to large  $n$ . Hence, we replace (3) with

$$\prod_{i=1}^n \prod_{j: Y_{ij}=1} \{d\Lambda_{0j} + \mathbf{x}_i^{*'} \boldsymbol{\alpha}(t_j - t_{j-1})\}^{dN_{ij}} \exp\left(-\{d\Lambda_{0j} + \mathbf{x}_i^{*'} \boldsymbol{\alpha}(t_j - t_{j-1})\}\right), \quad (6)$$

where  $Y_{ij} = 1$  if subject  $i$  is at risk at time  $t_j$  and  $Y_{ij} = 0$  otherwise. A related approximation to (6) was proposed by Clayton (1994) for proportional hazards frailty models.

#### 2.4 Prior Specification

To complete a Bayesian specification of the model, prior distributions are needed for the baseline hazard function  $\lambda_0(t)$ , model indicators  $\mathbf{M} = (M_1, \dots, M_p)'$ , and coefficients  $\boldsymbol{\alpha}$ . For the model indicators, we choose multinomial priors:

$$\pi(\mathbf{M}) = \prod_{k=1}^p (p_{k,-1}^+)^{1_{(M_k=-1)}} (p_{k,0}^+)^{1_{(M_k=0)}} (p_{k,1}^+)^{1_{(M_k=1)}}, \quad (7)$$

where  $\mathbf{p}_k^+ = (p_{k,-1}^+, p_{k,0}^+, p_{k,1}^+)'$  are prior model probabilities for the  $k$ th predictor, with  $p_{k,-1}^+ = \Pr(M_k = -1)$  the prior probability of a negative effect,  $p_{k,0}^+ = \Pr(M_k = 0)$  the prior probability of no effect, and  $p_{k,1}^+ = \Pr(M_k = 1)$  the prior probability of a positive effect. When prior information is limited, a reasonable default would be to set  $\mathbf{p}_k^+ = (0.25, 0.5, 0.25)'$  to correspond to 0.5 prior probability of no effect of the  $k$ th predictor and equal probability of a positive or negative effect. In addition, the prior is flexible enough to incorporate prior information on the direction of the effect. For example, in applications involving potentially adverse exposures, it may be reasonable to assume *a priori* that the hazard of disease is non-decreasing with level of exposure, though one may be uncertain whether there is a no effect or there is an increase. In such cases, one can let  $\mathbf{p}_k^+ = (0, 0.5, 0.5)'$ .

The model indicators  $\mathbf{M}$  index the direction and occurrence of effects for the different predictors, but do not quantify the magnitude of the effect. The slopes are characterized by the coefficients  $\boldsymbol{\alpha}$ . When  $M_k = 0$ , there is no effect of the  $k$ th predictor and  $\alpha_k$  effectively drops out of expression (2). When  $M_k \neq 1$ ,  $\alpha_k$  is a strictly positive parameter measuring

the absolute value of the slope on the standardized scale of the predictor. We choose

$$\pi(\alpha_k | M_k \neq 0) = \mathcal{G}(\alpha_k; a_k^+, b_k^+), \text{ for } k = 1, \dots, p, \quad (8)$$

with  $\mathcal{G}(\cdot; a, b)$  denoting the gamma density having mean  $a/b$  and variance  $a/b^2$ , with prior independence assumed across  $k$ , and with  $\pi(\alpha_k | M_k = 0)$  left unspecified. The hyperparameters  $\mathbf{a}^+ = (a_1^+, \dots, a_p^+)'$  and  $\mathbf{b}^+ = (b_1^+, \dots, b_p^+)'$  can be chosen based on one's expectation for the magnitude of the slope on the standardized scale, with large values of  $a_k^+/b_k^{+2}$  corresponding to a diffuse prior. Our prior is related to priors for Bayesian variable selection in linear regression (e.g., Geweke, 1996) in that we use a mixture of point masses at 0, corresponding to predictors being excluded, and continuous densities, quantifying the slopes for predictors in the model.

Although the methodology can be applied for other choices of prior for the baseline hazard function, we focus on a piecewise constant model for simplicity in exposition, with  $\lambda_0(t) = \lambda_{0h}$  for  $t \in (s_{h-1}, s_h]$ , where  $s_0 < s_1 < \dots < s_H < s_{H+1} = \infty$  are pre-specified interval endpoints, which can be chosen to be arbitrarily close together. We then assume independent gamma priors for  $\boldsymbol{\lambda}_0 = (\lambda_{01}, \dots, \lambda_{0H})'$ :

$$\pi(\boldsymbol{\lambda}_0) = \prod_{h=1}^H \mathcal{G}(\lambda_{0h}; a_{0h}, b_{0h}), \quad (9)$$

where  $a_{0h}/b_{0h}$  is the prior expectation for  $\lambda_{0h}$ , and  $a_{0h}/b_{0h}^2$  is the prior variance.

### 2.5 Auxiliary Variables Gibbs Sampler

To simplify efficient posterior computation, we use a data augmentation approach. First note that under (6) it is as if the  $dN_{ij}$  are independent Poisson distributed random variables:

$$dN_{ij} \overset{ind}{\sim} \text{Poisson}\left(d\Lambda_{0j} + \sum_{k=1}^p x_{ik}^* \alpha_k\right) \text{ for all } i, j : Y_{ij} = 1. \quad (10)$$

Unfortunately, the additive form of the Poisson sum results in non-standard full conditional posterior distributions, and it is difficult to develop an efficient approach to computation.

However, we can solve this problem by re-expressing (10) in an augmented form involving independent Poisson latent variables corresponding to each term in the expression for the Poisson mean. In particular, we let

$$\begin{aligned}
dN_{ij} &= dN_{ij0} + \sum_{k=1}^p \mathbf{1}_{(M_k \neq 0)} dN_{ijk}, \quad \text{for all } i, j : Y_{ij} = 1, \\
\pi(dN_{ij0}) &= \text{Poisson}(dN_{ij0}; d\Lambda_{0j}) \\
\pi(dN_{ijk}) &= \text{Poisson}(dN_{ijk}; (t_j - t_{j-1})x_{ik}^* \alpha_k), \quad k = 1, \dots, p.
\end{aligned} \tag{11}$$

Using the property that the sum of independent Poisson random variables is Poisson, it is straightforward to show that this expression is equivalent to (10) integrating out the latent variables. This formulation allows us to take advantage of Poisson-gamma conjugacy to obtain simple conditional posterior distributions.

Derivation of the conditional posterior distributions, to be used in Gibbs sampling, is outlined in Appendix A. The sampler iterates through the following steps:

**Step 1.** Sample the latent variables  $(dN_{ij0}, dN_{ij1}, \dots, dN_{ijp})'$ , for  $i, j : Y_{ij} = 1$ , jointly from their full conditional posterior distribution (16) as follows:

1. If  $dN_{ij} = 0$  then let  $dN_{ij0} = dN_{ijk} = 0$  for all  $k : M_k \neq 0$ ,
2. If  $dN_{ij} > 0$  then sample  $(dN_{ij0}, dN_{ijk} : M_k \neq 0, k = 1, \dots, p)$  from

$$\text{Multinomial}(dN_{ij}; P_{ij0}, P_{ijk} : M_k \neq 0, k = 1, \dots, p),$$

where  $P_{ij0} = d\Lambda_{0j} / (d\Lambda_{0j} + \sum_{k: M_k \neq 0}^p (t_j - t_{j-1})x_{ik}^* \alpha_k)$  and

$$P_{ijk} = \frac{(t_j - t_{j-1})x_{ik}^* \alpha_k}{d\Lambda_{0j} + \sum_{k: M_k \neq 0}^p (t_j - t_{j-1})x_{ik}^* \alpha_k} \quad \text{for } k : M_k \neq 0, k = 1, \dots, p.$$

3. For all  $k : M_k = 0$ , sample  $dN_{ijk}$  from  $\text{Poisson}((t_j - t_{j-1})x_{ik}^* \alpha_k)$ ,  $k = 1, \dots, p$ .

**Step 2.** Sample  $\lambda_{0h}$ ,  $h = 1, \dots, H$ , from expression (17).

**Step 3.** Sample  $M_k$ ,  $k = 1, \dots, p$ , from expression (18).

**Step 4.** For  $k : M_k \neq 0$ , sample  $\alpha_k$  from expression (19), and otherwise set  $\alpha_k$  equal to its previous value.

Under mild regularity conditions, samples from this algorithm will converge to a target distribution that is the joint posterior distribution.

## 2.6 Hypothesis Testing, Variable Selection, and Model Averaging

Posterior summaries calculated from the output of the Gibbs sampling algorithm can be used for hypothesis testing, variable selection, and model averaging. There is often interest in testing the null hypothesis of no additive effect of the  $k$ th predictor, which we express as  $H_{0k}^+ : M_k = 0$ , against a one- or two-sided alternative. Posterior model probabilities can be used as a basis for inferences of this type. The posterior probability of  $H_{0k}^+$  can be calculated by averaging  $1_{(M_k=0)}$  across the Gibbs iterations. The estimated posterior probability can be used directly as a weight of evidence in favor of the null hypothesis (i.e., as a Bayesian alternative to the p-value), or in calculating a Bayes factor (Kass and Raftery, 1995). One may set  $p_{k,-1}^+ = 0$  or  $p_{k,1}^+ = 0$  to conduct a one-sided hypothesis test.

This approach has advantages over classical testing procedures for additive hazard models, which rely on large sample approximations that may not be reasonable when sample sizes are moderate. Our Bayesian approach does not rely on large sample theory (except with respect to the number of iterations of the Gibbs sampler, which can be increased easily). In addition, the non-negative hazards constraint is incorporated, not only for a point estimate, but for all samples from the posterior. Hence, credible regions for the hazard function will not include negative values.

If interest focuses not on hypothesis testing but instead on selecting the best subset from the vector of candidate predictors,  $\mathbf{x}_i$ , one can rely on the posterior model probabilities for the  $2^p$  possible subsets of  $\mathbf{x}_i$ . The Gibbs sampler adaptively adds or drops predictors one at a time according to the conditional model probabilities given the current values for the

coefficients of the other predictors. Across the iterations, samples will tend to congregate on the preferred models according to their posterior probability, with the penalty for model complexity determined by the values of  $\mathbf{p}_k^+$ , for  $k = 1, \dots, p$ . Setting  $p_{k,0}^+ = 0.5$  corresponds to equal prior probabilities for inclusion or exclusion of the predictor. The next section generalizes the approach to allow both additive and multiplicative effects.

### 3. Inference in Additive-Multiplicative Hazard Models

For a more general hazard model that has proportional and additive components, let

$$\lambda(t; \mathbf{x}_i) = \lambda_0(t) \exp(\mathbf{x}_i' \boldsymbol{\beta}) + \mathbf{x}_i^{*'} \boldsymbol{\alpha}, \quad (12)$$

where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  are proportional hazards coefficients, and the remaining parameters are as defined in Subsection 2.2. For notational convenience, we define the model so that the same predictors are in the proportional and additive components. However, by choosing a prior distribution that assigns positive probability to zero values for the regression coefficients, we can allow predictors to be excluded from one or both of the component models. We adapt our prior to accommodate the multiplicative component and then generalize the Gibbs sampler accordingly.

In choosing the prior for  $(\boldsymbol{\lambda}_0, \mathbf{M}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ , we assume  $\pi(\boldsymbol{\lambda}_0, \mathbf{M}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \pi(\boldsymbol{\lambda}_0) \pi(\mathbf{M}, \boldsymbol{\alpha}) \pi(\boldsymbol{\beta})$ , where  $\pi(\mathbf{M}, \boldsymbol{\alpha})$  and  $\pi(\boldsymbol{\lambda}_0)$  are as shown in expressions (7)-(9), respectively. To induce a prior on the proportional hazards regression coefficients,  $\boldsymbol{\beta}$ , we let  $\gamma_k = \exp(\beta_k)$  and  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)'$  with

$$\pi(\boldsymbol{\gamma}) = \prod_{k=1}^p \text{I}_1\text{-}\mathcal{G}_{\mathcal{A}_k}(\gamma_k; p_{k,0}^\times, a_k^\times, b_k^\times), \quad (13)$$

where  $\text{I}_1\text{-}\mathcal{G}_{\mathcal{A}_k}(p, a, b)$  denotes the one-inflated truncated gamma density consisting of the mixture of a point mass at one (with probability  $p$ ) and a  $\mathcal{G}(a, b)$  gamma density truncated so that  $\gamma_k \in \mathcal{A}_k$ . We use the  $\times$  superscript to index the hyperparameters corresponding to the multiplicative component.

The one-inflated truncated gamma density was proposed by Dunson and Herring (2003) for Bayesian inference in the Cox model. The point mass at one for  $\gamma_k$  corresponds to the model with  $\beta_k = 0$  and  $x_{ik}$  effectively excluded from the proportional hazards component. Setting  $\mathcal{A}_k$  equal to  $(0, \infty)$  avoids restrictions on the possible values of  $\beta_k$ . To restrict  $\beta_k \leq 0$  or  $\beta_k \geq 0$  in conducting a one-sided hypothesis test, set  $\mathcal{A}_k$  equal to  $(0, 1]$  or  $[1, \infty)$ , respectively.

Under the proposed prior, the following possibilities exist for the  $k$ th predictor: (a) proportional hazards only (prior probability =  $(1 - p_{k,0}^\times)p_{k,0}^+$ ); (b) additive hazards only ( $p_{k,0}^\times(1 - p_{k,0}^+)$ ), (c) both proportional and additive hazards ( $(1 - p_{k,0}^\times)(1 - p_{k,0}^+)$ ); and (d) neither proportional nor additive hazards ( $p_{k,0}^\times p_{k,0}^+$ ). Allowing the predictor to be in both components of the model adds flexibility to the regression function. To express *a priori* ignorance about whether the  $k$ th predictor has any effect and, given an effect, whether it is additive or proportional, one could set

$$\Pr(H_{0k} : \beta_k = \alpha_k = 0) = \Pr(H_{0k}^\times : \beta_k = 0) \times \Pr(H_{0k}^+ : \alpha_k = 0) = p_{k,0}^\times p_{k,0}^+ = 0.5,$$

and then let  $p_{k,0}^\times = p_{k,0}^+ = \sqrt{0.5}$ . This choice is reasonable when the focus is on testing  $H_{0k} : \text{no effect of the } k\text{th predictor vs } H_{1k} : \text{some effect}$ .

The Gibbs sampler of Subsection 2.5 and Appendix A can be generalized by incorporating minor modifications of the conditional densities used in steps 1-4 and by incorporating an additional step for sampling  $\gamma_k = \exp(\beta_k)$  from its full conditional distribution, for  $k = 1, \dots, p$ . We again utilize the data augmentation approach, but with the following prior density for the latent variable  $dN_{ij0}$ :

$$\pi(dN_{ij0}) = \text{Poisson}(dN_{ij0}; d\Lambda_{0j} \exp(\mathbf{x}'_i \boldsymbol{\beta})). \quad (14)$$

Details needed to implement the algorithm are provided in Appendix B. Hypothesis testing, variable selection, and model averaged estimation can proceed using the approach described

in Subsection 2.6 with straightforward modifications. The method is illustrated through application to simulated and real data examples in Sections 4 and 5, respectively.

#### 4. Simulation Study

We conducted a small simulation study to verify that the method can distinguish between additive and multiplicative models, that the approach has good frequentist operating characteristics, and that our code is working properly. We focused on the case in which  $p = 4$  and  $n = 651$ , values taken from the data example in Section 5. We assumed piecewise constant baseline hazards, with  $\lambda_0(t)$  fixed at 0.005, 0.01, 0.02, 0.04, and 0.08 within the age intervals  $(45, 50]$ ,  $(50, 55]$ ,  $(55, 60]$ ,  $(60, 65]$ , and  $(65, 71]$ , respectively. All subjects were assumed to begin the study at 45 years of age disease free.

We simulated failure times under model (12), with four predictors  $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})'$  generated from independent Bernoulli(0.5) distributions and with the regression coefficients taking one of the following sets of values:

Model	Additive				Multiplicative			
	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$
1	0.03	0	0	0	1	1	1	1
2	0	0	0	0	2	1	1	1
3	0.03	0	0	0	1	2	1	1
4	0	0	0	0	1	1	1	1

where  $\alpha_k = 0$  if the  $k$ th predictor is excluded from the additive component, and  $\gamma_k = 1$  if the predictor is excluded from the multiplicative component. Values of  $\alpha_k = 0.03$  and  $\gamma_k = 2.0$  were chosen to correspond to moderate but biologically important (in the context of the data example of Section 5) increases in the hazard. Censoring times were generated independently of the failure times from a uniform(45,71) distribution. 50 data sets were generated for each choice of model.

We analyzed each data set separately using the approach described in Section 3 with  $p_{k,0}^\times = p_{k,0}^+ = \sqrt{0.5}$  for  $k = 1, 2, 3, 4$ . In addition, to express ignorance about whether a predictor has a negative or positive effect, we let  $p_{k,-1}^\times = p_{k,1}^\times = (1 - \sqrt{0.5})/2$  and  $p_{k,-1}^+ =$

$p_{k,1}^+ = (1 - \sqrt{0.5})/2$ . We let  $a_{0h} = b_{0h} = 0.1$  to specify a diffuse prior for the baseline hazard. Finally, we chose  $a_k^+ = 1$ ,  $b_k^+ = 25$ ,  $a_k^\times = .1$ , and  $b_k^\times = .1$  to assign high prior probability to a wide range of plausible values for the additive and multiplicative coefficients. This prior was chosen to be weakly informative in the context of the application of Section 5.

For each data set, we ran the MCMC algorithm for 5,000 iterations, starting at arbitrary parameter values which differed from the true values and discarding the first 1,000 iterations as a burn-in. Based on examination of trace plots of the sampled parameters, convergence occurred rapidly and mixing of the chain was excellent, with low autocorrelation in the samples. Table 1 presents simulation results, including the estimated bias in the model-averaged posterior median, empirical probabilities of rejecting  $H_{0k}$ , the average of the Rao-Blackwellized estimates of  $\Pr(H_{0k} | \text{data})$ , and empirical coverage rates of 95% credible intervals. The empirical rejection rates are calculated as the proportion of simulated data sets for which the estimated posterior probability of  $H_{0k}$  is less than 0.05.

Based on Table 1, there was no evidence of an inflated type I error rate for our Bayesian approach using the hyperparameter values described above. These hyperparameters were chosen prior to any simulation analyses, and we did not consider alternative values due to the extreme computation involved in implementing simulation studies of MCMC methods. The empirical type I error rates for the additive model coefficients ranged from  $0/50=0.00$  to  $2/50=0.04$ , with the average rate being 0.014. Hence, our approach may be slightly conservative at detecting additive effects, though power was high when  $\alpha_k = 0.03$  in our simulations. For the multiplicative coefficients, the empirical type I error rates ranged from  $0/50=0.00$  to  $5/50=0.10$ , with an average rate equal to the nominal level of 0.05. In addition, power to detect a multiplicative effect was high when the true hazard ratio was  $\gamma_k = 2.0$ .

Similar results were obtained when repeating the simulation study with  $p_{k,-1}^\times = p_{k,-1}^+ = 0$  to express *a priori* belief that the effect is non-negative and conduct one-sided tests. In particular, type I error rates were identical and coverage was similar, but (as expected) the

one-sided analysis yielded higher power. In summary, our approach had excellent frequentist operating characteristics in the cases we considered.

## 5. Data Example

### 5.1 Data, Model, and Prior Specification

Data are drawn from the Framingham heart study (Dawber et al., 1951), which followed 1571 individuals who were disease free at roughly age 45. We focus on age at onset of coronary heart disease (CHD) in female subjects in relation to the following predictors: (1) hypertension, (2) above normal cholesterol, (3) overweight (BMI > 25), and (4) obesity (BMI  $\geq$  30) (with 0=no, 1=yes in each case). We assume that the hazard of onset of CHD follows the additive-multiplicative model shown in expression (12), with  $\mathbf{x}_i = (x_{i1}, \dots, x_{i4})'$ , and  $x_{i1}, x_{i2}, x_{i3}$  and  $x_{i4}$  denoting the 0/1 indicators described above, respectively.

Based on a Cox regression model analysis of the data, the hazard ratios for the four predictors were estimated as 1.50, 1.41, 1.36 and 1.31, respectively, with the corresponding 1-sided p-values being 0.042, 0.043, 0.094 and 0.300. It is of interest to compare these results, which are obtained under an assumed proportional hazards model, with results under our approach.

We used the prior specification described in Section 4, but with  $p_{k,-1}^\times = p_{k,-1}^+ = 0$  to formalize our assumption that values of one for these predictors are not associated with a decreased risk of CHD. It is biologically reasonable and consistent with the previous literature to consider a one-sided analysis in which we assess whether each of the predictors has no effect or results in an increased risk. The resulting posterior probabilities of the null hypothesis estimated under our approach represent Bayesian alternatives to the one-sided p-values produced in the Cox analysis. We will assess the sensitivity of our inferences to the prior specification by repeating the analysis for reasonable alternative values of the hyperparameters.

## 5.2 Analysis

We used the Gibbs sampler, outlined in Appendix B, for posterior computation. We discarded the first 1,000 iterations as a burn-in, and calculated posterior summaries based on an additional 15,000 samples. To assess the sensitivity of the results to the choice of prior, we repeated our analysis with (i)  $a_k^+, b_k^+, a_k^\times, b_k^\times, a_{0h}, b_{0h}$  multiplied by 2 to decrease the prior variance; (ii) with these hyperparameters divided by 2 to increase the prior variance; (iii) with  $p_{k,0}^\times = p_{k,0}^+ = 0.5$  instead of  $\sqrt{0.5}$ ; and (iv) with  $p_{k,-1}^\times = p_{k,1}^\times = p_{k,-1}^+ = p_{k,1}^+ = (1 - \sqrt{0.5})/2$ , as in Section 4, to conduct a two-sided analysis allowing increasing or decreasing effects with equal prior probabilities.

For each of the four predictors, Table 2 presents the estimated posterior probabilities of only additive effects ( $H_{1k}^+ \cap H_{0k}^\times$ ), only multiplicative effects ( $H_{0k}^+ \cap H_{1k}^\times$ ), both additive and multiplicative effects ( $H_{1k}^+ \cap H_{1k}^\times$ ), and no association ( $H_{0k}$ ). In addition to the results from the primary analysis, the range of values of  $\Pr(H_{0k} \mid \text{data})$  estimated in the sensitivity analyses (i)-(iv) is presented.

In the primary analysis, the estimated posterior probability that there was no increase in the hazard of CHD for individuals with hypertension, controlling for other factors, was 0.016. The corresponding posterior probabilities for the above normal cholesterol, overweight, and obesity covariates were 0.014, 0.014, and 0.091, respectively. Hence, considering the posterior probability of the null as a Bayesian alternative to the p-value and using the standard 0.05 cutoff for significance, we conclude that hypertension, above normal cholesterol, and being overweight are all statistically significant risk factors for CHD. From Table 2 it is clear that these results are robust to the prior specification. The posterior probabilities of the null hypothesis are slightly higher in the 2-sided analysis that allows negative effects (as expected), but the conclusions are consistent.

The estimated posterior probabilities of the null hypothesis in the 1-sided analysis are substantially lower than the 1-sided p-values from the Cox model analysis. Although poste-

rior probabilities and p-values are not directly comparable, this may suggest an improvement in sensitivity for our approach relative to a simple frequentist proportional hazards analysis. One reason for this apparent improvement may be the greater flexibility of the model averaging approach. Although the proportional hazards model has the highest posterior probability for each of the predictors (as shown in Table 2), the results are far from definitive, and substantial probability is assigned to an additive model for the 2nd predictor and to an additive-multiplicative model for predictors 1-3. We anticipate that it is typically true in applications that neither the multiplicative nor the additive models fits perfectly, and by using a combination we may more flexibly characterize the effect of the predictor.

In general, in conducting inferences on the effect of a predictor on the hazard, it may be better to rely on model averaging and to avoid selecting a single best model. However, if results are definitively in favor of an additive or multiplicative effect for a given predictor, then presentation of results can be simplified by relying on the preferred model. In particular, one can report the proportional or additive hazard regression coefficient as a single measure of the effect. For the Framingham study, we cannot conclude, with a high probability, that any one of the covariates has only additive or only multiplicative effects. For this reason, we present effect estimates stratified by age for each of the covariates. In particular, Table 3 presents posterior means and 95% credible intervals for the increase in the probability of developing CHD within a given age interval for individuals with any one of the risk factors relative to individuals with none of the risk factors (i.e., those in the baseline category).

## 6. Discussion

Although there is always uncertainty about the most appropriate form for the regression function for time to event data, proportional hazards models are typically used for convenience. As a different but equally simple formulation, one may consider additive hazards. Estimation and inference for additive hazards models is complicated by the need to place

constraints on the regression coefficients to avoid negative hazard estimates for certain times and values of the predictors. We have developed a novel Bayesian approach for addressing this problem. Although the proposed method for additive hazards regression represents an advance in itself, we also propose an approach that facilitates selection and formally accommodates uncertainty between proportional and additive hazards. Due to the structure of the prior, the full conditional posterior distributions are conjugate after augmentation with Poisson latent variables, simplifying efficient computation.

Based on a small simulation study, our computational algorithm is highly efficient in terms of number of iterations until convergence and rate of mixing of the Markov chain. This efficiency is a byproduct of the carefully tailored prior distribution and the associated conjugacy properties. We also observed excellent frequentist operating characteristics in the cases we have considered thus far. Another appealing feature of our approach is its generalizability. For example, it is straightforward to account for multivariate or correlated event times by including a gamma frailty multiplier in the hazard function. In this case, our MCMC algorithm can be generalized by incorporating simple modifications to the conditional distributions and by adding additional steps for sampling the frailties from their gamma conditional distributions and for updating the frailty variance (e.g., by Metropolis-Hastings or adaptive-rejection sampling).

## ACKNOWLEDGEMENTS

The authors thank Jianwen Cai for providing the data. We also thank Shyamal Peddada and David Umbach for their critical reading of the manuscript, and the anonymous reviewers for comments resulting in substantial improvements in the presentation.

## APPENDIX A

### *Posterior and Full Conditional Posterior Distributions for Additive Hazards Analysis*

The joint posterior density of the parameters and latent variables is proportional to

$$\left\{ \prod_{i=1}^n \prod_{j=1}^J 1 \left( dN_{ij} = dN_{ij0} + \sum_{k=1}^p 1_{(M_k \neq 0)} dN_{ijk} \right) \pi(dN_{ij0} \mid \boldsymbol{\lambda}_0) \prod_{k=1}^p \pi(dN_{ijk} \mid \mathbf{M}, \boldsymbol{\alpha}, \mathbf{X}) \right\} \pi(\boldsymbol{\lambda}_0) \pi(\mathbf{M}) \pi(\boldsymbol{\alpha} \mid \mathbf{M}). \quad (15)$$

**Step 1.** It follows from expression (15) that the full conditional distribution of the latent variables  $(dN_{ij0}, dN_{ij1}, \dots, dN_{ijp})$  is proportional to

$$\begin{aligned} & 1 \left( dN_{ij} = dN_{ij0} + \sum_{k: M_k \neq 0} dN_{ijk} \right) \left\{ \frac{(d\Lambda_{0j})^{dN_{ij0}}}{dN_{ij0}!} \right\} \left\{ \prod_{k=1}^p \frac{\left( (t_j - t_{j-1}) x_{ik}^* \alpha_k \right)^{dN_{ijk}}}{dN_{ijk}!} \right\} \\ & \propto \left\{ \frac{dN_{ij}!}{dN_{ij0}! \prod_{k: M_k \neq 0} dN_{ijk}!} (P_{ij0})^{dN_{ij0}} \prod_{k: M_k \neq 0} (P_{ijk})^{dN_{ijk}} \right\} \left\{ \prod_{k: M_k = 0} \frac{\left( (t_j - t_{j-1}) x_{ik}^* \alpha_k \right)^{dN_{ijk}}}{dN_{ijk}!} \right\} \\ & \propto \text{Multinomial} \left( \{dN_{ij0}, dN_{ijk} : M_k \neq 0\}; dN_{ij}, \{P_{ij0}, P_{ijk} : M_k \neq 0\} \right) \\ & \quad \times \prod_{k: M_k = 0}^p \text{Poisson} \left( dN_{ijk}; (t_j - t_{j-1}) x_{ik}^* \alpha_k \right), \end{aligned} \quad (16)$$

where  $P_{ij0}$  and  $P_{ijk}$  are defined in subsection 2.3.

**Step 2.** The full conditional distribution of  $\lambda_{0h}$ ,  $h = 1, \dots, H$ , is proportional to

$$\begin{aligned} & \left\{ \prod_{i,j: Y_{ij}=1} (d\Lambda_{0j})^{dN_{ij0}} \exp(-d\Lambda_{0j}) \right\} (\lambda_{0h})^{a_{0h}-1} \exp(-\lambda_{0h} b_{0h}) \\ & \propto \left\{ \prod_{i,j: Y_{ij}=1} \left( \sum_{l=1}^H \lambda_{0l} (t_j - t_{j-1}) 1_{(t_j \in (s_{l-1}, s_l])} \right)^{dN_{ij0}} \exp \left( - \sum_{l=1}^H \lambda_{0l} (t_j - t_{j-1}) 1_{(t_j \in (s_{l-1}, s_l])} \right) \right\} \\ & \quad (\lambda_{0h})^{a_{0h}-1} \exp(-\lambda_{0h} b_{0h}) \\ & \propto (\lambda_{0h})^{\sum_{i,j: Y_{ij}=1, t_j \in (s_{h-1}, s_h]} dN_{ij0}} \exp \left( - \lambda_{0h} \sum_{i,j: t_j \in (s_{h-1}, s_h]} Y_{ij} (t_j - t_{j-1}) \right) \lambda_{0h}^{a_{0h}-1} \exp(-\lambda_{0h} b_{0h}) \\ & \propto \mathcal{G} \left( \lambda_{0h}; a_{0h} + \sum_{i=1}^n \sum_{j: t_j \in (s_{h-1}, s_h]} J dN_{ij0}, b_{0h} + \sum_{i=1}^n \sum_{j: t_j \in (s_{h-1}, s_h]} J Y_{ij} (t_j - t_{j-1}) \right). \end{aligned} \quad (17)$$

**Step 3.** To derive the conditional distribution of the model indicator,  $M_k$ , we start with the joint posterior density of the parameters prior to augmentation, which is propor-

tional to

$$\begin{aligned}
& \left[ \prod_{i,j:Y_{ij}=1} \left( d\Lambda_{0j} + (t_j - t_{j-1}) \sum_{k=1}^p x_{ik}^* \alpha_k \right)^{dN_{ij}} \exp \left\{ - \left( d\Lambda_{0j} + (t_j - t_{j-1}) \sum_{k=1}^p x_{ik}^* \alpha_k \right) \right\} \right] \\
& \quad \pi(\boldsymbol{\lambda}_0) \pi(\mathbf{M}) \pi(\boldsymbol{\alpha} \mid \mathbf{M}) \\
& \propto \left( p_{k,-1}^+ \prod_{i,j:Y_{ij}=1} \text{Poisson} \left( dN_{ij}; d\Lambda_{0j} + (t_j - t_{j-1}) \{ \mathbf{x}_{i(-k)}^{*'} \boldsymbol{\alpha}_{(-k)} + (1 - x_{ik}) \alpha_k \} \right) \right)^{1(M_k=-1)} \\
& \times \left( p_{k,0}^+ \prod_{i,j:Y_{ij}=1} \text{Poisson} \left( dN_{ij}; d\Lambda_{0j} + (t_j - t_{j-1}) \mathbf{x}_{i(-k)}^{*'} \boldsymbol{\alpha}_{(-k)} \right) \right)^{1(M_k=0)} \\
& \times \left( p_{k,1}^+ \prod_{i,j:Y_{ij}=1} \text{Poisson} \left( dN_{ij}; d\Lambda_{0j} + (t_j - t_{j-1}) \{ \mathbf{x}_{i(-k)}^{*'} \boldsymbol{\alpha}_{(-k)} + x_{ik} \alpha_k \} \right) \right)^{1(M_k=1)} \quad (18)
\end{aligned}$$

which is Multinomial( $M_k; \{-1, 0, 1\}, \tilde{\boldsymbol{\pi}}_k$ ), with  $\tilde{\pi}_{kl}$  equal to the  $l$ th element in the product (18) standardized so that  $\sum_{l=1}^3 \tilde{\pi}_{kl} = 1$ .

**Step 4.** For  $M_k = 0$ , the posterior distribution of  $\alpha_k$  is not defined and we set  $\alpha_k$  equal to its previous value. For  $M_k \neq 0$ , the full conditional posterior distribution is

$$\begin{aligned}
& \propto (\alpha_k)^{a_k^+ - 1} \exp(-\alpha_k b_k^+) \prod_{i,j:Y_{ij}=1} \left\{ (t_j - t_{j-1}) x_{ik}^* \alpha_k \right\}^{dN_{ijk}} \exp \left\{ - (t_j - t_{j-1}) x_{ik}^* \alpha_k \right\} \\
& \propto (\alpha_k)^{a_k^+ - 1 + \sum_{i,j:Y_{ij}=1} dN_{ijk}} \exp \left[ - \alpha_k \left\{ b_k^+ + \sum_{i,j:Y_{ij}=1} (t_j - t_{j-1}) x_{ik}^* \right\} \right] \\
& \propto \mathcal{G} \left( \alpha_k; a_k^+ + \sum_{i,j:Y_{ij}=1} dN_{ijk}, b_k^+ + \sum_{i,j:Y_{ij}=1} (t_j - t_{j-1}) x_{ik}^* \right). \quad (19)
\end{aligned}$$

## APPENDIX B

### *Gibbs Sampler for Additive-Multiplicative Hazards Analysis*

The joint posterior density is the same as in the additive hazards case described in Appendix A, but we multiply by the prior density for the proportional hazard coefficients,  $\boldsymbol{\beta}$ , and condition on  $\boldsymbol{\beta}$  in the prior for  $dN_{ij0}$ . We focus on binary predictors,  $x_{ik}$ , for simplicity.

**Step 1.** Proceeds as in Appendix A, but with  $d\Lambda_{0j}$  replaced by  $d\Lambda_{0j} \exp(\mathbf{x}'_i \boldsymbol{\beta})$  and with the latent variable density in (14) used in place of (11).

**Step 2.** The full conditional distribution of  $\lambda_{0h}$  is as shown in Appendix A, but with  $Y_{ij} \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})$  substituted for  $Y_{ij}$  in the gamma conditional distribution.

**Step 3.** As shown in Appendix A, but with  $d\Lambda_{0j}$  replaced by  $d\Lambda_{0j} \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})$ .

**Step 4.** As shown in Appendix A.

**Step 5.** Sample  $\gamma_k = \exp(\beta_k)$  from its full conditional posterior distribution,

$$\pi(\gamma_k | \mathbf{N}, \mathbf{Y}, \boldsymbol{\lambda}_0, \mathbf{M}, \boldsymbol{\alpha}, \boldsymbol{\gamma}_{(-k)}) = \mathbb{I}_{1-\mathcal{G}_{\mathcal{A}_k}}(\gamma_k; \tilde{\pi}_k^\times, \tilde{a}_k^\times, \tilde{b}_k^\times),$$

where the conditional posterior probability of  $\gamma_k = 1$  is

$$\tilde{\pi}_k^\times = \frac{p_{k,0}^\times \exp \left\{ - \sum_{i,j} x_{ik} Y_{ij} d\Lambda_{0j} \exp(\mathbf{x}_{i(-k)} \boldsymbol{\beta}_{(-k)}) \right\}}{p_{k,0}^\times \exp \left\{ - \sum_{i,j} x_{ik} Y_{ij} d\Lambda_{0j} \exp(\mathbf{x}_{i(-k)} \boldsymbol{\beta}_{(-k)}) \right\} + (1 - p_{k,0}^\times) \frac{C(a_k^\times, b_k^\times) P_{\mathcal{A}_k}(\tilde{a}_k^\times, \tilde{b}_k^\times)}{C(\tilde{a}_k^\times, \tilde{b}_k^\times) P_{\mathcal{A}_k}(a_k^\times, b_k^\times)}},$$

where  $C(a, b) = a^b / \Gamma(a)$ ,  $\tilde{a}_k^\times = a_k^\times + \sum_{i,j} x_{ik} dN_{ij0}$ ,  $P_{\mathcal{A}_k}(a, b) = \int_{\mathcal{A}_k} \mathcal{G}(\gamma; a, b) d\gamma$ , and

$$\tilde{b}_k^\times = b_k^\times + \sum_{i,j} x_{ik} Y_{ij} d\Lambda_{0j} \exp(\mathbf{x}'_{i(-k)} \boldsymbol{\beta}_{(-k)}).$$

Note that this conditional distribution is conjugate, and sampling can proceed by first sampling from the point mass at one using  $1_{(\gamma_k=1)} \sim \text{Bernoulli}(\tilde{\pi}_k^\times)$  and then sampling from  $\mathcal{G}(\tilde{a}_k^\times, \tilde{b}_k^\times)$  if  $\gamma_k \neq 1$ .

## REFERENCES

- Aalen, O.O. (1980). A model for non-parametric regression analysis of counting processes. In *Lecture Notes in Statistics-2: Mathematical Statistics and Probability Theory*, W. Klonecki, A. Kozek and J. Rosinski (eds), 1-25.
- Beamonte, E. and Bermúdez, J.D. (2003). A Bayesian semiparametric analysis for additive hazard models with censored observations. *Test*, in press (available for download at [www.uv.es/~beamonte/investigacion/test.pdf](http://www.uv.es/~beamonte/investigacion/test.pdf)).

- Chipman, H, George, E.I. and McCulloch, R.E. (2001). The practical implementation of Bayesian model selection. *IMS Lecture Notes - Monograph Series Volume 38*.
- Clayton, D. (1994). Bayesian analysis of frailty models. Technical Report, Medical Research Council Biostatistics Unit, Cambridge.
- Cox, D.R. (1972). Regression models and life-tables (with discussions). *Journal of the Royal Statistical Society: B* **34**, 187-202.
- Cox, D.R. and Oakes, D.A. *Analysis of Survival Data*. London: Chapman & Hall.
- Dawber, T.R., Meadurs, G.F., and Moore, F.E.J. (1951). Epidemiological approaches to heart disease: The Framingham study. *American Journal of Public Health* **41**, 279-286.
- Dunson, D.B. and Herring, A.H. (2003). Bayesian inferences in the Cox model for order restricted hypotheses. *Biometrics* **59**, 918-925.
- Faraggi, D. and Simon, R. (1998). Bayesian variable selection method for censored survival data. *Biometrics* **54**, 1475-1485.
- George, E.I. and McCulloch, R.E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7**, 339-373.
- Geweke, J. (1996). Variable selection and model comparison in regression. In *Bayesian Statistics 5*, J.O. Berger, J.M. Bernardo, A.P. Dawid, and A.F.M. Smith (eds.). Oxford University Press, 609-620.
- Kass, R.E. and Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773-795.
- Kim, S.W. and Sun, D.C. (2000). Intrinsic priors for model selection using an encompassing model with applications to censored failure time data. *Lifetime Data Analysis* **6**, 251-269.

- Ibrahim, J.G., Chen, M.-H., and MacEachern, S.N. (1999). Bayesian variable selection for proportional hazards models. *Canadian Journal of Statistics* **27**, 701-717.
- Ibrahim, J.G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer.
- Ibrahim, J.G., Chen, M.-H., and Sinha, D. (2001). Criterion-based methods for Bayesian model assessment. *Statistica Sinica* **11**, 419-443.
- Lin, D.Y. and Ying, Z.L. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.
- Lin, D.Y. and Ying, Z.L. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *Annals of Statistics* **23**, 1712-1734.
- Martinussen, T. and Scheike, T.H. (2002). A flexible additive multiplicative hazard model. *Biometrika* **89**, 283-298.
- McKeague, I.W. and Sasieni, P.D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501-514.
- Scheike, T.H. and Zhang, M.J. (2002). An additive-multiplicative Cox-Aalen regression model. *Scandinavian Journal of Statistics* **29**, 75-88.
- Sinha, D., Chen M.-H., and Ghosh, S.K. (1999). Bayesian analysis and model selection for interval-censored survival data. *Biometrics* **55**, 585-590.
- Volinsky, C.T., Madigan, D., Raftery, A.E. and Kronmal, R.A. (1997). Bayesian model averaging in proportional hazards: assessing the risk of a stroke. *Applied Statistics* **46**, 433-448.

Volinsky, C.T. and Raftery, A.E. (2000). Bayesian information criterion for censored survival models. *Biometrics* **56**, 256-262.

**Table 1**

*Results from simulation study of the frequentist operating characteristics of the proposed Bayesian approach.*

Model	Parameter	True Value	$\hat{\theta}$	$\widehat{RR}$	$\widehat{\pi}_{k,0}$	Coverage
1	$\alpha_1$	0.03	0.03	0.84	0.03	0.92
1	$\alpha_2$	0.00	0.00	0.04	0.56	0.96
1	$\alpha_3$	0.00	0.00	0.02	0.58	1.00
1	$\alpha_4$	0.00	0.00	0.00	0.61	1.00
1	$\gamma_1$	1.00	1.00	0.04	0.52	1.00
1	$\gamma_2$	1.00	1.02	0.06	0.42	0.96
1	$\gamma_3$	1.00	1.02	0.04	0.44	1.00
1	$\gamma_4$	1.00	1.01	0.10	0.40	1.00
2	$\alpha_1$	0.00	0.00	0.02	0.58	0.98
2	$\alpha_2$	0.00	0.00	0.00	0.65	1.00
2	$\alpha_3$	0.00	0.00	0.00	0.64	1.00
2	$\alpha_4$	0.00	0.00	0.00	0.63	1.00
2	$\gamma_1$	2.00	2.02	0.96	0.01	0.98
2	$\gamma_2$	1.00	1.09	0.04	0.41	1.00
2	$\gamma_3$	1.00	1.09	0.06	0.44	1.00
2	$\gamma_4$	1.00	1.09	0.06	0.45	0.98
3	$\alpha_1$	0.03	0.03	0.84	0.04	0.88
3	$\alpha_2$	0.00	0.00	0.04	0.53	0.98
3	$\alpha_3$	0.00	0.00	0.02	0.60	0.98
3	$\alpha_4$	0.00	0.00	0.02	0.58	0.98
3	$\gamma_1$	1.00	1.17	0.04	0.43	0.98
3	$\gamma_2$	2.00	1.90	0.84	0.04	0.98
3	$\gamma_3$	1.00	1.12	0.04	0.44	0.98
3	$\gamma_4$	1.00	1.13	0.06	0.43	0.96
4	$\alpha_1$	0.00	0.00	0.02	0.59	1.00
4	$\alpha_2$	0.00	0.00	0.00	0.58	1.00
4	$\alpha_3$	0.00	0.00	0.00	0.63	1.00
4	$\alpha_4$	0.00	0.00	0.02	0.62	1.00
4	$\gamma_1$	1.00	1.09	0.00	0.45	1.00
4	$\gamma_2$	1.00	1.08	0.02	0.48	1.00
4	$\gamma_3$	1.00	1.06	0.04	0.49	1.00
4	$\gamma_4$	1.00	1.11	0.06	0.48	0.96

$\hat{\theta}$  = Model-averaged posterior median

$\widehat{RR}$  = Empirical rejection rate based on  $\leq 0.05$  cutoff

$\widehat{\pi}_{k,0}$  = Average of Rao-Blackwellized estimator

Coverage = coverage of 95% credible interval

**Table 2**

*Estimated posterior probabilities of only additive ( $H_{1k}^+ \cap H_{0k}^\times$ ), only multiplicative ( $H_{0k}^+ \cap H_{1k}^\times$ ), both additive and multiplicative ( $H_{1k}^+ \cap H_{1k}^\times$ ), and no association ( $H_{0k}$ ) for the predictors: (1) hypertension, (2) above normal cholesterol, (3) overweight, and (4) obese.*

Predictor	1 or 2-sided <sup>†</sup>	Posterior Probabilities			
		$H_{1k}^+ \cap H_{0k}^\times$	$H_{0k}^+ \cap H_{1k}^\times$	$H_{1k}^+ \cap H_{1k}^\times$	$H_{0k}$
1	1	0.019	0.799	0.166	0.016 [0.015-0.016] <sup>‡</sup>
	2	0.024	0.620	0.328	0.028
2	1	0.225	0.422	0.339	0.014 [0.008-0.015]
	2	0.200	0.472	0.285	0.043
3	1	0.054	0.674	0.258	0.014 [0.013-0.019]
	2	0.104	0.467	0.411	0.018
4	1	0.022	0.833	0.054	0.091 [0.080-0.095]
	2	0.074	0.385	0.437	0.104

<sup>†</sup>1=1-sided ( $M_k \in \{0, 1\}$ ) or 2-sided ( $M_k \in \{-1, 0, 1\}$ )

<sup>‡</sup>Range of values in sensitivity analyses for priors (i) - (iii)

**Table 3**

*Estimated difference between individuals with a given risk factor and those in the baseline category with respect to the probability of developing CHD.*

Age	Hypertension		Cholesterol		Overweight		Obese <sup>†</sup>	
	Mean	95%CI	Mean	95%CI	Mean	95%CI	Mean	95%CI
(45, 50]	0.009	[0.000,0.034]	0.015	[0.001,0.037]	0.008	[0.000,0.026]	0.009	[0,0.034]
(50, 55]	0.023	[0.001,0.064]	0.030	[0.003,0.069]	0.019	[0.001,0.049]	0.022	[0,0.070]
(55, 60]	0.042	[0.002,0.094]	0.045	[0.005,0.090]	0.033	[0.003,0.070]	0.041	[0,0.113]
(60, 65]	0.055	[0.003,0.112]	0.053	[0.007,0.099]	0.043	[0.004,0.085]	0.048	[0,0.123]
(65, 71]	0.062	[0.004,0.120]	0.057	[0.009,0.103]	0.049	[0.004,0.098]	0.047	[0,0.106]

<sup>†</sup>for obesity, the reported difference is relative to overweight, but not obese, women