

# Modelling survival data to account for model uncertainty: a single model or model averaging?

Sri Astuti Thamrin<sup>1,2\*</sup>

\*Corresponding author

Email: [sri.thamrin@connect.qut.edu.au](mailto:sri.thamrin@connect.qut.edu.au)

James M McGree<sup>1</sup>

Email: [james.mcgree@qut.edu.au](mailto:james.mcgree@qut.edu.au)

Kerrie L Mengersen<sup>1</sup>

Email: [k.mengersen@qut.edu.au](mailto:k.mengersen@qut.edu.au)

<sup>1</sup>Mathematical Sciences, Queensland University of Technology, GPO Box 2434, 4001, Brisbane, Queensland, Australia

<sup>2</sup>Mathematical Sciences, Hasanuddin University, Makassa, Indonesia

## Abstract

This study considered the problem of predicting survival, based on three alternative models: a single Weibull, a mixture of Weibulls and a cure model. Instead of the common procedure of choosing a single “best” model, where “best” is defined in terms of goodness of fit to the data, a Bayesian model averaging (BMA) approach was adopted to account for model uncertainty. This was illustrated using a case study in which the aim was the description of lymphoma cancer survival with covariates given by phenotypes and gene expression. The results of this study indicate that if the sample size is sufficiently large, one of the three models emerge as having highest probability given the data, as indicated by the goodness of fit measure; the Bayesian information criterion (BIC). However, when the sample size is reduced, no single model was revealed as “best”, suggesting that a BMA approach would be appropriate. Although BMA approach can compromise on goodness of fit to the data (when compared to the true model), it can provide robust predictions and facilitate more detailed investigation of the relationships between gene expression and patient survival.

## Keywords

Bayesian modelling; Bayesian model averaging; Cure model; Markov Chain Monte Carlo; Mixture model; Survival analysis; Weibull distribution.

## Introduction

Modelling survival data plays an important role in the application of statistics in medicine and health science. In addition to a nonparametric formulation, there are many parametric models available for describing survival, including models based on a single distribution such as the Exponential and Weibull, mixture models based for example on mixtures of distributions and a mixture of susceptible and insusceptible individuals or so-called cure models which account for a fraction of the patients being cured from the disease. Given the wealth of models, the dilemma that is faced by many practitioners is the choice of a survival model.

The problem of model selection is abundant throughout the literature. This includes both covariate selection and choice of the model itself. Some of the methods are based on a series of significance tests while others fit more comprehensive models; some include prior information; some use analytic or approximate methods of estimation while others use Markov Chain Monte Carlo (MCMC) methods; different approaches use different optimisation or model comparison criteria such as Bayes factors (Raftery 1996). For example, McGrory and Titterton (2007) showed how variational techniques can be used to extend the deviance information criterion (DIC) to include the comparison of mixture models, a period of Basu and Tiwari (2010) used Bayes factors to compare the various model structures in breast cancer survival data.

Recently, Bonato et al. (2011) proposed Bayesian ensemble methods to obtain better survival prediction in high-dimensional gene expression data. Regardless of the method, the most common approach is to choose a single model based on the adapted optimisation or model choice criterion. However, if a single model is selected, then inferences are conditional on the selected model, and model uncertainty is ignored which often leads to excessively narrow or misleading inferences (Adrian et al. 1997; Hjort and, Claeskens 2003). This difficulty can be overcome by combining the information provided by all suitable models into the analysis. The most common way of achieving this is to use a form of model averaging. From a Bayesian point of view, this averaging is applied such that the posterior distribution of the quantity of interest is obtained over the set of suitable models, weighted by the respective posterior model probabilities (Raftery 1996).

Draper (1995) and Raftery (1995) reviewed Bayesian model averaging (BMA) and the cost of ignoring model uncertainty. Madigan and Raftery (1994) also considered BMA by using Occam's razor and Occam's window approaches to reduce the number of candidate models. Yuan and Yin (2011) used model averaging procedures to make more robust inferences regarding the dose-finding design for phase I clinical trials. Pramana et al. (2012) focused on the case in which several parametric models are fitted to gene expression data and discussed model averaging techniques for the estimation of dose-response models.

In this paper, we consider the problem of predicting survival, based on three alternative models; a single Weibull, a mixture of Weibulls and a cure model. The Weibull distribution is a popular parametric distribution for describing survival times (Dodson 1994). Given the variety of shapes that can be described by the probability density function (pdf) and the convenient representation of the distribution/survival function, the Weibull distribution has been used very effectively for analysing lifetime data, particularly when the data are censored, which is very common in most life testing experiments (Collet 1994; Kundu 2008).

Given the nature of microarray data to describe biological systems and outcomes of patients, and the potential of these covariates to produce more precise inferences about survival, the use of a single parametric distribution to describe survival time may not be adequate. Microarray data may enable the description of several homogeneous subgroups of patients with respect to survival time. This paper therefore also considered a mixture of Weibull models for precise estimation and prediction of survival. Mixture models can be used to describe a population consisting of several disjoint groups, where each group is assigned its own distribution, weighted by the probability of an individual from the overall population belonging to that group. This model thus provides a convenient and flexible mechanism for identification and estimation of distributions which are not well modelled by any standard parametric family (Stephens 1997). In the study considered here, the mixture is assumed to comprise a known number of Weibull distributions, with potentially different parameters. Most approaches to the analysis of time to event data implicitly assume all individuals will experience the event of interest. However, there are situations when a proportion of individuals are not expected to experience the event of interest; that is, those individuals are often referred to as immune, cured or nonsusceptible (Ibrahim et al. 2001).

To address this issue, cure rate models are considered, which are survival models incorporating a cure fraction. These models, which can be considered as a form of mixture model with one component degenerating to a point mass, extend the understanding of time to event data by allowing the formulation of more accurate and informative conclusions.

Finally, instead of adopting the usual practice of choosing a single “best” model, where “best” is defined in terms of the probability of the model given the data, a BMA approach was adopted to account for model uncertainty in the prediction of the response. We illustrate the approach using a microarray dataset.

The paper is organised as follows. In Section “Methods”, we define BMA. The three competing models are described in a Bayesian framework in Section “Models”. The computational approach for estimation is also presented in this section. In the Section “Application to gene expression data”, we illustrate the model using a case study. The results are discussed further in Section “Discussion”.

## Methods

The key elements of BMA were discussed by Raftery (1995). He suggested weighting each model by the posterior model probabilities derived from a Bayesian analysis. Assume that there are  $S$  models being considered, for  $s = 1, 2, \dots, S$ , each with parameter set  $\theta_s$  based on data  $D$ . Let  $\Delta$  be the quantity of interest; this could represent, for example, the posterior predictive distribution of  $y$ . Hence, the posterior distribution of  $\Delta$  given data  $D$  (Hoeting et al. 1999) is

$$p(\Delta | D) = \sum_{s=1}^S p(\Delta | S = s, D)p(S = s | D),$$

where  $p(M = s | D)$  is the posterior probability of a particular model being true, defined as

$$p(S = s | D) = \frac{p(D | S = s)p(S = s)}{\sum_{s=1}^S p(D | S = s)p(S = s)}, s = 1, 2, \dots, S,$$

where  $p(D | S = s) = \int p(D | \theta_s, S = s)p(\theta_s | S = s)d\theta_s$ .

Here,  $p(D | S = s)$  is the marginal likelihood of the data  $D$  given model  $S$  and  $p(\theta_s | S = s)$  is the prior density of  $\theta_s$  given model  $S = s$ .  $p(S = s)$  is the prior probability for models being true (Hoeting et al. 1999).

Given a model selection problem in which we have to choose between two models, the plausibility of the two different models  $S_1$  and  $S_2$  is assessed by the Bayes factor as the ratio of posterior model probabilities.

The main detractor from using Bayes factors is that they are, in general, difficult to compute. Raftery (1995) proposed using Bayesian information criterion (BIC) (Schwarz 1978) as an approximation. Buckland et al. (1997) and Claeskens and Hjort (2008) discussed the utilization of BIC in BMA. Buckland et al. (1997) proposed simpler methods where weights are based upon the penalized likelihood functions formed from the AIC (Akaike 1973).

The starting point for Burnham and Anderson’s model selection theory is the Kullback-Leibler (KL) information given by Burnham and Anderson (2002) and Claeskens and Hjort (2008):

$$I(f | q) = \int f(x) \log \frac{f(x)}{q(x | \theta_s)} dx,$$

where  $f$  represents the density function of the true and unknown model,  $q$  represents the density function of the model that is used to approximate  $f$ , and  $\theta_s$  is a vector of the unknown parameters to be estimated. The notation  $I(f | q)$  denotes the information lost when  $q$  is used to approximate  $f$  or the distance from  $q$  to  $f$ . For a given set of models, one can compare the KL information for each model and select the model that minimizes the information loss across the considered set of models (Burnham and Anderson 2002, 2004). However, in practice  $I(f | q)$  cannot be computed since the true model  $f$  is unknown. Schwarz (1978) and Burnham and Anderson (2002) made the link between the KL information and likelihood theory, and showed that the expected KL information can be expressed as

$$E(KL) = -\log p(D | \hat{\theta}_s, S = s) + d_s \log(n),$$

where  $p(D | \theta_s)$  is the likelihood,  $d_s$  is the number of parameters in the model and  $n$  is the number of uncensored observations in a survival context (Volinsky and Raftery 2000). A Laplace approximation, typically the BIC (Schwarz 1978), can be used to approximate  $p(D | S = s)$  (Clyde 2000; Hoeting et al. 1999; Jackson et al. 2009; Yuan and Yin 2011):

$$\log(p(D | S = s)) \approx \log p(D | \hat{\theta}_s, S = s) - d_s \log(n),$$

$$BIC = -2 \log p(D | \hat{\theta}_s, S = s) + d_s \log(n). \quad (1)$$

Here  $\log p(D | \hat{\theta}_s, S = s)$  is the maximised log-likelihood of model  $s$ , which estimates goodness of fit of the data.

Schwarz (1978) and Burnham and Anderson (2002) proposed the likelihood of the model given the data, using  $\hat{\theta}_s$  defined by

$$p(D | \hat{\theta}_s, S = s) \propto e^{0.5 \times BIC}. \quad (2)$$

The BMA weight for the  $s^{th}$  model (Jackson et al. 2009; Yuan and Yin 2011) is therefore given by

$$p(S = s | D) = \frac{\exp(-\frac{1}{2} BIC_s) p(S = s)}{\sum_{s=1}^S \exp(-\frac{1}{2} BIC_s) p(S = s)}.$$

The BMA weight can be interpreted as the weight of the evidence that model  $s$  is true model given a set of  $S$  models. For the case with non-informative prior probabilities, we can let  $p(S = s)$  be equal for all candidate models ( $1/S$ ), indicating no prior preference for any of the models (Jackson et al. 2009; Pramana et al. 2012). The model with the highest BMA weight will be considered as the best model. Therefore,  $p(S = s | D)$  is also an approximation to the posterior probability of the model  $s$  being correct (Schwarz 1978). A smaller BIC values indicates a better model fit, accounting for model complexity.

Let  $\tilde{f}_{sj}$  be the  $j^{th}$  simulated observation from the  $s^{th}$  model. Then, the mean of survival from the BMA model ( $\bar{f}_{MA}$ ), can be calculated as follows

$$\bar{f}_{MA} = \left( \sum_{j=1}^N \sum_{s=1}^S w_s \tilde{f}_{sj} \right) / N,$$

where  $N$  is the number of simulated observation and  $w_s = p(S = s | D)$  is the BMA weight, defined previously.

## Models

### Weibull model

In this section, we define the Weibull model for analysing survival of patients in the context of human health. We confine ourselves to survival times that are the difference between a nominated start time and a declared failure (uncensored data) or a nominated end time (censored time). Let  $T$  be a nonnegative random variable for a person's survival time and  $t$  be a realisation of the random variable  $T$ . Kleinbaum and Klein (2005) give some reasons for the occurrence of right censoring in survival studies, including termination of the study, drop outs, or loss to follow-up. For the censored observations, one could impute the missing survival times or assume that they are event-free. The former is often difficult, especially if the censoring proportion is large, and extreme imputation assumptions (such as all censored cases fail right after the time of censoring) may distort inferences (Leung et al. 1997; Stajduhar et al. 2009). In this study, we treat all censored cases as event-free regardless of observation time.

Initially, we assume that we observe survival times  $t$  of patients possibly from a heterogeneous population. The two-parameter Weibull density function for survival time is given by

$$W(t | \alpha, \gamma) = \alpha \gamma t^{\alpha-1} \exp(-\gamma t^\alpha),$$

for  $\alpha > 0$  and  $\gamma > 0$ , where  $\alpha$  is a shape parameter and  $\gamma$  is a scale parameter (Ibrahim et al. 2001).

Since the logarithm of the Weibull hazard is a linear function of the logarithm of time, it is more convenient to write the model in terms of the parameterisation  $\lambda = \log(\gamma)$  (Ibrahim et al. 2001), so that:

$$f(t | \alpha, \lambda) = \alpha t^{\alpha-1} \exp(\lambda - \exp(\lambda)t^\alpha),$$

where  $t > 0$ ,  $\alpha > 0$  and  $\gamma > 0$ .

The corresponding survival function and the hazard function, using the  $\lambda$  parameterization, are as follows:

$$\begin{aligned} S(t | \alpha, \lambda) &= \exp(-\exp(\lambda)t^\alpha), \\ h(t | \alpha, \lambda) &= f(t | \alpha, \lambda)/S(t | \alpha, \lambda) = \alpha \exp(\lambda)t^{\alpha-1}. \end{aligned}$$

We now assume that we observe possibly right-censored data for  $n$  patients;  $\mathbf{y} = (y_1, \dots, y_n)$  where  $y_i = (t_i, \delta_i)$  and  $\delta_i$  is an indicator function such that (Marin et al. 2005):

$$\delta_i = \begin{cases} 1, & \text{if the lifetime is uncensored, i.e., } T_i = t_i. \\ 0, & \text{if the lifetime is censored, i.e., } T_i > t_i. \end{cases} \quad (3)$$

Let  $x_{ij}$  be the  $j^{\text{th}}$  covariate associated with  $t_i$  for  $j = 1, 2, \dots, p+1$ . In our case study,  $x_{ij}$  indicates the  $p$  gene expressions from DNA microarray data, and  $x_{i0}$  indicates the multi-category phenotype covariate. The data structure is as follows:

$$\begin{bmatrix} \text{Survival time} \\ t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} \begin{bmatrix} \text{Category} & \text{Gene 1} & \dots & \text{Gene p} \\ x_{10} & x_{11} & \dots & x_{1p} \\ x_{20} & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n0} & x_{n1} & \dots & x_{np} \end{bmatrix}.$$

The gene expression data can be included into the model through  $\lambda$  (Thamrin et al. 2013). Given that  $\lambda$  must be positive, one option is to include the covariates as follows:

$$\begin{aligned}\gamma_i &= \exp(\mathbf{x}'_i \boldsymbol{\beta}), \text{ so that} \\ \lambda_i &= \log(\gamma_i) = \mathbf{x}'_i \boldsymbol{\beta}.\end{aligned}\tag{4}$$

Thus, the log-likelihood function becomes:

$$\begin{aligned}\log L(\alpha, \boldsymbol{\beta} \mid D) &= \sum_{i=1}^n \delta_i (\log(\alpha) + (\alpha - 1) \log(t_i) \\ &\quad + \mathbf{x}'_i \boldsymbol{\beta}) - \exp(\mathbf{x}'_i \boldsymbol{\beta}) t_i^\alpha.\end{aligned}$$

We assume that  $(\alpha, \lambda)$  are independent a priori (Marin et al. 2005), and assign Gamma distributions. Thus, the priors are now given by:

$$\begin{aligned}\alpha &\sim \text{Gamma}(u_\alpha, v_\alpha) \\ \lambda_i &\sim \text{Normal}(\mathbf{x}'_i \boldsymbol{\beta}, \sigma^2) \\ \boldsymbol{\beta} &\sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}),\end{aligned}$$

and we allow  $\boldsymbol{\Sigma}$  to be diagonal with elements  $\sigma_j^2, j = 1, 2, \dots, p$ .

Diffuse priors are represented by large positive values for  $\sigma^2$ , and small positive values for  $u_\alpha$  and  $v_\alpha$ .

The joint posterior distribution of  $(\alpha, \boldsymbol{\beta})$  is given by:

$$\begin{aligned}p(\boldsymbol{\beta}, \alpha \mid D) &\propto L(\alpha, \boldsymbol{\beta} \mid D) p(\alpha) p(\boldsymbol{\beta}) \\ &\propto \alpha^{\alpha_0 + d - 1} \exp \left\{ \sum_{i=1}^n (\delta_i \mathbf{x}'_i \boldsymbol{\beta} + \delta_i (\alpha - 1) \log(t_i) - t_i^\alpha \exp(\mathbf{x}'_i \boldsymbol{\beta})) \right. \\ &\quad \left. - b_0 \alpha - \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) \right\}.\end{aligned}$$

MCMC analysis is performed by sampling from the conditional distributions of the parameters. The conditional distribution of  $\alpha$  does not have an explicit form and as such can be sampled from approximately algorithms such as Metropolis Hastings or slice sampling (Gilks et al. 1996).

### Weibull mixture model

We define the Weibull mixture model for analysing survival data. A mixture of  $K$  Weibull densities (Marin et al. 2005) is defined by

$$f(t \mid K, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = \sum_{m=1}^K w_m W(t \mid \alpha_m, \gamma_m),\tag{5}$$

where  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$ ,  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)$ , are the parameters of each Weibull distribution and  $\mathbf{w} = (w_1, \dots, w_K)$  is a vector of nonnegative weights which sum to one.

The corresponding survival function  $S(t \mid K, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$  and hazard function  $h(t \mid K, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$  are as

follows:

$$S(t | K, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = \sum_{m=1}^K w_m \exp(-\gamma_m t^{\alpha_m}),$$

$$h(t | K, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = f(t | K, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) / S(t | K, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\gamma}).$$

We now assume that we observe possibly right-censored data for  $n$  patients;  $\mathbf{y} = (y_1, \dots, y_n)$  where  $y_i = (t_i, \delta_i)$  and  $\delta_i$  is an indicator function as described in Section “Weibull model”.

Let  $x_{ij}$  be the  $j^{\text{th}}$  covariate associated with patient  $i$ , for  $j = 1, 2, \dots, p$ . In our application,  $x_{ij}$  could indicate, for example, the gene expressions. The covariates can be included in the model as follows (Farmomeni and Nardi 2010)

$$\log(\gamma_m) = \mathbf{x}_i' \boldsymbol{\beta}_m = \lambda_m, \quad (6)$$

where  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ ,  $\boldsymbol{\gamma}_m = (\gamma_{1m}, \dots, \gamma_{pm})$  and  $\boldsymbol{\beta}_m = (\beta_{1m}, \dots, \beta_{pm})$ , for  $i = 1, 2, \dots, n$  and  $m = 1, 2, \dots, K$ .

Thus, the likelihood function becomes:

$$L(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\gamma} | K, t_i, \delta_i, \mathbf{x}) \propto \prod_{i=1}^n \left[ f(t_i | K, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{x})^{\delta_i} S(t_i | K, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{x})^{1-\delta_i} \right]$$

Here, the incomplete information is modelled via the survivor function, which reflects the probability that the patient was alive for duration greater than  $t_i$ .

The following prior distributions are placed on the parameters  $\mathbf{w}$  and  $\boldsymbol{\alpha}$ :

$$\mathbf{w} | K \sim \text{Dirichlet}(\phi_1, \dots, \phi_K), \phi_m = \phi, \forall m = 1, 2, \dots, K.$$

$$\alpha_m \sim \text{Gamma}(u_\alpha, v_\alpha), m = 1, 2, \dots, K.$$

For a model without covariates, we employ the following prior for  $\gamma_m$ .

$$\gamma_m \sim \text{Gamma}(u_\gamma, v_\gamma), m = 1, 2, \dots, K.$$

We chose small positive values for  $u_\alpha, v_\alpha, u_\gamma, v_\gamma$  to express vague prior knowledge about these parameters and we set  $\phi = 1$  (Marin et al. 2005). For a model with covariates, we employ a multivariate normal prior on  $\boldsymbol{\beta}_m$ , so that

$$\boldsymbol{\beta}_m | K \sim N(\mathbf{0}, \boldsymbol{\Sigma}),$$

and we allow  $\boldsymbol{\Sigma}$  to be diagonal with elements  $\sigma_j^2, j = 1, 2, \dots, p$ . Again, we express a vaguely informative prior by setting a large positive value for  $\sigma_j^2$ .

The model described in this section can be fitted using MCMC sampling with latent values  $Z_i$  to indicate component membership of the  $i^{\text{th}}$  observation (Diebolt and Robert 1994; Robert and Casella 2000). Since  $w_m = Pr(Z_i = m)$ , we can write  $Z_i \sim M(w_1, \dots, w_K)$ . In this scheme, the  $Z_i$  are sampled by computing posterior probabilities of membership, and the other parameters are sampled from their full conditional distributions. This was implemented in the WinBUGS software package (Spiegelhalter et al. 2002).

The WinBUGS software (Lunn et al. 2000; Ntzoufras 2009; Spiegelhalter et al. 2002), is an interactive Windows version of the BUGS program for Bayesian analysis of complex statistical models using

MCMC techniques.

Label switching, caused by non-identifiability of the mixture components, was dealt with post-MCMC using the reordering algorithm of Marin et al. (2005). The algorithm proceeded by selecting the permutation of components at each iteration that minimised the vector dot product with the so-called ‘‘pivot’’, a high density point from the posterior distribution. The MCMC output was then reordered according to each selected permutation. In this paper, the approximate maximum a posteriori (MAP) (i.e. the realization of parameters corresponding to the MCMC iterate that maximised the unnormalised posterior) was chosen as the pivot.

## Cure model

As in Section ‘‘Weibull model’’, we observe time to the event of interest for  $n$  independent subjects, and we let  $(t_i, \delta_i)$  denote the observed time and the event indicator for the  $i$ -th observation. Let  $S_1(t)$  be the survivor function for the entire population,  $S^*(t)$  be the survivor function for the non-cured group in the population, and  $\pi$  be the cure rate function. Then the standard cure rate model is given by:

$$S_1(t) = \pi + (1 - \pi)S^*(t). \quad (7)$$

The commonly used parametric distributions include Exponential and Weibull for  $S^*(t)$ .

As in Yakovlev and Tsodikov (1996), Chen et al. (1999) and Ibrahim et al. (2001), for an individual in a population, let  $N$  denote the number of latent variables. Assume that  $N$  has a Poisson distribution with mean  $\theta$ . Let  $Z_i, i = 1, \dots, N$  denote the random time, where  $Z_i$  are independently and identically distributed (i.i.d.) with a common distribution function  $F(t) = 1 - S(t)$ . Also, assume that  $Z_i$  are independent of  $N$ . The time to event can be defined by the random variable  $Y = \min(Z_i, 0 \leq i \leq N)$ , where  $P(Z_0 = \infty) = 1$ . Hence, the survival function for the population is given by

$$\begin{aligned} S_{pop}(t) &= P(N = 0) + P(Z_1 > t, \dots, Z_N > t, N \geq 1) \\ &= \exp(-\theta) + \sum_{k=1}^{\infty} [S(t)]^k \frac{\theta^k}{k!} \exp(-\theta) \\ &= \exp(-\theta F(t)). \end{aligned} \quad (8)$$

A corresponding cure fraction in model (8) is  $\lim_{t \rightarrow \infty} S_{pop}(t) = \exp(-\theta) > 0$ . We also know from (8) that the cure fraction is given by  $S_{pop}(\infty) = P(N = 0) = \exp(-\theta)$ . As  $\theta \rightarrow \infty$ , the cure fraction tends to 0, whereas as  $\theta \rightarrow 0$ , the cure fraction tends to 1. Corresponding population density and hazard functions are  $f_{pop}(t) = -\frac{d}{dt}S_{pop}(t) = \theta f(t) \exp(-\theta F(t))$  and  $h_{pop}(t) = \theta f(t)$ , respectively.

The proportional hazards structure with the covariates is modelled through  $\theta$  (Chen et al. 1999; Ibrahim et al. 2001). The population survival function (7) can be written as

$$S_{pop}(t) = \exp(-\theta) + [1 - \exp(-\theta)]S^*(t),$$

where  $S^*(t) = \frac{\exp(-\theta F(t)) - \exp(-\theta)}{1 - \exp(-\theta)}$ , and  $f^*(t) = \frac{\exp(-\theta F(t))}{1 - \exp(-\theta)} \theta f(t)$ .

Following Chen et al. (1999) and Ibrahim et al. (2001), we construct the likelihood function. Suppose we have  $n$  subjects and we assume that  $N_i$ 's are i.i.d with Poisson distributions with means  $\theta_i, i = 1, \dots, n$ . Let  $Z_{i1}, \dots, Z_{iN}$  denote the times for the  $N_i$  competing causes, which are unobserved, and which have a cumulative distribution function,  $F(\cdot)$ . In this section, we will specify a parametric form for  $F(\cdot)$  that is a Weibull distribution. Let  $\psi = (\alpha, \lambda)'$ , where  $\alpha$  is the shape parameter and  $\lambda$  is the scale parameter. We incorporate covariates for the cure rate model through the cure parameter  $\theta$  and we



have a different cure rate parameter,  $\theta_i$ , for each subject.

Let  $\mathbf{x}'_i = (x_{i1}, \dots, x_{ik})$  denote the  $k \times 1$  vector of covariates for the  $i$ th subject, and let  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)$  denote the corresponding vector of regression coefficients. We relate  $\theta$  to the covariates by  $\theta_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$ . Let  $t_i$  denote the survival time for subject  $i$ , which is right censored, let  $C_i$  be the censoring time, and let  $\delta_i$  be the censoring indicator, assuming 1 if  $T_i$  is a failure time and 0 if it is right censored. The observed data are  $D = (n, \mathbf{t}, \boldsymbol{\delta}, \mathbf{X})$ , where  $\mathbf{t} = (t_1, \dots, t_n)'$ ,  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)'$  and  $\mathbf{X} = (x_1, \dots, x_n)'$ . The complete data are given by  $D_c = (n, \mathbf{t}, \boldsymbol{\delta}, \mathbf{X}, \mathbf{N})$ , where  $\mathbf{N} = (N_1, \dots, N_n)'$ . The complete-data likelihood function of the parameter  $(\boldsymbol{\psi}, \boldsymbol{\beta})$  can be written as

$$L(\boldsymbol{\psi}, \boldsymbol{\beta} | D_c) = \left\{ \prod_{i=1}^n S(t_i | \boldsymbol{\psi})^{N_i - \delta_i} (N_i f(t_i | \boldsymbol{\psi}))^{\delta_i} \right\} \times \exp \left\{ \sum_{i=1}^n N_i \log(\theta_i) - \log(N_i!) - n\theta_i \right\}. \quad (9)$$

Again, we assume independent priors for  $\boldsymbol{\beta}$  and  $\boldsymbol{\psi}$ , where  $\alpha \sim \text{Gamma}(a_\alpha, b_\alpha)$ ,  $\lambda \sim N(\boldsymbol{\mu}_\lambda, \boldsymbol{\Sigma}_\lambda)$  and  $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$ . We also assume  $p(\alpha, \lambda) = p(\alpha | \delta_0, \tau_0)p(\lambda)$ ,  $p(\alpha | \delta_0, \tau_0) \propto \alpha^{\delta_0 - 1} \exp(-\tau_0 \alpha)$ , and the hyperparameters  $(\delta_0, \tau_0)$  are specified (Chen et al. 1999; Ibrahim et al. 2001).

Combining these specifications with the likelihood function (9), the joint posterior distribution of  $(\alpha, \lambda, \boldsymbol{\beta})$  becomes

$$p(\alpha, \lambda, \boldsymbol{\beta} | D) \propto \prod_{i=1}^n (\theta_i f(t_i | \alpha, \lambda))^{\delta_i} \exp(-\theta_i(1 - S(t_i | \alpha, \lambda))) \times p(\alpha | \delta_0, \tau_0)p(\alpha, \lambda)p(\boldsymbol{\beta}). \quad (10)$$

The joint posterior density of  $(\alpha, \lambda, \boldsymbol{\beta})$  in equation (10) is analytically intractable because the integration of the joint posterior density is not easy to perform. Hence, inferences are based on MCMC simulation methods. We can use the Metropolis-Hastings algorithms or slice sampling to simulate samples of  $\alpha$ ,  $\lambda$  and  $\boldsymbol{\beta}$ . MCMC computations were implemented using the WinBUGS system (Spiegelhalter et al. 2002).

## Application to gene expression data

### DLBCL dataset

We applied the proposed method of model averaging across the three candidate survival models to a dataset containing gene expression of Diffuse Large B-cell Lymphoma (DLBCL). The dataset comprises gene expression measurements and survival times of patients with DLBCL (Rosenwald et al. 2002). DLBCL (Lenza et al. 2008) is a type of cancer of the lymphatic system in adults which can be cured by anthracycline-based chemotherapy in only 35 to 40 percent of patients (Rosenwald et al. 2002). In general, types of this disease are very diverse and their biological properties are largely unknown, meaning that this is a relatively difficult cancer to cure and prevent. Rosenwald et al. (2002) proposed that there are three phenotypes subgroups of patients of DLBCL: activated B-like DLBCL, germinal centre (GC)-B like and type III DLBCL. The GC B-like DLBCL is less dangerous than the others in the progression of the tumour; the activated B-like DLBCL is more active than the others and the type III DLBCL is the most dangerous in the progression of tumour (Alizadeh et al. 2000). These groups were defined using microarray experiments and hierarchical clustering. The authors showed that these phenotypes subgroups were differentiated from each other by distinct gene expressions of hundreds of

different genes and had different survival time patterns. This dataset contains 219 patients with DLBCL, including 138 patient deaths during follow-up. Patients with missing values for a particular microarray element were excluded from all analyses involving that element.

Based on patterns of gene expression in biopsy specimens of the lymphoma, Rosenwald et al. (2002) analysed this dataset to predict the likelihood of patients' survival after chemotherapy for DLBCL. By using a Cox proportional-hazards model, Rosenwald et al. (2002) identified five individual gene expressions which correlated with the survival after chemotherapy. These gene expressions are germinal center B-cell (GC-B), lymphoma node, proliferation, BMP6 and MHC. In this study, these five gene expressions are used as covariates for estimating survival times based on the three competing models in Section "Models".

## Results

As discussed in Section "Methods", to account for model uncertainty, the model averaging technique which combines estimates from different survival models was carried out. This was accomplished through a weighted average of the survival considered in the analysis. First, we calculated the Kaplan-Meier estimates of overall survival according to the gene expression and the relation between the gene expression score and the subgroups phenotype of DLBCL. We confirmed that these phenotypes had different survival time patterns (Figures 1). Following this, we fitted the three models either to all gene expression data and to the three phenotype subgroups. We then applied the BMA approach described in Section "Methods". For each model, we ran the corresponding MCMC algorithm for 100 000 iterations, discarding the first 10 000 iterations as burn-in.

---

### **Figure 1 Kaplan-Meier estimates of overall survival according to the gene-expression subgroups.**

---

The model averaged prediction for the response was calculated according to equation (). Table 1 shows the estimated posterior mean of the parameters, the 95% credible intervals (CI), the BIC values and the BMA weights for each of the fitted models for the whole dataset. The BMA weights reflect the relative posterior probability of the models. As can be seen from Table 1, for the Weibull model, there are three genes that substantially describe patients' survival times, namely GC-B ( $\beta_1$ ), lymphoma node ( $\beta_2$ ) and MHC ( $\beta_5$ ). These three genes have a negative effect on the expected survival time. For the mixture model, GC-B ( $\beta_1$ ), lymphoma node ( $\beta_2$ ) and proliferation ( $\beta_3$ ) accounted for patients' survival times in the first component. In the second component, GC-B ( $\beta_1$ ), lymphoma node ( $\beta_2$ ) and MHC signature ( $\beta_5$ ) substantially explained patients' survival times. All these genes have negative effects on the expected survival time for their respective component. For the cure model, four of these genes substantially describe patients' survival times, namely GC-B ( $\beta_1$ ), lymphoma node ( $\beta_2$ ), BMP6 ( $\beta_4$ ) and MHC ( $\beta_5$ ) signature. Three of these, namely GC-B ( $\beta_1$ ), lymphoma node ( $\beta_2$ ) and MHC signature ( $\beta_5$ ), have a negative effect on the expected survival time. Under the cure model, approximately 33.8% of the patients are cured of DLBCL (Figure 2).

**Table 1** The estimated posterior mean of the parameters, the 95% credible intervals (CI), the BIC values and the BMA weights for each of the fitted models for the full DLBCL dataset

Model	Parameter	Mean	95% CI	BIC	Weight
Weibull	$\alpha$	0.7305	(0.626, 0.840)	687.0953	0.0009
	$\beta_0$	-1.578	(-1.84, -1.33)		
	$\beta_1$	-0.3446	(-0.516, -0.172)		
	$\beta_2$	-0.2844	(-0.454, -0.116)		
	$\beta_3$	0.2097	(-0.049, 0.468)		
	$\beta_4$	0.3292	(0.115, 0.537)		
	$\beta_5$	-0.3019	(-0.488, -0.112)		
Mixture	$\alpha_1$	4.029	(2.411, 6.631)	734.0054	$\approx 0$
	$\alpha_2$	0.7707	(0.662, 0.885)		
	$\beta_{01}$	6.857	(5.479, 8.205)		
	$\beta_{02}$	-1.724	(-2.007, -1.457)		
	$\beta_{11}$	-11.62	(-12.88, -10.35)		
	$\beta_{12}$	-0.3956	(-0.575, -0.216)		
	$\beta_{21}$	-2.087	(-3.54, -0.689)		
	$\beta_{22}$	-0.3172	(-0.495, -0.143)		
	$\beta_{31}$	-2.241	(-3.425, -1.059)		
	$\beta_{32}$	0.1972	(-0.064, 0.461)		
	$\beta_{41}$	-0.2849	(-1.434, 0.854)		
	$\beta_{42}$	0.3594	(0.141, 0.574)		
	$\beta_{51}$	-0.7928	(-2.107, 0.477)		
	$\beta_{52}$	-0.3102	(-0.500, -0.115)		
	$\pi_1$	0.01992	(0.002, 0.053)		
$\pi_2$	0.9801	(0.946, 0.997)			
Cure	$\alpha$	0.9884	(0.828, 1.145)	673.1359	0.9991
	$\beta_0$	0.1611	(-0.124, 0.560)		
	$\beta_1$	-0.3151	(-0.484, -0.144)		
	$\beta_2$	-0.2821	(-0.451, -0.115)		
	$\beta_3$	0.189	(-0.070, 0.442)		
	$\beta_4$	0.3303	(0.118, 0.539)		
	$\beta_5$	-0.3039	(-0.490, -0.112)		

**Figure 2** Box-plots of the cure rates (posterior distribution of  $\pi$ ) for the full DLBCL dataset, and to each of the three phenotypes (ABC, GCB and Type III).

This is clearly exhibited in Table 1, which shows that the cure model has the largest posterior model probability (or BMA weight). To evaluate the model fit, a comparison of the density prediction for the models to the observed data was carried out.

Table 2 show the 95% CI, BIC values and the BMA weights for each of the models based on phenotype for the DLBCL dataset. In general, for all phenotypes, the mixture model is not favourable as its weight value is approximately equal to zero and it has the largest BIC value. On the other hand, the BIC values of the other two models are close to each other, suggesting a combination of these two model in order to account for the uncertainty in the prediction of survival.

**Table 2 The estimated posterior mean of parameters, the 95% CI, BIC values and the BMA weights for each of the models based on phenotype for the DLBCL dataset**

Phenotype	Model	Variable	Parameter	Mean	95% CI	BIC	Weight			
GCB	Weibull		$\alpha$	0.692	(0.5365, 0.8595)	341.212	0.497			
		Intercept	$\beta_0$	-1.649	(-2.185, -1.17)					
		GCB	$\beta_1$	-0.179	(-0.5859, 0.239)					
		Lymphoma	$\beta_2$	-0.118	(-0.3958, 0.1607)					
		Proliferation	$\beta_3$	0.459	(-0.0306, 0.934)					
	Mixture		BMP6	$\beta_4$	0.414	(0.01773, 0.809)	377.759	$\approx 0$		
			MHC	$\beta_5$	-0.325	(-0.6389, -0.01228)				
			$\alpha_1$	4.252	(2.591, 7.175)					
			$\alpha_2$	0.816	(0.6209, 1.032)					
		Intercept	$\beta_{01}$	6.491	(5.246, 7.781)					
			$\beta_{02}$	-2.152	(-2.798, -1.567)					
		GCB	$\beta_{11}$	-11.81	(-13.05, -10.53)					
			$\beta_{12}$	-0.030	(-0.5104, 0.4592)					
		Lymphoma	$\beta_{21}$	-1.839	(-3.082, -0.6744)					
			$\beta_{22}$	-0.134	(-0.48, 0.2254)					
		Proliferation	$\beta_{31}$	-2.165	(-3.313, -0.9932)					
			$\beta_{32}$	0.588	(-0.07796, 1.407)					
		BMP6	$\beta_{41}$	-0.242	(-1.357, 0.8482)					
			$\beta_{42}$	0.654	(0.17, 1.161)					
		MHC	$\beta_{51}$	-0.629	(-1.993, 0.5117)					
			$\beta_{52}$	-0.382	(-0.7687, -0.002227)					
		Cure		$\phi_1$	0.090	(0.02007, 0.1863)			341.188	0.503
				$\phi_2$	0.91	(0.8137, 0.9799)				
				$\alpha$	0.845	(0.6075, 1.1)				
				Intercept	$\beta_0$	0.604				
	GCB			$\beta_1$	-0.173	(-0.5754, 0.2402)				
	Lymphoma			$\beta_2$	-0.116	(-0.3891, 0.1579)				
	Proliferation			$\beta_3$	0.433	(-0.0522, 0.9041)				
	BMP6			$\beta_4$	0.396	(-0.0007, 0.788)				
	MHC			$\beta_5$	-0.330	(-0.6422, -0.0209)				
ABC	Weibull		$\alpha$	0.894	(0.695, 1.115)	215.564	0.013			
		Intercept	$\beta_0$	-1.86	(-2.562, -1.217)					
		GCB	$\beta_1$	-0.509	(-0.9948, -0.03871)					
		Lymphoma	$\beta_2$	-0.626	(-0.9568, -0.3099)					
		Proliferation	$\beta_3$	-0.487	(-1.118, 0.1422)					
	Mixture		BMP6	$\beta_4$	0.645	(0.2725, 1.021)	256.552	$\approx 0$		
			MHC	$\beta_5$	-0.479	(-0.7955, -0.1598)				
			$\alpha_1$	2.427	(1.083, 4.152)					
			$\alpha_2$	0.960	(0.7525, 1.189)					
		Intercept	$\beta_{01}$	6.636	(5.301, 7.959)					
			$\beta_{02}$	-2.572	(-3.346, -1.865)					
		GCB	$\beta_{11}$	-12.11	(-13.36, -10.86)					
			$\beta_{12}$	-0.925	(-1.438, -0.4356)					
		Lymphoma	$\beta_{21}$	-3.155	(-4.578, -1.75)					
			$\beta_{22}$	-0.768	(-1.114, -0.4341)					
		Proliferation	$\beta_{31}$	-2.377	(-3.561, -1.188)					
			$\beta_{32}$	-0.480	(-1.099, 0.1353)					
		BMP6	$\beta_{41}$	0.079	(-1.064, 1.232)					
			$\beta_{42}$	0.690	(0.3249, 1.061)					
		MHC	$\beta_{51}$	-0.644	(-1.919, 0.6499)					
			$\beta_{52}$	-0.515	(-0.8176, -0.2047)					
			$\phi_1$	0.037	(0.0046, 0.09883)					
			$\phi_2$	0.963	(0.9012, 0.9953)					

Cure		$\alpha$	1.189	(0.8906, 1.483)	206.961	0.987	
	Intercept	$\beta_0$	0.019	(-0.6417, 0.7362)			
	GCB	$\beta_1$	-0.432	(-0.8874, 0.01376)			
	Lymphoma	$\beta_2$	-0.587	(-0.905, -0.2867)			
	Proliferation	$\beta_3$	-0.484	(-1.076, 0.1012)			
	BMP6	$\beta_4$	0.607	(0.2557, 0.9631)			
	MHC	$\beta_5$	-0.446	(-0.7481, -0.1346)			
Type III	Weibull	$\alpha$	0.834	(0.5958, 1.101)	162.27	0.538	
		Intercept	$\beta_0$	-1.75			(-2.736, -0.9093)
		GCB	$\beta_1$	-0.404			(-1.028, -0.19)
		Lymphoma	$\beta_2$	-0.274			(-0.7404, 0.1644)
		Proliferation	$\beta_3$	0.506			(-0.0897, 1.084)
		BMP6	$\beta_4$	0.017			(-0.5301, 0.5206)
		MHC	$\beta_5$	-0.199			(-0.6839, 0.3098)
	Mixture	$\alpha_1$	11.82	(8.609, 15.14)	196.271	$\approx 0$	
		$\alpha_2$	0.596	(0.43, 0.7757)			
		Intercept	$\beta_{01}$	6.002			(3.682, 8.336)
			$\beta_{02}$	-5.005			(-7.19, -2.812)
		GCB	$\beta_{11}$	-9.32			(-12.02, -6.611)
			$\beta_{12}$	0.564			(0.1829, 1.004)
		Lymphoma	$\beta_{21}$	-2.913			(-5.716, -0.02716)
			$\beta_{22}$	-0.558			(-1.015, -0.1525)
		Proliferation	$\beta_{31}$	-2.021			(-4.547, 0.455)
			$\beta_{32}$	0.893			(0.3455, 1.515)
		BMP6	$\beta_{41}$	0.320			(-2.466, 3.373)
			$\beta_{42}$	0.140			(-0.2735, 0.5384)
		MHC	$\beta_{51}$	-0.336			(-2.733, 2.323)
			$\beta_{52}$	-0.293			(-0.7741, 0.1504)
	Cure	$\phi_1$	0.072	(0.0108, 0.1805)	162.578	0.462	
		$\phi_2$	0.928	(0.8195, 0.9891)			
		$\alpha$	0.969	(0.6534, 1.339)			
		Intercept	$\beta_0$	0.989			(-0.5077, 4.153)
		GCB	$\beta_1$	-0.349			(-0.973, -0.25)
		Lymphoma	$\beta_2$	-0.269			(-0.7375, 0.1687)
		Proliferation	$\beta_3$	0.502			(-0.0955, 1.084)
BMP6		$\beta_4$	0.046	(-0.4801, -0.1801)			
MHC		$\beta_5$	-0.183	(-0.6625, 0.3207)			

From Tables 1 and 2, we can see that the Weibull model is better than a two-component Weibull mixture model, suggesting the two-component mixture Weibull is optimal as competing model.

As can be seen in Figure 3, in the full DLBCL dataset, the predicted curve for the cure model quite close to the observed data, suggesting a good fit of the data. Specifically, in this model, 94.354% of observed survival times in the dataset fall in the corresponding 95% prediction intervals. As expected, this is quite similar to the result obtained from model averaging (91.935%) (Table 3).

**Figure 3 The posterior densities of the three models and the model averaged density for the full DLBCL dataset and each of the three phenotypes.** For comparison, the observed data is also represented as a histogram.

**Table 3 The percentage of observed values that lay in the corresponding 95% credible interval for the individual models and BMA model based on the full DLBCL dataset and each of the three phenotypes**

Model	All DLBCL	GCB	ABC	Type III
Weibull	87.516	90	89.130	89.286
Mixture	85.903	88	86.956	82.142
Cure	94.354	92	91.304	85.714
BMA	91.935	94	93.478	92.857

Furthermore, in the GCB phenotype, the genes corresponding to the BMP6 ( $\beta_4$ ) and MHC signature ( $\beta_5$ ) in the Weibull model and MHC signature ( $\beta_5$ ) in the cure model substantially affect patients' survival time. In the ABC phenotype, in the Weibull model, with the exception of proliferation ( $\beta_3$ ), all genes were involved substantially in the description of patients' survival and lymphoma node ( $\beta_2$ ), BMP6 ( $\beta_4$ ) and MHC signature ( $\beta_5$ ) are potentially important prognostic factors for predicting survival in the cure model. For the type III phenotype, the GC-B gene ( $\beta_1$ ) in both models and only the BMP6 gene ( $\beta_4$ ) in the cure model are substantial in explaining the survival times of the patients.

Under the cure model, in the GCB phenotype, approximately 33.2% of the patients are estimated to be cured of DLBCL. In the ABC and type III phenotypes, the respective cure rates are approximately 26.57% and 18.7% (Figure 2).

The results of the posterior densities prediction for the individual models and the model averaged prediction based on these three phenotypes are presented in Figure 3. In comparison to other models, the mixture model fitted the data poorly for each phenotype. In detail, using model averaging, for the GCB phenotype, 94% of the observed survival times in the dataset lie in the respective 95% prediction intervals. For the other two phenotypes, namely the ABC and the type III, 93.478% and 92.857% of the observed survival times in the dataset are in the corresponding 95% prediction intervals, respectively (Table 3).

## Discussion

This study has adopted a Bayesian model averaging approach to account for model uncertainty in the prediction of survival. The case study that we considered involved lymphoma cancer survival, with covariates given by phenotypes and gene expressions. Here, we proposed three competing models and used BMA to combine these models to account for model uncertainty.

Overall, the results of this study indicate that if using the full dataset without further grouping, selecting a single model that best fits the data was adequate. The reason is that there is clear support for one model (i.e. only one model has a relatively larger BIC value and dominates based on this criterion). However, the results were different when the model selection process took into account the phenotype subgroups of the patients. A single model appeared to be inadequate. This was due to the fact that the values of BIC for the Weibull and the cure had nearly equal weight, indicating the absence of a dominant model based on this criterion and the presence of uncertainty issues in the model selection. As suggested and shown in this study, BMA was used to address this problem. The applicability of BMA was also associated with the smaller number of samples in each phenotype subgroup (Annest et al. 2009; Volinsky et al. 1997; Yeung et al. 2005).

This study also revealed that in each phenotype, the expression and number of predictor genes substantially describing the survival times of the patients varied across models. Overall, in both of the favourable models, none of the genes existed consistently as substantial predictors for the patients' sur-

vival. For example, in the Weibull model, the MHC and BMP genes in the GCB and ABC phenotypes and the GCB genes in the ABC and Type III phenotypes were important predictors of survival. In contrast, in the cure model, BMP was substantially associated with predicted survival in the ABC and Type III phenotypes. For both models, only three genes i.e. lymphoma node, BMP6 and MHC signature in the ABC phenotype were highly associated with the survival times of the patients.

This study has indicated that the application of BMA to combine competing models overcomes the problem of uncertainty. Comparison of different survival models has allowed the identification and analysis of the more detailed relationships between gene expressions in given phenotypes and the survival times of the patients. An advantage of BMA is more accurate and precise prediction of patient survival. However, this study only involved three candidate models. More models can be obviously included in the analysis. This study has also focused on the marginal likelihood  $p(D | Q_s)$  estimation methods based on the Laplace approximation. However, other approaches are also possible. Indeed marginal likelihood estimation is possible using nested sampling (Skilling 2006), where the marginal likelihood is viewed as the expectation, with respect to the prior, of the likelihood. The other approach is Chib's method (Chib 1995), that presented a generic method which can be applied to output from the Gibbs sampler. Applying BMA to other datasets or other applications is desired to obtain robust predictions.

### **Competing interests**

The authors declare that they have no conflict of interest.

### **Acknowledgements**

The first author acknowledges the Ministry of National Education Indonesia (DIKTI) for providing funding for this research as well as Dr. Christopher Drovandi from Queensland University of Technology for his generous feedback.

### **References**

- Raftery AE, Madigan D, Hoeting JA (1997) Bayesian model averaging for linear regression models. *J Am Stat Assoc* 92(437):179–191.
- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csaki (Eds). *Second Int Symp Inform Theory* :267–281.
- Alizadeh AA, Eisen MB, Davis RE (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403:503–511.
- Annest A, Bumgarner RE, Raftery AE, Yeung KE (2009) Iterative Bayesian Model Averaging: A method for the application of survival analysis to high-dimensional microarray data. *BMC Bioinformatics* 10(72):1471–2105.
- Basu S, Tiwari RC (2010) Breast cancer survival, competing risk and mixture cure model: A Bayesian analysis. *J R Stat Soc* 173(2):307–329.
- Bonato V, Baladandayuthapani V, Broom BM, Sulman EP, Aldape KD, Do KA (2011) Bayesian Ensemble methods for survival prediction in gene expression data. *Bioinformatics* 27(3):359–367.
- Buckland S, Burnham K, Augustin N (1997) Model selection: An integral part of inference. *Biometrics* 53:603–618.

- Burnham K, Anderson D (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New-York.
- Burnham K, Anderson D (2004) Multimodel inference: Understanding AIC and BIC in model selection. *Sociol Methods Res* 33:261–304.
- Chen MH, Ibrahim JG, Sinha D (1999) A new Bayesian model for survival data with a surviving fraction. *J Am Stat Assoc* 94:909–919.
- Chib S (1995) Marginal likelihood from the Gibbs output. *J Am Stat Assoc* 90:1313–1321.
- Claeskens G, Hjort NL (2008) *Model Selection and Model Averaging*. Cambridge University Press.
- Clyde M (2000) Model uncertainty and health effect studies for particulate matter. *Environmetrics* 11:745–763.
- Collet D (1994) *Modelling Survival Data in Medical Research*. Chapman and Hall, 1st edition.
- Diebolt J, Robert CP (1994) Estimation of finite mixture distributions through Bayesian sampling. *J R Stat Soc Series B* 56:363–375.
- Dodson B (1994) *Weibull Analysis*. American Society Quality, Milwaukee
- Draper D (1995) Assessment and propagation of model uncertainty. *J R Stat Soc Series B* 57:45–97.
- Farmomeni A, Nardi A (2010) A two-component Weibull mixture to model early and late mortality in a Bayesian framework. *Comput Stat Data Anal* 54:416–428.
- Gilks WR, Richardson S, Spiegelhalter DJ (1996) *Markov chain monte carlo in practice*. Chapman and Hall.
- Hjort NL, Claeskens G (2003) Frequentist model average estimators (with discussion). *J Am Stat Assoc* 98:879–899.
- Hoeting J, Madigan D, Raftery AE, Volinsky C (1999) Bayesian Model Averaging: A Tutorial. *Stat Sci* 14:382–417.
- Ibrahim JG, Chen MH, Sinha D (2001) *Bayesian Survival Analysis*. Springer.
- Jackson CH, Thompson SG, Sharples LD (2009) Accounting for uncertainty in health economic decision models by using model averaging. *J R Stat Soc Series A* 172(2):383–404.
- Kundu D (2008) Bayesian inference and life testing plan for the Weibull distribution in presence of progressive censoring. *Technometrics* 50(2):144–154.
- Kleinbaum DG, Klein M (2005) *Survival Analysis: A Self-learning Text*. Springer.
- Lenza G, Wrightb GW, Emrea NT, Kohlhammera H, Davea SS, Davisa RE, Cartya S, Lama LT, Shaffera AL, Xiaoc W, Powellc J, Rosenwald A, Ott G, Muller-Hermelink HK, Gascoynef RD, Connorsf JM, Campog E, Jaffe h ES, JDelabiei, Smelandj EB, Rimsza LM, Fisherm RI, Weisenburgero DD, Chano WC, Staudta LM (2008) Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci U S A* 105(36):13520–13525.
- Leung KM, Elashoff RM, Afifi AA (1997) Censoring issues in survival analysis. *Ann Rev Public Health* 18:83–104.
- Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS – A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat Comput* 10:325–337.



- Madigan D, Raftery AE (1994) Model selection and accounting for model uncertainty in graphical models using Occam's window. *J Am Stat Assoc* 89:1535–1546.
- Marin JM, Bernal MR, Wiper MP (2005) Using Weibull mixture distributions to model heterogeneous survival. *Commun Stat Simul Comput* 34(3):673–684.
- Marin JM, Mengersen K, Robert CP (2005) *Bayesian Modelling and Inference on Mixtures of Distributions. Handbook of Statistics 25, D. Dey and C.R. Rao (eds)*. Elsevier-Sciences.
- McGrory CA, Titterington DM (2007) Variational approximations in Bayesian model selection for finite mixture distributions. *Comput Stat Data Anal* 51(11):5352–5367.
- Ntzoufras I (2009) *Bayesian Modelling using WinBUGS*. Wiley, New Jersey.
- Pramana S, Shkedy Z, Göehlmann HW, Talloen W, Bondt AD, Straetmans R, Lin D, Pinheiro J (2012) Model-based Parametric Approaches. In D. Lin, Z. Shkedy, D. Yekutieli, D. Amaratunga and L. Bijnsens (Ed.). *Modeling Dose-response Microarray Data in Early Drug Development Experiments Using R* (1st ed., pp. 231-249)
- Raftery AE (1995) Bayesian model selection in social research (with discussion). *Soc Methodol* 25:111–163.
- Raftery AE (1996) Approximate Bayes factors and accounting for model uncertainty in generalized linear models. *Biometrika* 83:251–266
- Robert CP, Casella G (2000) *Monte Carlo Statistical Methods*. Springer, New York.
- Rosenwald A, Wright G, Wiestner A, Chan WC, Connors JM, Campo E, Gascoyne RD, Grogan TM, Muller HK, Smeland EB, Chiorazzi M, Giltnane JM, Hurt EM, Zhao H, Averett L, Henrickson S, Yang L, Powell J, Wilson WH, Jaffe ES, Simon R, Klausner RD, Montserrat E, Bosch F, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J et al. (2002) The use Of molecular profiling to predict survival after chemotherapy for diffuse large B-cell lymphoma. *N Engl J Med* 346(25):1937–1947.
- Schwarz GE (1978) Estimating the dimension of a model. *Ann Stat* 6(2):461–464.
- Skilling J (2006) Nested sampling for general Bayesian computation. *Bayesian Anal* 1:833–860.
- Spiegelhalter D, Best N, Carlin B, VanderLinde A (2002) Bayesian measures of model complexity and fit. *J R Stat Soc Series B* 64(4):583–639.
- Stephens M (1997) *Bayesian Methods for Mixtures of Normal Distributions*. PhD thesis, The University of Oxford.
- Stajduhar I, Basic BD, Bogunovic N (2009) Impact of censoring on learning Bayesian networks in survival modelling. *Artif Intell Med* 47:199–217.
- Thamrin SA, McGree JM, Mengersen KL (2013) Bayesian Weibull Survival Model for gene expression data. In C. L. Alston, K. L. Mengersen and A. N. Pettitt (Ed.). *Case Studies in Bayesian Statistical Modelling and Analysis* (1st ed., pp. 171-185).
- Volinsky CT, Raftery AE (2000) Bayesian Information Criterion for Censored Survival Models. *Biometrics* 56:256–262.
- Volinsky C, Madigan D, Raftery AE, Kronmal R (1997) Bayesian model averaging in proportional hazard models: Assessing the risk of a stroke. *Appl Stat* 46(4):443–448.

- Yakovlev AY, Tsodikov AD (1996) *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, Singapore.
- Yeung KE, Bumgarner R, Raftery AE (2005) Bayesian model averaging: Development of an Improved Multi-Class, Gene selection and classification tool for microarray. *Bioinformatics* 21:2394–2402.
- Yuan Y, Yin G (2011) Robust EM continual reassessment method in oncology dose finding. *J Am Stat Assoc* 108(495):818–831.