Published in final edited form as:
Comput Stat Data Anal. 2014 July ; 75: 53–65. doi:10.1016/j.csda.2014.02.009.

Bayesian Variable Selection under the Proportional Hazards Mixed-effects Model

Kyeong Eun Lee,
Department of Statistics, Kyungpook National University, Daegu, 702-701, Korea

Yongku Kim, and
Department of Statistics, Kyungpook National University, Daegu, 702-701, Korea

Ronghui Xu*
Division of Biostatistics and Bioinformatics, Department of Family and Preventive Medicine and Department of Mathematics, University of California, San Diego, USA

Abstract

Over the past decade much statistical research has been carried out to develop models for correlated survival data; however, methods for model selection are still very limited. A stochastic search variable selection (SSVS) approach under the proportional hazards mixed-effects model (PHMM) is developed. The SSVS method has previously been applied to linear and generalized linear mixed models, and to the proportional hazards model with high dimensional data. Because the method has mainly been developed for hierarchical normal mixture distributions, it operates on the linear predictor under the Cox type models. The PHMM naturally incorporates the normal distribution via the random effects, which enables SSVS to efficiently search through the candidate variable space. The approach was evaluated through simulation, and applied to a multi-center lung cancer clinical trial data set, for which the variable selection problem was previously debated upon in the literature.

Keywords

correlated survival data; MCMC; model selection; multi-center clinical trial; proportional hazards mixed-effects model; stochastic search variable selection

1 Introduction

Correlated survival data arise in various practical applications including multi-center clinical trials, genetic studies, and recurrent events. In many such applications the data consist of clusters and observations within the clusters. A number of statistical methods have been

* Corresponding author: 9500 Gilman Drive, La Jolla, CA 92093-0112, USA. rxu@ucsd.edu, tel: (858) 534-6380, fax: (858) 534-5273..

developed over the last decade to analyze such data. The proportional hazards mixed-effects model (PHMM) was proposed by Ripatti and Palmgren (2000) and Vaida and Xu (2000) to model clustered survival data, which allows cluster specific random effects of arbitrary covariates. Suppose that T_{ij} is the random variable representing the failure time of individual j in cluster i. The PHMM assumes that the hazard function of T_{ij} follows

$$\lambda_{ij}(t) = \lambda_0(t)\ exp\left(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{b}_i\right), \quad (1)$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects, $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ is a $q \times 1$ vector of cluster specific random effects, $\mathbf{x}_{ij}$ is a $p \times 1$ vector of covariates, and $\mathbf{z}_{ij}$ is typically a $q \times 1$ subvector of $\mathbf{x}_{ij}$, except that $\mathbf{z}_{ij}$ is allowed to contain an element of '1' for a random cluster effect on the baseline hazard.

Under model (1) various inference procedures have been proposed in the literature. Ripatti and Palmgren (2000) considered a penalized partial likelihood approach, which is similar to the penalized quasi-likelihood (PQL) under the generalized linear mixed models. Vaida and Xu (2000) proposed a nonparametric maximum likelihood estimator (NPMLE), obtained using a Monte Carlo EM algorithm. Cortiñas-Abrahantes *et al.* (2007) considered a Laplace EM algorithm for the NPMLE. A comprehensive comparison of these methods can be found in Gamst *et al.* (2009). Although it is reasonably clear to see the advantages and limitations of the different inference procedures, only very recently attention has started to focus on model selection. Under model (1) this concerns the selection of fixed as well as random effects.

Xu *et al.* (2009) considered the likelihood ratio test under model (1), as well as a profile Akaike information criterion for model selection. Donohue *et al.* (2011) developed a conditional Akaike information criterion, where the focus is on the estimation of the fixed as well as the random effects. Under the special case of frailty models where $\mathbf{z}_{ij}$ is restricted to either 0 or 1, Fan and Li (2002) considered selection of the fixed effects. Gray (1995) and Commenges and Andersen (1995) developed score tests for no random effects in the frailty model, although it is also possible to generalize the score tests to test for no random effects of additional covariates under model (1) via stratification (Gray, 2006). Dunson and Chen (2004) also considered selection of random effects under the gamma frailty model, using a Bayesian approach. Interestingly Dunson and Chen (2004) arrived at a different conclusion from the score tests of Gray (1995), on the data from a multi-center clinical trial in lung cancer, which will be further discussed in this paper.

Stochastic search variable selection (George and McCulloch, 1993, SSVS) is an approach based on the Bayesian hierarchical normal mixture setup under a regression model, where latent variables are used to indicate the inclusion or exclusion of a potential predictor. It uses Gibbs sampler to sample from a multinomial distribution on the set of possible subset choices, and the promising subsets of predictors are identified as those with high posterior probabilities. As will be described below, SSVS avoids the overwhelming problem of calculating the posterior probabilities of all 2^p subsets, and is computationally fast and efficient. The SSVS method has been extended to linear and generalized linear mixed models (Chen and Dunson, 2003; Kinney and Dunson, 2007), and to survival models (Lee

and Mallick, 2004). Because of its ability to select among a larger number of potential predictors, it has been applied to high dimensional data including genomics and other complex disease risk factor studies (Beattie *et al.*, 2002; Lee *et al.*, 2003; Swartz *et al.*, 2008; Lin and Huang, 2008).

In the following we develop the SSVS under the general PHMM (1), for selection of both fixed and random effects of arbitrary covariates. There has been no Bayesian approach to this problem in the literature, which has the advantage of subsequent model averaging that can take into account model uncertainty and selection bias. In Section 3 we examine the performance of SSVS using simulations. We apply the approach to the multi-center lung cancer clinical trial data set that was previously analyzed in Gray (1995) and Dunson and Chen (2004) in Section 4. The last section contains further discussion, and all the posterior computation details are given in the Appendix.

2 Variable Selection under the PHMM

For clusters $i = 1, \cdots, n$, and observations $j = 1, \cdots, n_i$, denote t_{ij} the observed, possibly right-censored failure time, $\delta_{ij} = 1$ if t_{ij} is an observed failure time, and 0 otherwise. Let N be the total number of observations, that is, $N = \sum_{i=1}^{n} n_i$.

Under model (1) $\mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{\mathbf{ij}}^{'}\mathbf{b}_{\mathbf{i}}$ is the linear predictor, or the prognostic index, which determines the relative risk of an individual. It is an intermediate quantity analogous to the response in a linear model, which in this case associates the predictors with the ultimate survival outcome. Since the SSVS was initially developed for the hierarchical normal mixture distributions, Lee and Mallick (2004) considered adding a small random quantity $\varepsilon_{ij} \sim N(0, \sigma^2)$ to the linear predictor. The resulting model is then

$$\lambda_{ij}(t) = \lambda_0(t)\ exp\left(\mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{\mathbf{ij}}^{'}\mathbf{b}_{\mathbf{i}} + \epsilon_{\mathbf{ij}}\right). \quad (2)$$

The ε_{ij}'s may be viewed as an individual heterogeneity term which can improve the fit of the model to the data (O'Quigley and Stare, 2002). But the consideration here is mainly computational, because it simplifies the posterior computation as described below and allows the Gibbs sampler to efficiently search through the model space. We should still consider data as generated under model (1), while model (2) is a working model; this is also reflected in our simulations later: while data were generated under model (1), we follow the approach described below to do variable selection and estimation. We should mention that the identifiability of model (2) is similar to the individual frailty models considered in Kosorok *et al.* (2001), and can also be more intuitively seen from the equivalent transformation model formulation: $h(T_{ij}) = -\mathbf{x}_{ij}^{'}\boldsymbol{\beta} - \mathbf{z}_{\mathbf{ij}}^{'}\mathbf{b}_{\mathbf{i}} + \mathbf{e}_{\mathbf{ij}}$, where $e_{ij} = e_{0ij} - \varepsilon_{ij}$, e_{0ij} has a fixed, known extreme value distribution with $\mathrm{Var}(e_{0ij}) = 1.645$, and $h(t) = \log \Lambda_0(t)$ where $\Lambda_0(t) = \int_0^t \lambda_0(s)\, ds$ is the cumulative baseline hazard function.

For notational purposes, let $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{in_i})'$, $\mathbf{Z}_i = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, \ldots, \mathbf{z}_{in_i})'$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in_i})'$ for $i = 1, 2, \ldots, n$. Also let $\mathbf{X} = \left(\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_n'\right)'$, $\mathbf{Z} = \mathrm{diag}\{\mathbf{Z}_1, \mathbf{Z}_2, \ldots,$

$\mathbf{Z}_n\}$, $\mathbf{b}=\left(\mathbf{b}_1^{'},\mathbf{b}_2^{'},\ldots,\mathbf{b}_n^{'}\right)^{'}$, and $\epsilon=\left(\epsilon_1^{'},\epsilon_2^{'},\ldots,\epsilon_n^{'}\right)^{'}$. Finally let $W_{ij}=\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+\mathbf{z}_{ij}^{'}\mathbf{b}_i+\epsilon_{ij}$, $\mathbf{W}$ = $(W_{11}, W_{12}, \ldots, W_{nn_n})'$, $\mathbf{t}$ = $(t_{11}, \ldots, t_{nn_n})'$, $\boldsymbol{\delta}$ = $(\delta_{11}, \ldots, \delta_{nn_n})'$, and $\mathbf{Y}$ = $(\mathbf{t}, \boldsymbol{\delta})$ which denotes the observed survival data. Then we have:

$$\mathbf{W}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{Z}\mathbf{b}=\epsilon,\quad \epsilon\sim\mathbf{N}\left(0,\sigma^2\mathbf{I}_\mathbf{N}\right),\quad \mathbf{b}\sim\mathbf{N}\left(0,\mathbf{I}_\mathbf{n}\otimes\boldsymbol{\Sigma}\right), \quad (3)$$

where $\boldsymbol{\Sigma}$ is positive semi-definite as it may include variance components that should be excluded from the final selected models, $\otimes$ denotes the Kronecker product, and $\mathbf{I}_n$ denotes an $n \times n$ identity matrix.

The SSVS uses latent binary variables $\boldsymbol{\gamma} = (\gamma_1, \ldots \gamma_p)'$ to indicate the inclusion or exclusion of a fixed effect: $\gamma_k = 1$ if $\beta_k \neq 0$ and 0 otherwise, $k = 1, \ldots, p$. Given $\boldsymbol{\gamma}$, let $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ consist of all nonzero elements of $\boldsymbol{\beta}$, and let $\mathbf{X}_{\boldsymbol{\gamma}}$ be the columns of $\mathbf{X}$ corresponding to the elements of $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$. After specifying the prior distribution for $\boldsymbol{\gamma}, \boldsymbol{\beta}_{\boldsymbol{\gamma}}$ and other parameters, one uses the observed data likelihood and Markov chain Monte Carlo (MCMC) to sample from the posterior distribution of $\boldsymbol{\gamma} = (\gamma_1, \ldots \gamma_p)'$. After burn-in, i.e. convergence of the MCMC chain, denote $s = 1, \ldots, S$ the MCMC samples. Then the marginal probability of inclusion for each fixed effect can be estimated by

$$\hat{\pi}_k=\frac{1}{S}\sum_{s=1}^{S}\gamma_k^{(s)}, \quad (4)$$

where $\gamma_k^{(s)}$ is the sampled value of γ_k in iteration s, $k = 1, \ldots, p$. For selection of the random effects, a re-parameterization of the covariance matrix $\boldsymbol{\Sigma}$ is applied, so that whether $a_l = 0$ ($l = 1, \ldots, q$, see below) plays the same role as the γ_k's above. This is described in full details in the following.

2.1 Prior specification

The priors for $\boldsymbol{\gamma}, \boldsymbol{\beta}_{\boldsymbol{\gamma}}$ and σ^2 are:

$$\gamma_k\sim\text{Bernoulli}\left(\pi_k\right),\quad k=1,\ldots,p; \quad (5)$$

$$\boldsymbol{\beta}_\gamma|\boldsymbol{\gamma}\sim\mathbf{N}\left(\mathbf{0},\sigma^2\left(\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma\right)^{-1}/\mathbf{g}\right); \quad (6)$$

$$g\sim\text{Gamma}\left(1/2,N/2\right);$$

$$\sigma^2\sim\frac{1}{\sigma^2}. \quad (8)$$

In the above, the γ_k's are assumed to be a priori independent with $P(\gamma_k = 1) = \pi_k$, $0 \leq \pi_k \leq 1$, for $k = 1, \ldots, p$. In practice we may take $\pi_k = 0.5$ if there is no prior knowledge about whether a fixed effect should be included, and we consider this as the 'non-informative'

prior; or we may take $\pi_k = 1$ if we want to force a fixed effect into the model. When all $\pi_k = 0.5$ ($k = 1, ..., p$), it is clear that each model γ for the fixed effects has a prior probability equal to 2^{-p}. The prior variance of β is taken to be proportional to $\sigma^2\left(\mathbf{X}_\gamma'\mathbf{X}_\gamma\right)^{-1}$, as it results in a fast computing algorithm for the Gibbs sampler; this is also called Zellner's g-prior (Zellner, 1986; Smith and Kohn, 1996). Finally, the improper prior for σ^2 is commonly used such that $\log(\sigma^2)$ is uniform.

To specify the priors for the variance components, Chen and Dunson (2003) considered a modified Cholesky Decomposition of $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \boldsymbol{\Psi}\boldsymbol{\Omega}\boldsymbol{\Omega}'\boldsymbol{\Psi}, \quad (9)$$

where $\boldsymbol{\Psi} = \text{diag}(\psi_1, \ldots, \psi_q)$, and $\boldsymbol{\Omega}$ is a lower triangular matrix with diagonal elements equal to 1. When $\psi_l = 0$ in $\boldsymbol{\Psi}$, the l-th diagonal element of $\boldsymbol{\Sigma}$ is also equal to 0, implying that the l-th random effect is excluded from model (1). The off-diagonal elements of $\boldsymbol{\Omega}$, denoted by ω, represent the dependency among the random effects. Using decomposition (9) we have $\mathbf{W} = \mathbf{X}_\gamma\beta_\gamma + \boldsymbol{Z}\left(\mathbf{I}_n \otimes \boldsymbol{\Psi}\boldsymbol{\Omega}\right)\mathbf{a} + \epsilon$, where $\mathbf{a} = \left(\mathbf{a}_1', \mathbf{a}_2', \ldots, \mathbf{a}_n'\right)'$, $\mathbf{a_i} \sim N\left(\mathbf{0}, \mathbf{I_q}\right)$. Kinney and Dunson (2007) further considered the parameter-expansion (PX) approach of Gelman (2006) for variance components. The over-parameterization in PX reduces dependence among the parameters in a hierarchical model and improves the Gibbs convergence (Liu *et al.*, 1998). Using the PX approach (3) becomes

$$\mathbf{W} = \mathbf{X}_\gamma\beta_\gamma + \boldsymbol{Z}\left(\mathbf{I}_n \otimes \mathbf{A}\boldsymbol{\Omega}\right)\boldsymbol{\xi} + \epsilon, \quad \epsilon \sim \mathbf{N}\left(\mathbf{0}, \sigma^2\mathbf{I_N}\right), \quad \boldsymbol{\xi_i} \sim \mathbf{N}\left(\mathbf{0}, \mathbf{D}\right), \quad (10)$$

where $\mathbf{A} = \text{diag}(\alpha_1, \ldots, \alpha_q)$, $\mathbf{D} = \text{diag}(d_1, \ldots, d_q)$, and $\boldsymbol{\xi} = \left(\boldsymbol{\xi}_1', \boldsymbol{\xi}_2', \ldots, \boldsymbol{\xi}_n'\right)'$. Following Kinney and Dunson (2007) the priors are:

$$\alpha_l \sim ZI\text{-}\,\mathrm{N}^+\left(0, 1, p_{l0}\right), \quad l = 1, \ldots, q; \quad (11)$$

$$\boldsymbol{\omega} \sim \mathbf{N}\left(\boldsymbol{\omega_0}, \mathbf{V}_\omega\right); \quad (12)$$

$$d_l \sim IG\left(\frac{1}{2}, \frac{N}{2}\right), \quad (13)$$

where ZI-$N^+(0, 1, p_{l0})$ represents the mixture distribution putting point mass p_{l0} on $\alpha_l = 0$, and probability $1 - p_{l0}$ on $N^+(0, 1)$ which is the positive part of N(0,1), and 'IG' denotes inverse Gamma. Just like for the fixed effects β, we can set the hyperparameters $p_{l0} = 0.5$ for equal prior probabilities to include or exclude a random effect, hence the 'non-informative' prior; or we can set $p_{l0} = 0$ to force a random effect in the model. For the other hyperparameters we set $\boldsymbol{\omega_0} = \mathbf{0}$ and $\mathbf{V}_{\boldsymbol{\omega}} = 0.5\mathbf{I}$.

Finally for the baseline cumulative hazard function $\Lambda_0(t)$ we consider a Gamma process (GP) prior (Kalbfleisch, 1978; Clayton, 1991; Ibrahim *et al.*, 2001):

$$\Lambda_0(t) \sim GP(a\Lambda^*(t), a), \quad (14)$$

where Λ^* is the mean process, and the variance of $\Lambda_0(t)$ is given by $\Lambda^*(t)/a$. When there are no random effects in the proportional hazards model and a is close to zero, the resulting marginal posterior of $\boldsymbol{\beta}$ is approximately proportional to the partial likelihood of Cox (1975), while as $a \to \infty$ the Gamma process is effectively replaced by Λ^*, and it becomes the likelihood function of ($\boldsymbol{\beta}$, Λ^*) (Ibrahim *et al.*, 2001). Typically Λ^* is assumed to be a known cumulative baseline hazard function with hyperparameters, and $\lambda^* = d\Lambda^*/dt$ denotes its corresponding baseline hazard function. In this paper we mostly take $\Lambda^*(t) = \eta t^{\kappa}$ from the Weibull distribution, and we can estimate the hyperparameters η and κ from the data by fitting a Weibull regression model including all covariates. Figure 1 plots the pointwise 5th and 95th quantiles of the Gamma process based on the lung cancer data, with various values of a. Following Lee and Mallick (2004) we use $a = 10$ for the rest of the paper.

2.2 The likelihood and posterior computation

Conditional on the random effects, we can integrate out $\Lambda_0(t) \sim G(a\Lambda(t), a)$ at each t, and obtain the likelihood of the survival data $\mathbf{Y}$ marginalized over the prior distribution of the baseline hazard function (Lee and Mallick, 2004). The resulting likelihood is

$$p(\mathbf{Y}|\mathbf{W},\boldsymbol{\theta}) = exp\left(-\sum_{i=1}^{n}\sum_{j=1}^{n_i} aB_{ij}\Lambda^*(t_{ij})\right)\prod_{i=1}^{n}\prod_{j=1}^{n_i}\{a\lambda^*(t_{ij})B_{ij}\}^{\delta_{ij}}, \quad (15)$$

where $B_{ij} = -\log\{1-\exp(W_{ij})/(a+A_{ij})\}$, $A_{ij} = \sum_{kl \in R(t_{ij})} exp(W_{kl})$, and $R(t_{ij})$ is the set of individuals at risk at time t_{ij} – ($j = 1, \ldots, n_i$; $i = 1, \ldots, n$). If we let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \boldsymbol{\omega}', \mathbf{d}', \sigma^2)'$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)'$ and $\mathbf{d} = (d_1, \ldots, d_q)'$, the above likelihood involves $\boldsymbol{\theta}$ only through $\mathbf{W}$.

We can obtain the posterior distribution of interest by

$$p(\boldsymbol{\theta}, \mathbf{W}|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\mathbf{W},\boldsymbol{\theta})\, p(\mathbf{W}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}). \quad (16)$$

As mentioned before $\mathbf{W}$ is an intermediate quantity that associates the predictors with the survival outcome, and here it is viewed more like a parameter in the posterior computation.

To draw inferences about all the parameters of interest as well as model selection, Gibbs samplers or Metropolis-within-Gibbs algorithms are typically implemented. To compute the model posterior distribution, we consider the composite parameter space method of Green and O'Hagan (1998), and tailor it to the context of candidate models with fixed and random effect structures. During an iteration of the procedure, parameters belonging to one part of the model are updated using a standard method, such as a Gibbs or Metropolis-Hastings step, while the other parameters are left unchanged. Our scheme moves around among the indicators for the fixed and the random effects, and the parameters for the fixed and the random effects, as detailed below:

1. Move from a selection of the fixed effects to the next selection of fixed effects by a standard MCMC step. The selection of fixed effects is indexed by the latent binary variables $\boldsymbol{\gamma} = (\gamma_1, \ldots \gamma_p)'$ to indicate the inclusion or exclusion of a fixed effect.
2. Update all fixed effect parameters by a standard MCMC procedure, holding all other parameters unchanged. That is, generate $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ from the full conditional distribution.
3. Move from a selection of random effects to another selection of random effects by a standard MCMC step. Just like for the fixed effect, the index for the random effects is determined by $a_l = 0$ or not, $l = 1, \ldots, q$.
4. Update all random effect parameters, holding all other parameters unchanged.

The proof of convergence properties as shown by Green and O'Hagan carries over to the algorithm above. All the relevant posterior computations are given in the Appendix. Note that we update each γ_k individually, $k = 1, \ldots, p$. Here we actually integrate out $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ in (10). A similar approach integrating out both $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ and σ^2 was used in Smith and Kohn (1996) to accelerate the convergence of the MCMC chain. We investigated both approaches, and the results were similar. Each a_l is also updated individually, $l = 1, \ldots, q$. The zero-inflated truncated normal prior for a_l yields a conjugate posterior.

3 Simulation Experiments

We simulated data under model (1) for various numbers of clusters and cluster sizes (n, n_i). Here we show the results with relatively small n and n_i, to illustrate the type of sample sizes required for the SSVS to properly select the fixed and the random effects. We set $\lambda_0(t) = 1$. Censoring was generated as Uniform(0, τ), where τ was chosen so that about 20% of the observations were censored in each case. We had $p = 4$ potential covariates, and $\mathbf{x}_{ij} = (x_{i1}, x_{i2}, x_{i3}, x_{i4})'$ where each component of $\mathbf{x}$ was generated independently from Uniform (−2, 2). For the random effects, we had $q = 3$, and $\mathbf{z}_{ij} = (1, x_{i1}, x_{i2})'$. The true value of the parameters were $\boldsymbol{\beta} = (0.8, 0.4, 0.4, 0)'$, and $\boldsymbol{\Sigma} = \text{diag}(0.4, 0.2, 0)$. In the tables we used subscript 0, 1 and 2 to indicate the random effects for the baseline hazard, x_1 and x_2, respectively. We also gave the empirical variances of the simulated random effects in parenthesis in addition to the true values of $\boldsymbol{\Sigma}$; the accuracy of the estimated variances can be better reflected when compared to these empirical variances than to the true values. We first used 'non-informative' prior for selecting any of the fixed or random effects, that is, $\pi_k = p_{l0} = 0.5$, $k = 1, \ldots, p$, $l = 1, \ldots, q$. The sample sizes were 20, 30 and 50 clusters, with cluster sizes 10 and 20, respectively. The MCMC chain consisted of 10,000 iterations, with the first 10% for burn-in so that $S = 9,000$ in this case.

The SSVS gives the marginal posterior probability for selecting each of the fixed and random effects as described in the previous section. It also gives the posterior probability of each potential model in a similar way. In Tables 1 and 2 we present the results for the three models with the highest posterior probabilities, namely B1, B2 and B3, as well as the averaged estimates from all $S = 9,000$ iterations and the corresponding 95% credible intervals. The latter have the interpretation of model averaged inference. Alternatively, a

user may choose a model based on its posterior probability, say, and carry out conditional inference; this will be illustrated in our data example of the next section.

For the smallest sample size of 20 clusters with 10 observations each, the top one-third of Table 1 shows that all of the top three models missed the random effect on x_1 which had a variance of 0.2. The first two fixed effects had marginal posterior probability one, while the third fixed effect had a posterior probability of 0.764. The random effect on the baseline hazard with a variance of 0.4 had a posterior probability of 0.635. The 95% credible intervals contained the true values of all parameters except β_3 and Σ_{11}.

When the number of clusters increased to 30 in the middle of Table 1, the true model had a posterior probability of 0.678. All the 'true' fixed effects had posterior probability one, and two random effects had posterior probability 0.77 and 0.934, respectively. The 4th fixed effects had a posterior probability of 0.212, which was slightly high. The 95% credible intervals contained the true values of all parameters except Σ_{00}, which was under-estimated.

The results for 50 × 10 are in the bottom one-third of Table 1, where the 4th fixed effect had a high posterior probability of 0.418. But the true random effects had much higher posterior probabilities (1 and 0.91, respectively) than the previous two scenarios, and the 95% credible intervals contained the true values of all parameters except Σ_{00}, which was over-estimated in this case.

When the cluster size increased to 20 observations per cluster, even with only 20 clusters in Table 2, the results were quite good: the true model had a posterior probability of 0.891, all the true fixed and random effects had posterior probabilities of one or very close to one (0.968 for Σ_{11}), the null fixed (β_4) and random (Σ_{22}) effects had very low posterior probabilities, and the 95% credible intervals contained the true values of all parameters. With larger numbers of clusters as in the middle and bottom of Table 2, the results were even better, with generally tighter credible intervals, and the true model had a posterior probability of 0.94 when there were 50 clusters of 20 observations each.

The above used 'non-informative' prior for selecting any of the fixed or random effects, that is, $\pi_k = p_{l0} = 0.5$, $k = 1, ..., 4$, $l = 0, 1, 2$. In Table 3 we consider the sensitivity of the prior probabilities p_{l0}'s for selecting random effects. As we understand in general, when the sample sizes are small, the prior should have more influence on the posterior distribution, as compared to larger sample sizes. This is indeed the case in Table 3. Note that p_{l0} is the probability of mass zero in the prior (11) for the random effects. When $n_i = 10$, $p_{l0} = 0.2$ ($l = 0, 1, 2$) gave rise to high posterior probabilities (1 and 0.887) for the two random effects corresponding to Σ_{00} and Σ_{11}, while $p_{l0} = 0.8$ ($l = 0, 1, 2$) gave only high posterior probability for one random effect. On the other hand, when $n_i = 20$, both $p_{l0} = 0.2$ and 0.8 ($l = 0, 1, 2$) gave high posterior probabilities for the two random effects. In fact, when $p_{l0} = 0.8$ only two models (B1 and B2) had positive posterior probabilities, and both contain the two true random effects.

4 An Example

We apply our proposed model to a multi-center advanced stage non-small cell lung cancer clinical trial data which was analyzed in Gray (1994) and Vaida and Xu (2000). The study was conducted by the Eastern Coorperative Oncology Group. There were two randomized treatment arms: a standard chemotherapy (CAV) and an alternating regimens (CAV-HEM) where cycles of CAV were alternated with HEM. The outcome of interest was overall survival, and the longest follow-up time was about 8.4 years. Five binary covariates were found to be significantly associated with survival in the previous published analyses: treatment assignment, presence or absence of bone metastases, presence or absence of liver metastases, performance status at study entry (ambulatory or not), and whether there was weight loss prior to study entry. Gray (1995) found significant institution-to-institution variation in the treatment effects using a score test under the frailty model. Vaida and Xu (2000) fitted model (1) to the data with potential random effects for all five covariates, and found that those for bone metastases were even stronger than the random effects for treatment, while the variances of the random effects of the rest three covariates converged towards zero. Dunson and Chen (2004) considered selection of frailty terms using a Bayesian approach by putting a mixture prior on the frailty variances with point mass at zero and inverse Gamma, and concluded that after accounting for the random bone metastases effects, there was no direct evidence of institutional variation in treatment effects. This then led to a correspondence by Gray (2006) pointing out the statistical significance of the random treatment effects by a score test even after accounting for the random bone metastases effects, together with a reply by Dunson and Chen who did a separate analysis to support their original conclusion published in 2004.

Here we take another look at the data using the SSVS approach. We consider the 22 institutions with more than 7 enrolled subjects each; this gives a total of 546 patients, and the actual numbers of patients per institution are between 11 and 56. In addition to the fixed effects for the five covariates described above, we also consider six potential random effects, on the baseline hazard as well as for the five covariates. We first set the prior probabilities for the fixed and the random effects to be $\pi_k = 0.5$, $k = 1, ..., 5$, and $p_{l0} = 0.5$, $l = 0, 1, ..., 5$. As mentioned before, we consider these as 'non-informative' priors for the probabilitilies of inclusion of the fixed and random effects, respectively. We ran three Gibbs samplers for 10,000 iterations, after a burn-in period of 5000 iterations from dispersed initial values. The Gelman-Rubin statistic (Gelman and Rubin, 1992) was computed to check the convergence and the chains were shown to converge. The results are given in Table 4. The three models with the highest posterior probabilities contain only the random bone metastases effect, and the random treatment effect has a very low inclusion probability of 0.002. The fixed effects of treatment and bone metastases have slightly low inclusion probabilities of 0.81 and 0.84, respectively. In the table we also provide 95% credible intervals conditional on B1, if that is the selected model, as well as averaged over all the models in the 'Estimate (95% CI)' column, like we did in the simulations of the last section.

To better understand the behavior of SSVS in this case, we carry out further simulations in Table 5 mimicking the lung cancer data. The covariates as well as the sample sizes including the number of clusters and the numbers of observations in each cluster for both tables are the

same as in the lung cancer data, and the baseline hazard function is estimated nonparametrically from the lung cancer data with R package 'phmm'. Recall that in the simulations of Section 3 all the covariates were continuously distributed as Uniform(−2, 2), with a variance of 4/3. For binary (0, 1) covariates, however, the variance is only 1/4. We can only compare the strength of any effect when the corresponding covariates are on the same scale, since we can otherwise always multiple the effect by a non-zero constant and divide the covariate by the same constant and the model is unchanged. In the top half ('weak effects') of Table 5 the strength of the random effects as reflected in their variances Σ_{11} and Σ_{22} are comparable to those estimated from the lung cancer data, while in the bottom half ('strong effects') they are increased to be equivalent to those for the Uniform(−2,2) covariates as in Section 3 ($0.2 \times 16/3 = 16/15$, $0.4 \times 16/3 = 32/15$). It is clear from the table that when both random effects are strong, both of them have a posterior probability of one; but with the level of strength as in the lung cancer data, only the stronger of the two random effects has a high probability of inclusion.

The above investigation might provide some explanation for the discrepancy between the frequentist score test as mentioned before and the Bayesian variable selection for the lung cancer data. While the score test detects significant institutional variation in treatment effects after having accounted for the random bone metastases effects, the random treatment effects are relatively weak such that in simulation studies the Bayesian variable selection chooses not to model it. From the point of view of model selection, it then depends on the criterion that is important to the question of concern according to which one chooses to model the random treatment effect or not. We also note that Cai (2010) analyzed this data set using Dirichlet process prior for the frailty distribution, and reached a conclusion that was somewhat a compromise between Dunson and Chen (2004) and Gray (1995).

Finally we carry out sensitivity analysis of the lung cancer data in Table 6 with respect to the prior probabilities of both the fixed (π_k's) and the random (p_{l0}'s) effects. Since there has been consensus in the literature about the clinical importance of all five covariates, we set the prior probabilities $\pi_k = 1$ ($k = 1, ..., 5$) in the top third of Table 6. This also makes it more comparable with Dunson and Chen (2004), where the inclusion of the fixed effects was not questioned. The model with the highest posterior probability contains only the random bone metastases effect, while the random treatment effect has an inclusion probability of 0.01. We then set the prior probabilities p_{l0} for the (exclusion of) random effects to be 0.2 and 0.8 ($l = 0, 1, ..., 5$), and the results are in the rest of Table 6. Again note that p_{l0} is the probability of mass zero in the prior (11) for the random effects, so $p_{l0} = 0.8$ makes it unlikely for a random effect to be included in the model, as is the case in the bottom of Table 6. In this case, it might be the relatively weak random effect of bone metastases (see above), in addition to some of the small cluster sizes in the data, that causes $p_{l0} = 0.8$ to lead to basically no selected random effects. On the other hand, even with a low $p_{l0} = 0.2$, we still only have the random bone metastases effect with high posterior probability of 0.85, while the other random effects have very low posterior probabilities of inclusion.

5 Discussion

In this paper we have developed the Bayesian SSVS approach for selection of fixed as well as random effects under the PHMM. For computational purpose, we have added the ε_{ij}'s to the linear predictor in the PHMM, which also expands the model to allow for individual heterogeneity. Our simulation results show that this approach works well even when the data are generated as i.i.d., i.e. with Var(ε_{ij}) = 0. For the prior distribution of σ^2 = Var(ε_{ij}), we have also considered truncated inverse-Gamma, the simulation results (data not shown) depended on the range of truncation and were generally not better than the uniform prior described in details in this paper.

We have used the Gamma process prior, which allows marginalization over the prior distribution of the baseline hazard function, and greatly accelerates the MCMC convergence. There are more general and flexible stochastic processes, such as the generalized Gamma process (Brix, 1999). The use of such a more general prior will substantially increase the computational burden. Alternatively, one might also consider putting prior on the increments of the cumulative baseline hazard function, as was done in Gray (1994) and Dunson and Chen (2004). As suggested by a reviewer, and also confirmed in our empirical experiences, while the Gamma process prior is relatively restrictive, it appears to provide sufficient performance for the purpose of variable selection. On the other hand, if more accurate parameter estimation or prediction is of interest, one might wish to investigate the more general prior processes and develop the corresponding computation techniques. This certainly makes an interesting future research topic.

We have carried out sensitivity analysis in both simulation and the lung cancer data analysis, with respect to the prior probabilities π_k and p_{l0} for selection of the fixed and random effects. We consider π_k = 0.5 and p_{l0} = 0.5 to be the 'non-informative' priors, to be used when we have no reference or prior knowledge of whether a fixed or random effect should be included in a model. On the other hand, we can set either of them to be one to force a fixed or random effect into a model. From our sensitivity analysis, it is seen that the smaller the sample size, the more influence the prior has on the posterior probabilities of selection in general. In addition, the strength of a random effect (see also below) also plays a role in its posterior probability of inclusion in the model.

A reviewer asked about the definition of weak versus strong random effects, which have led to different conclusions of variable selection in Section 4. We think that this is a relative concept, and perhaps largely depends on the underlying data structure which is often unknown. We have illustrated through the lung cancer data analysis how simulation can be used to help understand the conclusions of model selection methods, given the possibly complex real data structure at hand.

The SSVS provides the posterior probabilities of any fixed and random effects under the PHMM, as well as the posterior probability of any candidate model that consists of a combination of the fixed and random effects. A user may choose his or her own decision rule, together with conditional or model averaged inference, depending on the needs and interpretation of the data analysis. George and Foster (2000) discussed the connection

between Bayes variable selection and other commonly used types of information criteria such as AIC, BIC, etc; the explicit connection under the PHMM is an open problem to be explored.

Our simulation experiments were carried out with moderate sample sizes. A notable phenomenon has been that the cluster sizes appear to have more impact on the performance of SSVS than the number of clusters: larger cluster sizes have substantially improved the variable selection. We note that the approach has not worked well for clusters as small as five observations each. Therefore, this method would not apply to certain data types with small clusters, such as in a twin study. Model selection under the PHMM for those cases still remains an area to be further studied.

Acknowledgments

The research of Kyeong Eun Lee was supported by Basic Science Research Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Education, Science and Technology (2010-0023305). The research of Ronghui Xu was partially supported by the United States National Institutes of Health NCRR Clinical and Translational Science Award 1UL1RR031980.

Appendix

The survival function conditional on the random effects is

$$P\left(T_{ij} \geq t | W_{ij}, \Lambda_0\right) = exp\left\{-\Lambda_0\left(t\right) exp\left(W_{ij}\right)\right\}, \quad i=1, \cdots, n; j=1, \cdots, n_i$$

where Λ_0 is the unknown baseline cumulative hazard function. The joint survival function conditional on the random effects is then

$$P\left(T_{11} \geq t_{11}, \ldots, T_{nn_n} \geq t_{nn_n} | \mathbf{W}, \Lambda\right) = exp\left\{-\sum_{i=1}^{n}\sum_{j=1}^{n_i}\Lambda_0\left(t_{ij}\right) exp\left(W_{ij}\right)\right\}.$$

For notational convenience, $\boldsymbol{\theta}_{-a}$ denotes all $\boldsymbol{\theta}$ except a.

- Full conditional distribution of **W**:

$$
\begin{aligned}
p(\mathbf{W}|\boldsymbol{Y},\boldsymbol{\theta}) &\propto p(\boldsymbol{Y}|\mathbf{W})\, p(\mathbf{W}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}) \\
&= exp\left\{-\sum_{i=1}^{n}\sum_{j=1}^{n_1} aB_{ij}\lambda^*(t_{ij})\right\}\prod_{i=1}^{n}\prod_{j=1}^{n_i}\{a\lambda^*(t_{ij})\,B_{ij}\}^{\delta_{ij}} \\
&\times\prod_{i=1}^{n}\left[|\mathbf{D}|^{-1/2}exp\left(-\frac{1}{2}\boldsymbol{\xi}_i'\mathbf{D}^{-1}\boldsymbol{\xi}_i\right)\prod_{j=1}^{n_i}\sigma^{-1}exp\left\{-\frac{1}{2\sigma^2}\left(W_{ij}-\mathbf{X}_{ij}'\boldsymbol{\beta}-\mathbf{z}_{\mathbf{ij}}'\mathbf{A}\boldsymbol{\Omega}\boldsymbol{\xi}_{\mathbf{i}}\right)^2\right\}\right] \\
&\times p\left(\sigma^2\right)p(\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{g})\,p(\boldsymbol{\alpha},\boldsymbol{\omega})\,p(\mathbf{D}) \propto \prod_{i=1}^{n}\prod_{j=1}^{n_i}\{a\lambda^*(t_{ij})\,B_{ij}\}^{\delta_{ij}}\,exp\left\{-\sum_{i=1}^{n}\sum_{j=1}^{n_i}aB_{ij}\Lambda^*(t_{ij})\right\} \\
&\times exp\left[-\frac{1}{2\sigma^2}\{\mathbf{W}-\mathbf{X}\boldsymbol{\beta}-\boldsymbol{Z}(\mathbf{I_n}\otimes\mathbf{A}\boldsymbol{\Omega})\boldsymbol{\xi}\}'\{\mathbf{W}-\mathbf{X}\boldsymbol{\beta}-\boldsymbol{Z}(\mathbf{I_n}\otimes\mathbf{A}\boldsymbol{\Omega})\boldsymbol{\xi}\}\right] \\
&\times p\left(\sigma^2\right)p(\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{g})\,p(\boldsymbol{\alpha},\boldsymbol{\omega})\,p(\mathbf{D})
\end{aligned}
\tag{17}
$$

- Full conditional distribution of $\boldsymbol{\gamma}$:

Let $p_\gamma=\sum_{k=1}^{p}1(\gamma_k=1)$ and

$$
S(\gamma)=\left[(\mathbf{W}-\mathbf{W}_0)'\left\{\mathbf{I}-\frac{1}{1+g}\mathbf{X}_\gamma\left(\mathbf{X}_\gamma'\mathbf{X}_\gamma\right)^{-1}\mathbf{X}_\gamma'\right\}(\mathbf{W}-\mathbf{W}_0)\right]^{-\frac{N}{2}}
$$

where $\mathbf{W}_0=\boldsymbol{Z}(\mathbf{I}_n\otimes\mathbf{A}\boldsymbol{\Gamma})\boldsymbol{\xi}$.

Then, integrate out $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ and σ^2 as follows:

$$
\begin{aligned}
p\left(\mathbf{W}|\boldsymbol{\theta}_{-(\boldsymbol{\beta}_\gamma,\sigma^2)}\right) &\propto \int_{\sigma^2}\left\{\int_\gamma p(\mathbf{W}|\boldsymbol{\theta})\,p\left(\boldsymbol{\beta}_\gamma|\sigma^2,g\right)d\boldsymbol{\beta}_\gamma\right\}p\left(\sigma^2\right)d\sigma^2 \\
&\propto \left(1+\tfrac{1}{g}\right)^{-\frac{p_\gamma}{2}}S(\gamma)
\end{aligned}
$$

Then

$$
p\left(\gamma|\mathbf{W},\boldsymbol{\theta}_{-(\gamma_k,\boldsymbol{\beta}_\gamma,\sigma^2)},\boldsymbol{Y}\right)\propto\left(1+\tfrac{1}{g}\right)^{-\frac{p_\gamma}{2}}S(\gamma)\prod_{k=1}^{p}\pi_k^{\gamma_k}(1-\pi_k)^{1-\gamma_k},
$$

Let $\gamma_k^1=(\gamma_1,\ldots,\gamma_k=1,\ldots,\gamma_p)$, and $\gamma_k^0=(\gamma_1,\ldots,\gamma_k=0,\ldots,\gamma_p)$, k = 1, ..., p. Since

$p\left(\gamma_k=1|\mathbf{W},\boldsymbol{\theta}_{-(\gamma_k,\boldsymbol{\beta}_\gamma,\sigma^2)},\boldsymbol{Y}\right)\propto\left(1+\tfrac{1}{g}\right)^{-\frac{p_{\gamma_k^1}}{2}}S\left(\gamma_k^1\right)\pi_k$ and

$p\left(\gamma_k=0|\mathbf{W},\boldsymbol{\theta}_{-(\gamma_k,\boldsymbol{\beta}_\gamma,\sigma^2)},\boldsymbol{Y}\right)\propto\left(1+\tfrac{1}{g}\right)^{-\frac{p_{\gamma_k^0}}{2}}S\left(\gamma_k^0\right)(1-\pi_k)$, we have

$$
p\left(\gamma_k=1|\mathbf{W},\boldsymbol{\theta}_{-(\gamma_k,\boldsymbol{\beta}_\gamma,\sigma^2)},\boldsymbol{Y}\right)=\frac{1}{1+\left(1+\frac{1}{g}\right)^{\frac{1}{2}}\cdot\frac{S\left(\gamma_k^0\right)(1-\pi_k)}{S\left(\gamma_k^1\right)\pi_k}}.
$$

- Full conditional distribution of $\boldsymbol{\gamma}$ after integrating out $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$:

Let $p_\gamma = \sum_{k=1}^{p} 1\left(\gamma_k = 1\right)$ and

$$S\left(\gamma\right) = exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{W}-\mathbf{W}_0\right)'\left(\mathbf{I}-\frac{1}{1+g}\mathbf{X}_\gamma\left(\mathbf{X}_\gamma' X_\gamma\right)^{-1}\mathbf{X}_\gamma'\right)\left(\mathbf{W}-\mathbf{W}_0\right)\right\}$$

where $\mathbf{W}_0 = \boldsymbol{Z}\left(\mathbf{I}_n \otimes \mathbf{A}\boldsymbol{\Gamma}\right)\boldsymbol{\xi}$. Then, integrate out $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ as follows:

$$\begin{aligned} p\left(\mathbf{W}|\boldsymbol{\theta}_{\gamma,\sigma^2}\right) &\propto \int_\gamma p\left(\mathbf{W}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\beta}_\gamma|\sigma^2, g\right) d\boldsymbol{\beta}_\gamma \\ &\propto \left(1+\tfrac{1}{g}\right)^{-\frac{p_\gamma}{2}} S\left(\gamma\right) \end{aligned}$$

Then

$$p\left(\gamma|\mathbf{W}, \boldsymbol{\theta}_{-(\gamma,\boldsymbol{\beta}_\gamma)}, \boldsymbol{Y}\right) \propto \left(1+\tfrac{1}{g}\right)^{-\frac{p_\gamma}{2}} S\left(\gamma\right) \prod_{k=1} \pi_k^{\gamma_k}(1-\pi_k)^{1-\gamma_k}.$$

Let $\gamma_k^1 = \left(\gamma_1, \ldots, \gamma_k = 1, \ldots, \gamma_p\right)$ and $\gamma_k^0 = \left(\gamma_1, \ldots, \gamma_k = 0, \ldots, \gamma_p\right)$, k = 1, ..., p. Since

$p\left(\gamma_k = 1|\mathbf{W}, \boldsymbol{\theta}_{-(\gamma_k,\boldsymbol{\beta}_\gamma)}, \boldsymbol{Y}\right) \propto \left(1+\tfrac{1}{g}\right)^{-\frac{p_{\gamma_k^1}}{2}} S\left(\gamma_k^1\right)\pi_k$ and

$p\left(\gamma_k = 0|\mathbf{W}, \boldsymbol{\theta}_{-(\gamma_k,\boldsymbol{\beta}_\gamma)}, \boldsymbol{Y}\right) \propto \left(1+\tfrac{1}{g}\right)^{-\frac{p_{\gamma_k^0}}{2}} S\left(\gamma_k^0\right)\left(1-\pi_k\right)$, we have

$$P\left(\gamma_k = 1|\mathbf{W}, \boldsymbol{\theta}_{-(\gamma_k,\boldsymbol{\beta}_\gamma)}, \boldsymbol{Y}\right) \frac{1}{1+\left(1+\frac{1}{g}\right)^{\frac{1}{2}} \cdot \frac{S\left(\gamma_k^0\right)\left(1-\pi_k\right)}{S\left(\gamma_k^1\right)\pi_k}}.$$

- Full conditional distribution of $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$:

$$\begin{aligned} &p\left(\boldsymbol{\beta}_\gamma|\boldsymbol{\theta}_{-\boldsymbol{\beta}}, \mathbf{W}, \boldsymbol{Y}\right) \\ &= p\left(\boldsymbol{\beta}_\gamma|\boldsymbol{\theta}_{-\boldsymbol{\beta}}, \mathbf{W}\right) \propto p\left(\mathbf{W}|\boldsymbol{\theta}\right) \\ &\times p\left(\boldsymbol{\beta}_\gamma|\boldsymbol{\theta}_{-\boldsymbol{\beta}}\right) \propto exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{W}-\mathbf{X}_\gamma\boldsymbol{\beta}_\gamma-\boldsymbol{Z}\left(\mathbf{I}_n\otimes\mathbf{A}\boldsymbol{\Omega}\right)\boldsymbol{\xi}\right)'\left(\mathbf{W}-\mathbf{X}_\gamma\boldsymbol{\beta}_\gamma-\boldsymbol{Z}\left(\mathbf{I}_n\otimes\mathbf{A}\boldsymbol{\Omega}\right)\boldsymbol{\xi}\right)\right\} \\ &\times\left|\mathbf{X}_\gamma'\mathbf{X}_\gamma\right|^{1/2} exp\left(-\frac{g}{2\sigma^2}\boldsymbol{\beta}_\gamma'\mathbf{X}_\gamma'\mathbf{X}_\gamma\boldsymbol{\beta}_\gamma\right) \propto exp\left\{-\frac{1+g}{2\sigma^2}\boldsymbol{\beta}_\gamma'\mathbf{X}_\gamma'\boldsymbol{\beta}_\gamma+\frac{1}{2\sigma^2}\boldsymbol{\beta}_\gamma'\mathbf{X}_\gamma^1\left(\mathbf{W}-\boldsymbol{Z}\left(\mathbf{I}_n\otimes\mathbf{A}\boldsymbol{\Omega}\right)\boldsymbol{\xi}\right)\right\}, \end{aligned}$$

therefore

$$\boldsymbol{\beta}_\gamma|\boldsymbol{\theta}_{-\boldsymbol{\beta}}, \mathbf{W}, \boldsymbol{Y} \sim N\left(\hat{\boldsymbol{\beta}}_\gamma, \hat{\mathbf{V}}_\gamma\right),$$

where $\hat{\boldsymbol{\beta}}_\gamma = \frac{1}{\sigma^2}\hat{\mathbf{V}}_\gamma\mathbf{X}_\gamma'\left(\mathbf{W}-\boldsymbol{Z}\right)\left(\mathbf{I}_n\otimes\boldsymbol{Z}\boldsymbol{\Omega}\right)\boldsymbol{\xi}$ and $\hat{\mathbf{V}}_\gamma = \frac{\sigma^2}{1+g}\left(\mathbf{X}_\gamma'\mathbf{X}_\gamma\right)^{-1}$.

- Full conditional distribution of g:

$$p(g|\cdot) \propto p(g)\, p\left(\boldsymbol{\beta}_\gamma|g\right) \propto \frac{(N/2)^{1/2}}{\Gamma(1/2)} g^{1/2-1} e^{-Ng/2} \times \left|\frac{\sigma^2}{g}\left(\mathbf{X}_\gamma'\mathbf{X}_\gamma\right)^{-1}\right|^{-1/2} exp\left\{-\frac{g}{2\sigma^2}\boldsymbol{\beta}_\gamma'\left(\mathbf{X}_\gamma'\mathbf{X}_\gamma\right)\boldsymbol{\beta}_\gamma\right\} \propto g^{\frac{p_\gamma+1}{2}-1} exp\left\{-\frac{\boldsymbol{\beta}_\gamma'\left(\mathbf{X}_\gamma'\mathbf{X}_\gamma\right)\boldsymbol{\beta}_\gamma/\sigma^2+N}{2} g\right\},$$

  where $p_\gamma = \sum_{l=1}^{p} 1(\gamma_l = 1)$. Therefore,
  $g|\cdot \sim \text{Gamma}\left((p_\gamma+1)/2, \left(\boldsymbol{\beta}_\gamma', \mathbf{X}_\gamma'\mathbf{X}_\gamma\boldsymbol{\beta}_\gamma/\sigma^2+N\right)/2\right)$.

- Full conditional distribution of σ^2:

$$p\left(\sigma^2|\cdot\right) \propto p(\mathbf{W}|\theta)\, p\left(\boldsymbol{\beta}_J|\cdot\right) p\left(\sigma^2\right) \propto \left(\sigma^2\right)^{-N/2} exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\psi}'\boldsymbol{\psi}\right\} \times \left(\sigma^2\right)^{-p_\gamma/2} exp\left\{-\frac{g}{2\sigma^2}\boldsymbol{\beta}_J'\left(\mathbf{X}_\gamma'\mathbf{X}_\gamma\right)\boldsymbol{\beta}_J\right\} \cdot \frac{1}{\sigma^2}.$$

  Therefore

$$\sigma^2|\cdot \sim IG\left(\frac{N+p_J}{2}, \frac{\boldsymbol{\psi}'\boldsymbol{\psi}+\mathbf{g}\boldsymbol{\beta}_\mathbf{J}'\mathbf{X}_\mathbf{J}'\mathbf{X}_\mathbf{J}\boldsymbol{\beta}_\mathbf{J}}{2}\right)$$

  where $\boldsymbol{\psi} = \mathbf{W} - \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma - \boldsymbol{Z}\left(\mathbf{I_n} \otimes \mathbf{A\Omega}\right)\boldsymbol{\xi}$.

- The full conditional distribution $p(\boldsymbol{\omega}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\varepsilon}, \mathbf{W}, \mathbf{X}\ \mathbf{Z})$ is $N\left(\hat{\boldsymbol{\omega}}, \hat{\mathbf{V}}_{\boldsymbol{\omega}}\right) \cdot 1(\boldsymbol{\omega} \in \mathbf{R})$, where

$$\hat{\mathbf{V}}_{\boldsymbol{\omega}} = \left(\frac{\mathbf{U}'\mathbf{U}}{\sigma^2} + \mathbf{V}_{\boldsymbol{\omega}}^{-1}\right)^{-1}, \quad (18)$$

$$\hat{\boldsymbol{\omega}} = \hat{\mathbf{V}}_{\boldsymbol{\omega}}\left\{\mathbf{V}_{\boldsymbol{\omega}}^{-1}\boldsymbol{\omega}_0 + \frac{\mathbf{U}'\left(\mathbf{W} - \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma\right)}{\sigma^2}\right\}, \quad (19)$$

  $\mathbf{U} = (\mathbf{u}_{11}, \mathbf{u}_{12}, \ldots, \mathbf{u}_{nn_n})'$, and the $q(q-1)/2$ vector $\mathbf{u}_{ij}$ is defined as $(\xi_{il}\alpha_m Z_{ijm} : l = 1, \ldots, q, m = l+1, \ldots, q)'$, so that the random effects term $\mathbf{z}_{ij}'\mathbf{A\Omega\xi_i}$ can be written as $\mathbf{u}_{ij}'\boldsymbol{\omega}$.

- The latent variables $\boldsymbol{\xi_i}$ have conditional distribution $p(\boldsymbol{\xi}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{W}, \mathbf{X}, \mathbf{Z})$ given by $N\left(\hat{\boldsymbol{\xi}}_i, \hat{\mathbf{V}}_{\boldsymbol{\xi}_i}\right)$, where

$$\hat{\boldsymbol{\xi}}_i = \frac{1}{\sigma^2}\hat{\mathbf{V}}_{\boldsymbol{\xi}_i}\boldsymbol{\Omega}'\mathbf{A}'\boldsymbol{Z}_i'\left(\mathbf{W}_i - \mathbf{X}_{i\gamma}\boldsymbol{\beta}_\gamma\right), \quad (20)$$

$$\hat{\mathbf{V}}_{\boldsymbol{\xi}_i}=\left(\frac{\boldsymbol{\Omega}'\mathbf{A}'\boldsymbol{Z}_i'\boldsymbol{Z}_i\mathbf{A}\boldsymbol{\Omega}}{\sigma^2}+\mathbf{D}^{-1}\right)^{-1}. \quad (21)$$

- $p(\alpha_l|\boldsymbol{\alpha}_{-l},\boldsymbol{\beta},\gamma,\boldsymbol{\xi},\mathbf{W},\mathbf{X},\boldsymbol{Z})=ZI\text{-}\,\mathrm{N}^+\left(\hat{\alpha}_l,\hat{V}_{\alpha l},\hat{p}_l\right)$, l = 1, ..., q, where

$$\hat{\alpha}_l=\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}y_{ijl}Y_{ijl}}{\sigma^2}\right)\hat{V}_{\alpha l},\quad \hat{V}_{\alpha l}=\left(\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{y_{ijl}^2}{\sigma^2}+1\right)^{-1}$$
$$\hat{p}_l=\frac{p_{l0}}{p_{l0}+(1-p_{l0})\frac{\phi(0;0,1)}{\phi\left(0;\hat{\alpha}_l,\hat{V}_{\alpha l}\right)}\cdot\frac{1-\Phi\left(0;\hat{\alpha}_l,\hat{V}_{\alpha l}\right)}{1-\Phi(0;0,1)}},$$

  $Y_{ijl}=W_{ij}-\mathbf{x}_{ij\gamma}'\boldsymbol{\beta}_\gamma-\sum_{k\neq l}y_{ijk}\alpha_k$, $\phi(0;m,v)$ denotes the normal density with mean m and variance v evaluated at 0, and $\Phi(0; m, v)$ denotes the normal cumulative distribution function with mean m and variance v evaluated at 0. The q vector

$$\mathbf{y}_{ij}=\left(Z_{ijl}\left(\xi_{il}+\sum_{m=1}^{l-1}\xi_{im}\gamma_{lm}\right):l=1,\ldots,q\right)^T$$

  is defined so that the random effects term $\mathbf{z}_{ij}'\mathbf{A}\boldsymbol{\Omega}\boldsymbol{\xi}_\mathbf{i}$ can be written as $\mathbf{y}_{ij}'\boldsymbol{\alpha}$.

- The full conditional distribution of d_l, l = 1, ..., q is $IG\left((n+1)/2,\left(N+\sum_{i=1}^{n}\xi_{il}^2\right)/2\right)$.

References

Beattie SD, Fong DKH, Lin DKJ. A two-stage bayesian model selection strategy for supersaturated designs. Technometrics. 2002; 44:55–63.

Brix A. Generalized gamma measures and shot-noise cox processes. Adv. Appl. Probab. 1999; 31:929–953.

Cai B. Bayesian semiparametric frailty selection in multivariate event time data. Biometrical Journal. 2010; 52:171–185. [PubMed: 20358551]

Chen Z, Dunson DB. Random effects selection in linear mixed models. Biometrics. 2003; 59:762–769. [PubMed: 14969453]

Clayton DG. A Monte Carlo method for Bayesian inference in frailty models. Biometrics. 1991; 47:467–485. [PubMed: 1912256]

Commenges D, Andersen P. Score test of homogeneity for survival data. Lifetime Data Analysis. 1995; 1:145–156. [PubMed: 9385097]

Cortiñas-Abrahantes J, Legrand C, Burzykowski T, Janssen P, Ducrocq V, Duchateau L. Comparison of different estimation procedures for proportional hazards model with random effects. Computational Statistics and Data Analysis. 2007; 51:3913–3930.

Cox D. Partial likelihood. Biometrika. 1975; 62:269–276.

Donohue MC, Overholser R, Xu R, Vaida F. Conditional Akaike information under generalized linear and proportional hazards mixed models. Biometrika. 2011; 98:685–700. [PubMed: 22822261]

Dunson DB, Chen Z. Selecting factors predictive of heterogeneity in multivariate event time data. Biometrics. 2004; 60:352–358. [PubMed: 15180660]

Fan J, Li R. Variable selection for Cox's proportional hazards model and frailty model. Annals of Statistics. 2002; 30:74–99.

Gamst A, Donohue M, Xu R. Asymptotic properties and empirical evaluation of the npmle in the proportional hazards mixed-effects model. Statistica Sinica. 2009; 19:997–1011.

Gelman A. Prior distribution for variance parameters in hierarchical models. Bayesian Analysis. 2006; 1:515–533.

Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical Science. 1992; 7:457–511.

George EI, Foster DP. Calibration and empirical bayes variable selection. Biometrika. 2000; 87:731–747.

George EI, McCulloch RE. Variable selection via gibbs sampling. Journal of the American Statistical Association. 1993; 88:881–889.

Gray R. A Bayesian analysis of institutional effects in a multicenter cancer clinical trial. Biometrics. 1994; 50:244–253. [PubMed: 8086608]

Gray R. Tests for variation over groups in survival data. Journal of the American Statistical Association. 1995; 90:198–203.

Gray R. Correspondence (Re: Dunson and Chen, 2004). Biometrics. 2006; 62:623–624. [PubMed: 16918929]

Green P, O'Hagan A. Model choice with mcmc on product spaces without using pseudo-priors. Nottingham University Statistics Research Report. 1998:98–01.

Ibrahim, JG.; Chen, M.; Sinha, D. Bayesian Survival Analysis. Springer-Verlag; New York: 2001.

Kalbfleisch JD. Nonparametric bayesian analysis of survival time data. Journal of the Royal Statistical Society, Series B. 1978; 40:214–221.

Kinney SK, Dunson DB. Fixed and random effects selection in linear and logistic models. Biometrics. 2007; 63:690–698. [PubMed: 17403104]

Kosorok, MR.; Lee, BL.; Fine, JP. Semiparametric inference for proportional hazards frailty regression models. Technical Report 156. Department of Biostatistics, University of Wisconsin; 2001.

Lee KE, Mallick BK. Bayesian methods for variable selection in survival models with application to dna microarray data. Sankhya: The Indian Journal of Statistics. 2004; 66:756–778.

Lee KE, Sha N, Dougherty ER, Vannucci M, Mallick BK. Gene selection: a bayesian variable selection approach. Bioinformatics. 2003; 19:90–97. [PubMed: 12499298]

Lin E, Huang LC. Identification of significant genes in genomics using bayesian variable selection methods. Advances and Applications in Bioinformatics and Chemistry. 2008; 1:13–18. [PubMed: 21918603]

Liu C,B,RD, Wu YN. Parameter expansion to accelerate EM: the PX-EM algorithm. Biometrika. 1998; 85:755–770.

O'Quigley J, Stare J. Proportional hazards models with frailties and random effects. Statistics in Medicine. 2002; 21:3219–33. [PubMed: 12375300]

Ripatti S, Palmgren J. Estimation of multivariate frailty models using penalized partial likelihood. Biometrics. 2000; 56:1016–1022. [PubMed: 11129456]

Smith M, Kohn R. Nonparametric regression using bayesian variable selection. Journal of Econometrics. 1996; 75:317–343.

Swartz MD, Yu RK, Shete S. Finding factors influencing risk: comparing variable selection methods applied to logistic regression models of cases and controls. Statistics in Medicine. 2008; 27:6158–6174. [PubMed: 18937224]

Vaida F, Xu R. Proportional hazards model with random effects. Statistics in Medicine. 2000; 19:3309–3324. [PubMed: 11122497]

Xu R, Vaida F, Harrington D. Using profile likelihood for semiparametric model selection with application to proportional hazards mixed models. Statistica Sinica. 2009; 19:819–842.

Zellner A. On assessing prior distributions and Bayesian regression analysis with g-prior distribution. Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti. 1986:233–243.

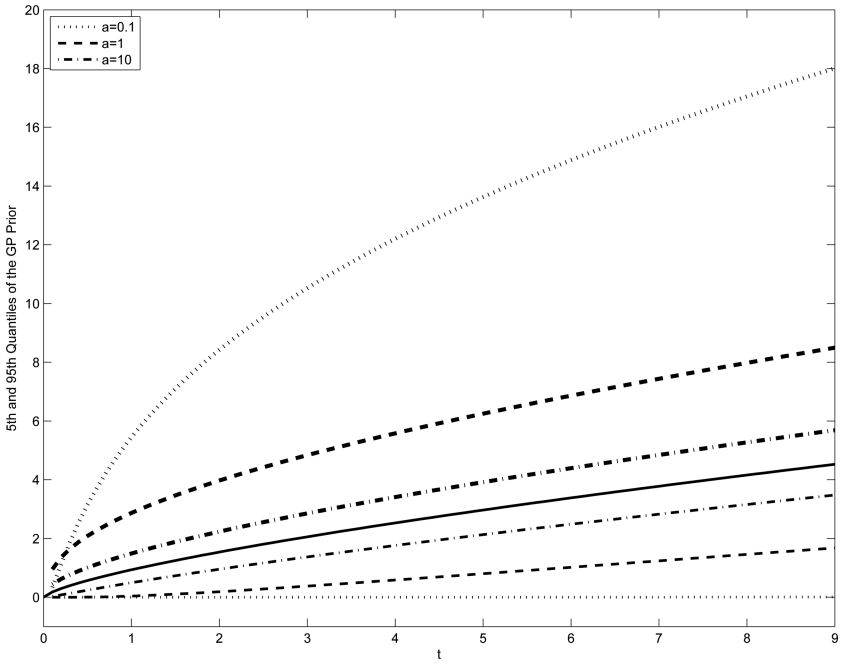

Figure 1.

The 5th and 95th quantiles of the Gamma process prior for the cumulative baseline hazard function, with $\hat{\eta}=0.94$ and $\hat{\kappa}=0.72$ estimated from the lung cancer data set. The solid line is the mean of the Gamma process.

Table 1

Simulation results with $n_i = 10$; B1, B2 and B3 are top three selected models.

	Parameter	True Value	B1	B2	B3	Estimate	95% CI	Pr(Inclusion)
$n = 20$	β_1	0.8	0.882	0.964	0.828	0.896	(0.738, 1.103)	1.000
	β_2	0.4	0.459	0.489	0.482	0.473	(0.313, 0.628)	1.000
	β_3	0.4	0.226	0.282		0.191	(0.000, 0.382)	0.764
	β_4	0				0.001	(−0.007, 0.028)	0.121
	Σ_{00}	0.4 (0.19)	0.313		0.327	0.203	(0.000, 0.605)	0.635
	Σ_{11}	0.2 (0.21)				0.001	(0.000, 0.000)	0.013
	Σ_{22}	0				0.000	(0.000, 0.000)	0.006
	Pr(Selection)		0.385	0.276	0.157			
$n = 30$	β_1	0.8	0.768	0.764	0.719	0.758	(0.603, 0.871)	1.000
	β_2	0.4	0.409	0.409	0.362	0.401	(0.305, 0.475)	1.000
	β_3	0.4	0.384	0.371	0.367	0.382	(0.295, 0.492)	1.000
	β_4	0		−0.075	−0.118	−0.021	(−0.142, 0.000)	0.212
	Σ_{00}	0.4 (0.17)	0.237	0.198		0.179	(0.000, 0.387)	0.770
	Σ_{11}	0.2 (0.17)	0.222	0.254	0.228	0.210	(0.000, 0.370)	0.934
	Σ_{22}	0				0.000	(0.000, 0.000)	0.001
	Pr(Selection)		0.678	0.091	0.091			
$n = 50$	β_1	0.8	0.850	0.867	0.796	0.851	(0.715, 0.999)	1.000
	β_2	0.4	0.464	0.471	0.447	0.465	(0.381, 0.541)	1.000
	β_3	0.4	0.500	0.496	0.438	0.494	(0.394, 0.597)	1.000
	β_4	0		0.095		0.039	(0.000, 0.144)	0.418
	Σ_{00}	0.4 (0.43)	0.843	0.798	0.666	0.808	(0.502, 1.233)	1.000
	Σ_{11}	0.2 (0.16)	0.306	0.306		0.279	(0.000, 0.465)	0.911
	Σ_{22}	0				0.000	(0.000, 0.000)	0.001
	Pr(Selection)		0.514	0.396	0.067			

Table 2

Simulation results with $n_i = 20$; B1, B2 and B3 are top three selected models.

	Parameter	True Value	B1	B2	B3	Estimate	95% CI	Pr(Inclusion)
$n = 20$	β_1	0.8	0.726	0.715	0.748	0.726	(0.564, 0.922)	1.000
	β_2	0.4	0.426	0.431	0.460	0.428	(0.351, 0.518)	1.000
	β_3	0.4	0.360	0.381	0.336	0.361	(0.283, 0.473)	1.000
	β_4	0		−0.012		−0.001	(0.000, 0.000)	0.078
	Σ_{00}	0.4 (0.41)	0.527	0.551	0.539	0.530	(0.259, 0.982)	1.000
	Σ_{11}	0.2 (0.22)	0.212	0.226		0.206	(0.071, 0.423)	0.968
	Σ_{22}	0				0.000	(0.000, 0.000)	0.001
	Pr(Selection)		0.891	0.077	0.030			
$n = 30$	β_1	0.8	0.775	0.771	0.776	0.775	(0.627, 0.940)	1.000
	β_2	0.4	0.403	0.397	0.405	0.402	(0.315, 0.483)	1.000
	β_3	0.4	0.469	0.476	0.501	0.470	(0.388, 0.559)	1.000
	β_4	0		−0.011		−0.001	(0.000, 0.000)	0.081
	Σ_{00}	0.4 (0.52)	0.718	0.721	0.707	0.718	(0.396, 1.174)	1.000
	Σ_{11}	0.2 (0.17)	0.242	0.256	0.274	0.244	(0.121, 0.425)	1.000
	Σ_{22}	0			0.002	0.000	(0.000, 0.000)	0.002
	Pr(Selection)		0.917	0.081	0.002			
$n = 50$	β_1	0.8	0.819	0.820		0.819	(0.685, 0.974)	1.000
	β_2	0.4	0.398	0.400		0.398	(0.330, 0.472)	1.000
	β_3	0.4	0.422	0.428		0.423	(0.354, 0.493)	1.000
	β_4	0		−0.004		−0.000	(0.000, 0.000)	0.060
	Σ_{00}	0.4 (0.29)	0.295	0.300		0.295	(0.176, 0.458)	1.000
	Σ_{11}	0.2 (0.28)	0.291	0.301		0.292	(0.189, 0.434)	1.000
	Σ_{22}	0					(0.000, 0.000)	0.000
	Pr(Selection)		0.940	0.060	0.000			

Table 3

Simulation results with $n = 20$; B1, B2 and B3 are top three selected models.

	Parameter	True Value	B1	B2	B3	Estimate	95% CI	Pr(Inclusion)
$n_i = 10$	β_1	0.8	0.446	0.426	0.642	0.430	(0.000, 0.739)	0.937
$p_{l0} = 0.2$	β_2	0.4	0.310	0.298	0.353	0.292	(0.018, 0.453)	0.952
	β_3	0.4	0.393	0.387	0.429	0.394	(0.258, 0.555)	1.000
	β_4	0		−0.025		−0.004	(−0.061, 0.015)	0.191
	Σ_{00}	0.4 (0.33)	0.637	0.636	0.908	0.702	(0.229, 1.554)	**1.000**
	Σ_{11}	0.2 (0.21)	0.269	0.275		0.267	(0.000, 0.683)	**0.887**
	Σ_{22}	0				0.011	(0.000, 0.090)	0.144
	Pr(Selection)		0.562	0.142	0.078			
$n_i = 10$	β_1	0.8	0.507	0.479	0.519	0.503	(0.338, 0.693)	1.000
$p_{l0} = 0.8$	β_2	0.4	0.287	0.269	0.251	0.281	(0.144, 0.409)	0.993
	β_3	0.4	0.358	0.352	0.366	0.357	(0.206, 0.522)	1.000
	β_4	0		−0.011		−0.002	(−0.044, 0.034)	0.160
	Σ_{00}	0.4 (0.33)	0.561	0.546	0.392	0.558	(0.194, 1.151)	**1.000**
	Σ_{11}	0.2 (0.21)			0.040	0.000	(0.000, 0.000)	**0.008**
	Σ_{22}	0				0.000	(0.000, 0.000)	0.002
	Pr(Selection)		0.825	0.159	0.007			
$n_i = 20$	β_1	0.8	0.738	0.715	0.696	0.733	(0.494, 0.976)	1.000
$p_{l0} = 0.2$	β_2	0.4	0.360	0.388	0.376	0.366	(0.268, 0.457)	1.000
	β_3	0.4	0.327	0.329	0.330	0.327	(0.243, 0.421)	1.000
	β_4	0		0.057		0.011	(0.000, 0.096)	0.201
	Σ_{00}	0.4 (0.32)	0.495	0.461	0.462	0.488	(0.205, 0.928)	**1.000**
	Σ_{11}	0.2 (0.28)	0.376	0.352	0.377	0.371	(0.199, 0.627)	**1.000**
	Σ_{22}	0			0.002	0.000	(0.000, 0.000)	0.008
	Pr(Selection)		0.793	0.200	0.006			
$n_i = 20$	β_1	0.8	0.696	0.791		0.705	(0.586, 0.837)	1.000
$p_{l0} = 0.8$	β_2	0.4	0.395	0.464		0.401	(0.335, 0.475)	1.000
	β_3	0.4	0.334	0.328		0.333	(0.297, 0.361)	1.000
	β_4	0		0.035		0.003	(0.000, 0.037)	0.090
	Σ_{00}	0.4 (0.32)	0.379	0.386		0.380	(0.215, 0.634)	**1.000**
	Σ_{11}	0.2 (0.28)	0.191	0.200		0.192	(0.117, 0.308)	**1.000**
	Σ_{22}	0				0.000	(0.000, 0.000)	0.000
	Pr(Selection)		0.910	0.090				

Table 4

Selection of fixed and random effects for the lung cancer data

Effect	Variable	B1 (95% CI)	B2	B3	Estimate (95% CI)	Pr(Inclusion)
Fixed	Treatment	−0.245 (−0.405, −0.109)	−0.229		−0.186 (−0.360, 0.000)	0.808
	Bone	0.248 (0.074, 0.435)		0.361	0.224 (0.000, 0.450)	0.835
	Liver	0.263 (0.132, 0.472)	0.276	0.408	0.317 (0.164, 0.521)	1.000
	P.S.	−0.430 (−0.591, −0.344)	−0.422	−0.574	−0.461 (−0.620, −0.347)	1.000
	W.L.	0.263 (0.173, 0.367)	0.287	0.309	0.273 (0.171, 0.386)	0.996
Random	Baseline				0.000 (0.000, 0.000)	0.001
	Treatment				0.000 (0.000, 0.000)	0.002
	Bone	0.158 (0.049, 0.305)	0.250	0.064	0.130 (0.000, 0.320)	0.822
	Liver				0.000 (0.000, 0.000)	0.000
	P.S.				0.000 (0.000, 0.000)	0.005
	W.L.				0.000 (0.000, 0.000)	0.002
Pr(Selection)		0.542	0.145	0.133		

Table 5

Simulated lung cancer data with weak and strong random effects

Effect	Parameter	True Value	B1	B2	B3	Estimate	95% CI	Pr(Inclusion)
	β_1	−0.5			−0.070	−0.010	(−0.082, 0.000)	0.168
	β_2	0.3	0.239		0.332	0.165	(0.000, 0.494)	0.609
	β_3	0.5	0.312	0.408	0.334	0.323	(0.000, 0.535)	0.907
	β_4	−0.8	−0.575	−0.546	−0.510	−0.554	(−0.788, −0.373)	1.000
	β_5	0.3	0.320	0.348	0.301	0.335	(0.201, 0.467)	1.000
	Σ_{00}	0				0.000	(0.000, 0.000)	0.022
Weak	Σ_{11}	0.1				0.000	(0.000, 0.000)	0.001
	Σ_{22}	0.2	0.320	0.367	0.322	0.343	(0.130, 0.658)	1.000
	Σ_{33}	0				0.000	(0.000, 0.000)	0.005
	Σ_{44}	0				0.000	(0.000, 0.000)	0.009
	Σ_{55}	0				0.001	(0.000, 0.000)	0.023
	Pr(Selection)		0.380	0.337	0.086			
	β_1	1	1.155	1.330	1.253	1.200	(0.921, 1.504)	1.000
	β_2	−2	−2.062	−2.100	−2.099	−2.074	(−2.297, −1.815)	1.000
	β_3	0			0.021	0.000	(0.000, 0.000)	0.038
	β_4	−1	−0.914	−1.028	−0.928	−0.943	(−1.117, −0.829)	1.000
	β_5	0		0.169		0.039	(0.000, 0.249)	0.235
	Σ_{00}	0				0.000	(0.000, 0.000)	0.001
Strong	Σ_{11}	16/15	0.619	0.900	0.677	0.688	(0.328, 1.346)	1.000
	Σ_{22}	32/15	1.548	1.341	1.550	1.492	(0.707, 2.566)	1.000
	Σ_{33}	0				0.003	(0.000, 0.000)	0.027
	Σ_{44}	0				0.000	(0.000, 0.000)	0.001
	Σ_{55}	0				0.000	(0.000, 0.000)	0.001
	Pr(Selection)		0.712	0.223	0.025			

Table 6

Sensitivity analysis of the lung cancer data: in all three cases, $k = 1, ..., 5$, and $l = 0, 1, ..., 5$.

Analysis	Effect	Variable	B1 (95% CI)	B2	B3	Estimate (95% CI)	Pr(Inclusion)
	Fixed	Trtmt	−0.230 (−0.375, −0.068)	−0.193	−0.273	−0.220 (−0.365, −0.063)	1.000
$\pi_k = 1$		Bone	0.273 (0.057, 0.466)	0.293	0.313	0.279 (0.069, 0.481)	1.000
$p_{l0} = 0.5$		Liver	0.377 (0.217, 0.555)	0.496	0.338	0.411 (0.224, 0.607)	1.000
		P.S.	−0.481 (−0.625, −0.323)	−0.538	−0.500	−0.497 (−0.638, −0.343)	1.000
		W.L.	0.289 (0.158, 0.432)	0.341	0.241	0.303 (0.168, 0.442)	1.000
	Random	Baseline				0.000 (0.000, 0.000)	0.001
		Trtmt			0.007	0.000 (0.000, 0.000)	0.013
		Bone	0.160 (0.045, 0.359)		0.170	0.115 (0.000, 0.318)	0.714
		Liver			0.001	0.000 (0.000, 0.000)	0.007
		P.S.				0.000 (0.000, 0.000)	0.001
		W.L.		0.001		0.000 (0.000, 0.000)	0.001
Pr(Selection)			0.699	0.278	0.012		
	Fixed	Trtmt	−0.186 (−0.319, −0.028)			−0.101 (−0.295, 0.000)	0.547
$\pi_k = 0.5$		Bone	0.292 (0.087, 0.516)	0.233		0.199 (0.000, 0.478)	0.734
$p_{l0} = 0.2$		Liver	0.371 (0.219, 0.531)	0.370	0.414	0.400 (0.248, 0.571)	1.000
		P.S.	−0.489 (−0.621, −0.361)	−0.525	−0.535	−0.504 (−0.642, −0.357)	1.000
		W.L.	0.267 (0.140, 0.394)	0.247	0.264	0.262 (0.127, 0.394)	0.992
	Random	Baseline				0.000 (0.000, 0.000)	0.007
		Trtmt				0.000 (0.000, 0.000)	0.014
		Bone	0.176 (0.036, 0.433)	0.168	0.118	0.141 (0.000, 0.383)	0.850
		Liver				0.000 (0.000, 0.000)	0.014
		P.S.				0.000 (0.000, 0.000)	0.005
		W.L.				0.000 (0.000, 0.000)	0.013
Pr(Selection)			0.354	0.213	0.176		
	Fixed	Trtmt	−0.223 (−0.367, −0.076)			−0.128 (−0.350, 0.000)	0.605
$\pi_k = 0.5$		Bone	0.208 (0.090, 0.351)		0.192	0.131 (0.000, 0.337)	0.636
$p_{l0} = 0.8$		Liver	0.401 (0.288, 0.527)	0.604	0.463	0.476 (0.305, 0.623)	1.000
		P.S.	−0.482 (−0.581, −0.393)	−0.584	−0.554	−0.520 (−0.621, −0.406)	1.000
		W.L.	0.308 (0.148, 0.440)	0.202	0.260	0.268 (0.156, 0.427)	1.000
	Random	Baseline				0.000 (0.000, 0.000)	0.000
		Trtmt				0.000 (0.000, 0.000)	0.000
		Bone				0.000 (0.000, 0.000)	0.001
		Liver				0.000 (0.000, 0.000)	0.000
		P.S.				0.000 (0.000, 0.000)	0.000
		W.L.				0.000 (0.000, 0.000)	0.000
Pr(Selection)			0.538	0.297	0.097		