Computational Statistics and Data Analysis 71 (2014) 324–334

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A hierarchical Bayes model for biomarker subset effects in clinical trials

Bingshu E. Chen [a,b,*], Wenyu Jiang [c], Dongsheng Tu [a,b]

[a] NCIC Clinical Trials Group, Queens University, Kingston, Ontario, Canada
[b] Department of Public Health Sciences, Queens University, Kingston, Ontario, Canada
[c] Department of Mathematics and Statistics, Queens University, Kingston, Ontario, Canada

ARTICLE INFO

Article history:
Received 17 May 2012
Received in revised form 18 January 2013
Accepted 18 May 2013
Available online 31 May 2013

Keywords:
Biomarker
Clinical trials
Gibbs sampling
Hierarchical Bayes model
Markov Chain Monte Carlo
Survival analysis
Subset treatment effect

ABSTRACT

Some baseline patient factors, such as biomarkers, are useful in predicting patients' responses to a new therapy. Identification of such factors is important in enhancing treatment outcomes, avoiding potentially toxic therapy that is destined to fail and improving the cost-effectiveness of treatment. Many of the biomarkers, such as gene expression, are measured on a continuous scale. A threshold of the biomarker is often needed to define a sensitive subset for making easy clinical decisions. A novel hierarchical Bayesian method is developed to make statistical inference simultaneously on the threshold and the treatment effect restricted on the sensitive subset defined by the biomarker threshold. In the proposed method, the threshold parameter is treated as a random variable that takes values with a certain probability distribution. The observed data are used to estimate parameters in the prior distribution for the threshold, so that the posterior is less dependent on the prior assumption. The proposed Bayesian method is evaluated through simulation studies. Compared to the existing approaches such as the profile likelihood method, which makes inferences about the threshold parameter using the bootstrap, the proposed method provides better finite sample properties in terms of the coverage probability of a 95% credible interval. The proposed method is also applied to a clinical trial of prostate cancer with the serum prostatic acid phosphatase (AP) biomarker.

1. Introduction

Human diseases such as cancer are often heterogeneous in their properties. Consequently, different patients may respond differently to the same treatment. In a randomized clinical trial evaluating a new treatment, after primary analysis is performed on all patients in the trial, subset analyses based on some pre-treatment characteristics of the patients are often conducted to identify a subset of patients who may benefit more from the new treatment than other patients. On the other hand, subset analysis can also be useful to identify patients who may not benefit from a treatment, thus making it possible to avoid a potentially toxic therapy which is destined to fail. For example, after a clinical trial conducted by the NCIC Clinical Trials Group showed that a new chemotherapy regimen (CEF) improved both the disease-free and overall survivals compared to the classical chemotherapy regimen (CMF) in pre-menopausal women with stage II breast cancer (Levine et al., 2005), a subset analysis was performed and demonstrated that only patients whose breast-cancer cells had amplification of a gene HER2 benefited from the CEF treatment (Pritchard et al., 2006). Based on this analysis, it was recommended that patients without HER2 amplification could be treated with the less toxic CMF regimen, whereas those with amplified HER2

* Correspondence to: Department of Community Health and Epidemiology, Queens University, Kingston, Ontario, Canada. Tel.: +1 613 533 6430.
E-mail address: bechen@ctg.queensu.ca (B.E. Chen).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.05.015

should receive the dose-intensive CEF treatment. An important aspect of subset analysis is to study the treatment–biomarker interaction, in order to describe the different responses to the treatment among the biomarker defined patient subsets (Wang et al., 2007; Werft et al., 2012).

A biomarker is a baseline patient characteristic that affects a patient's response to the treatment, and is often measured on a continuous scale. For example, in the study of a treatment for a specific type of cancer, the biomarker could be the expression level of a particular gene. Several approaches have been proposed to investigate the interaction between the treatment and a continuous biomarker. Gray (1992) used splines in additive models for the analysis of survival data under the proportional hazards assumption. Bonetti and Gelber (2000) presented a sub-population treatment effect pattern plot (STEPP) to graphically assess treatment–biomarker interactions using the Cox proportional hazards regression model. It is also possible to model the biomarker–treatment interaction in a continuous fashion. Royston and Sauerbrei (2004) proposed a new approach to model interactions between treatment and a biomarker by using fractional polynomials (FP). Treating a biomarker in a continuous fashion has the advantage of making use of all potential information provided by the data (Royston et al., 2006). However, for making medical and clinical decisions in practice, it is important to adopt a threshold parameter to define clearly the subset of patients that may benefit from a certain treatment, which is called the sensitive subset henceforth. The threshold could be determined by a simple grid search method, in which the threshold is obtained by the point yielding the smallest p-value in the interaction tests across an arbitrary finite number of candidate points. Faraggi and Simon (1996) suggested that this approach may overestimate the true treatment effect in the subsets and proposed a two-fold cross-validation method for estimating and testing the subset treatment effect. For sensitive subsets defined on an interval with two threshold parameters, Wacholder et al. (2010) developed a permutation test for the biomarker–treatment interaction. Jiang et al. (2007) studied a single candidate biomarker scenario with an unknown threshold parameter defining the sensitive subset. They focused primarily on test procedures, and if the test detects a significant subset effect, they proposed a grid searching approach similar to the profile likelihood method to estimate the threshold that defines the sensitive subset. The theoretical aspects of the profile likelihood method were studied by several authors (Jespersen, 1986; Luo and Boyett, 1997; Pons, 2003). However, the finite sample properties of the profile likelihood approach have not been fully investigated so far. In the paper of Jiang et al. (2007), the estimation problem was a secondary interest and the methods for point and interval estimation were not assessed through simulation or evaluated for theoretical validity. It remains an open research question to properly estimate the biomarker threshold that defines a sensitive patient subset and to estimate the treatment effect among the patients in the sensitive subset.

In this paper, we consider a typical clinical trial situation with censored survival outcome and a continuous biomarker and develop a hierarchical Bayes method under the framework of the Cox proportional hazards model (Cox, 1972). Other parametric or semi-parametric models, such as accelerated life time models and proportional odds models can also be used. The hierarchical Bayesian model has been previously proposed to deal with change point problems (Carlin et al., 1992). The hierarchical Bayesian method simultaneously makes a statistical inference on the biomarker threshold defining the sensitive patient subset and the interaction effect between the treatment and the biomarker. The proposed method is evaluated through Monte Carlo simulations. The method is also applied to a data set arising from a clinical trial on prostate cancer.

2. Hierarchical Bayes model

Let T_i and C_i be respectively the potential failure and censoring times for patient i in the study. Let $\delta_i = I(T_i < C_i)$ be a survival status indicator and $X_i = \min(T_i, C_i)$ the observed failure or censoring time, whichever occurs first. Let $z_{1i} = 1$ be the treatment indicator taking value 0 or 1 if patient i is assigned to a control or a new treatment group and z_{2i} a continuous biomarker variable. Given a threshold parameter c for the biomarker variable z_{2i}, the following proportional hazards model for the hazard function $h(t)$ of the survival time T_i can be used to assess the treatment effect on the subset defined by the threshold,

$$h(t|z_{1i}, z_{2i}, c; \boldsymbol{\beta}) = h_0(t) \exp\{\beta_1 z_{1i} + \beta_2 I(z_{2i} > c) + \beta_3 z_{1i} I(z_{2i} > c)\} \tag{1}$$

where $h_0(t)$ is the baseline hazard function. With column vectors $Z_i(c) = [z_{1i}, I(z_{2i} > c), z_{1i} I(z_{2i} > c)]'$ and $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3]'$, we can express the above model as $h_0(t) \exp\{Z_i'(c)\boldsymbol{\beta}\}$.

In the above model, c is an unknown threshold parameter that defines the sensitive subset and we assume that only patients with $z_{2i} > c$ benefit from the new treatment. Without loss of generality, we assume that z_{2i} takes values between 0 and 1 ($0 < z_{2i} < 1$). For a general continuous biomarker variable, we can apply an appropriate transformation so that the z_{2i} are re-scaled to the interval (0, 1). For example, if we apply the inverse function of the empirical cumulative distribution function of $\{z_{2i}\}$, then a threshold parameter, say, $c = 0.6$, implies patients are sensitive to the treatment when their biomarker measurements exceed the 60th percentile in the entire patient population. The regression parameters β_1, β_2 and β_3 model respectively the overall (new) treatment effect, the biomarker main effect and the treatment–biomarker interaction. The interest of the analysis lies mainly in determining the threshold parameter c for defining the sensitive subset and in estimating the size of the treatment effect β_3 restricted on the sensitive subset. The inference of this model cannot be handled as a typical regression problem in the survival analysis framework. This is because the likelihood for the data based on the model involves the unknown threshold parameter c.

In previous theoretical studies and applications (Jespersen, 1986; Luo and Boyett, 1997; Pons, 2003), the profile likelihood method was often applied for the inference of parameters in (1). For each given threshold parameter c, a maximum partial

likelihood estimate $\hat{\boldsymbol{\beta}}_c$ is obtained from model (1), and this leads to a profile likelihood function

$$\ell_p(c) = \ell(\hat{\boldsymbol{\beta}}_c, c), \tag{2}$$

for parameter c. The profile likelihood (2) is then maximized to obtain the estimate $\hat{c}_{\max}$ for c. The maximum likelihood estimate $\hat{\boldsymbol{\beta}}_{\max}$ for β can be obtained from model (1) by plugging in $\hat{c}_{\max}$ for c. Since the profile likelihood function $\ell_p(c)$ is not differentiable, it is difficult to apply the classical asymptotic theories directly for confidence intervals or hypothesis testing for the threshold parameter c. As a consequence, it is often necessary to rely on either the bootstrap method (Pons, 2003) or model simulation (Luo and Boyett, 1997) to make a statistical inference about c.

We will take a Bayesian approach for the statistical inference on both the threshold parameter c and the regression coefficients $\boldsymbol{\beta}$. We first assume that the threshold parameter c has a prior distribution Beta$(2, q)$ for a given $q > 1$. That is, conditioning on the hyper-parameter q, the probability density function (p.d.f.) of the threshold parameter c is given by

$$p_1(c|q) \propto q(q+1)c(1-c)^{q-1}. \tag{3}$$

Since the mode of this p.d.f. is $\frac{1}{q}$, with $q > 1$, this prior is flexible enough to accommodate any prior distribution in the family with its mode taking any specific value in the interval $(0, 1)$.

When specifying the prior distribution for c in the hierarchical Bayes model, instead of taking an arbitrary value for q, we assume a hyper-prior distribution for q with a density function of the form,

$$p_2(q) \propto \frac{(q-1)}{q(q+1)}, \quad \text{for } q > 1. \tag{4}$$

Furthermore, this density function allows us to simplify some distributional calculation involved in the Gibbs sampling described below. It is also possible to choose other prior distributions for q, for example, a noninformative prior $p_2(q) \propto 1$ for $q > 1$. Our experience from the numerical simulations suggests that the posterior distributions are all very similar for different choices of prior distributions.

When there is some prior knowledge about the parameter $\boldsymbol{\beta}$, we discuss in Section 6 that it is possible to assume a multivariate normal proper prior for $\boldsymbol{\beta}$. For simplicity in describing the algorithm, we assume henceforth that $\boldsymbol{\beta}$ has a uniform improper prior distribution $p(\boldsymbol{\beta}) \propto 1$. For any given $0 < c < 1$, the inference for the regression coefficients $\boldsymbol{\beta}$ can be based on the Cox proportional hazards regression model, and the corresponding partial likelihood function for $\boldsymbol{\beta}$ is given by,

$$p_3(\boldsymbol{\beta}|c) = \prod_{i=1}^{n} \left[\frac{\exp\{Z_i'(c)\boldsymbol{\beta}\}}{\sum_{j \in R(x_i)} \exp\{Z_j'(c)\boldsymbol{\beta}\}} \right]^{\delta_i},$$

where the risk set $R(t)$ is the index set of patients that are at risk of experiencing an event at time t. We focus on the partial likelihood approach for $\boldsymbol{\beta}$ because it is well accepted and widely used in biomedical research and studies. An alternative modeling approach is to assume a parametric baseline hazard function, for example, a piecewise constant hazard. Consequently, given the observed data, the joint posterior distribution for $\boldsymbol{\beta}$, c, q is

$$\begin{aligned} p(\boldsymbol{\beta}, c, q|\text{data}) &\propto p_1(c|q)p_2(q)p_3(\boldsymbol{\beta}|c) \\ &= \prod_{i=1}^{n} \left[\frac{\exp\{Z_i'(c)\boldsymbol{\beta}\}}{\sum_{j \in R(x_i)} \exp\{Z_j'(c)\boldsymbol{\beta}\}} \right]^{\delta_i} c(1-c)^{q-1}(q-1). \end{aligned}$$

For any given $0 < c < 1$, the posterior distribution is in fact the product of a partial likelihood function for $\boldsymbol{\beta}$ and a Gamma probability density function for q, with

$$\lim_{c \to 0} p(\boldsymbol{\beta}, c, q|\text{data}) = \lim_{c \to 1} p(\boldsymbol{\beta}, c, q|\text{data}) = 0.$$

Therefore, the posterior $p(\boldsymbol{\beta}, c, q|\text{data})$ is a proper distribution even when the improper priors are assumed for both q and $\boldsymbol{\beta}$. Statistical inference on the regression coefficient $\boldsymbol{\beta}$ and threshold parameter c can be made through the corresponding marginal distributions for $\boldsymbol{\beta}$ and c defined respectively as, $p(\boldsymbol{\beta}) = \int_{c,q} p(\boldsymbol{\beta}, c, q|\text{data})dcdq$ and $p(c) = \int_{\beta,q} p(\boldsymbol{\beta}, c, q|\text{data})d\boldsymbol{\beta}dq$. These marginal posterior distributions involve, however, complex numerical integrations and may be difficult to evaluate in application. In the following, we propose a Markov Chain Monte Carlo (MCMC) approach with Gibbs Sampling to obtain posterior samples from these marginal posterior distributions (Geman and Geman, 1984). Statistical inference such as point estimation, credible interval and hypothesis testing will be made based on these samples.

Specifically, for given initial values $(\boldsymbol{\beta}_0, q_0)$, we can sequentially draw samples c_k, $\boldsymbol{\beta}_k$, q_k for $k = 1, 2, \ldots$ using the following algorithm in three steps. In the simulation study and application example, we take $c_0 = 0.5$, $\boldsymbol{\beta}_0 = (0, 0, 0)'$ and $q_0 = 2$.

Step 1: Given the observed data and the parameters $\boldsymbol{\beta}_k$, q_k from the previous iteration, the conditional distribution $f(c|\cdot)$ is

$$f_1(c|\boldsymbol{\beta}_k, q_k) \propto \prod_{i=1}^{n} \left[\frac{\exp\{Z_i'(c)\boldsymbol{\beta}_k\}}{\sum_{j\in R(x_i)} \exp\{Z_j'(c)\boldsymbol{\beta}_k\}} \right]^{\delta_i} c(1-c)^{q_k-1}. \tag{5}$$

There is not a straightforward method to generate random samples from (5) directly. We use the following Metropolis–Hasting algorithm (Metropolis et al., 1953) to generate c_k from distribution (5). Given the current value c_k, we generate two uniformly distributed random variables u and $u_k \sim unif(0, 1)$. Let

$$\alpha_1 = \frac{f_1(u_k|\boldsymbol{\beta}_k, q_k)}{f_1(c_k|\boldsymbol{\beta}_k, q_k)},$$

and

$$c_{k+1} = \begin{cases} u_k & \text{if } u < \alpha_1 \\ c_k & \text{otherwise.} \end{cases}$$

In this way, the MCMC sample c_{k+1} has p.d.f. $f_1(c|\boldsymbol{\beta}_k, q_k)$. The conditional distribution (5) suggests that with relative large sample size, the posterior distribution of $f_1(c|\boldsymbol{\beta}_k, q_k)$ relies mainly on the partial likelihood function instead of the prior distribution Beta$(2, q_k)$.

Step 2: Given c_{k+1} and q_k, the likelihood function for the regression coefficient $\boldsymbol{\beta}$ has the form of the partial likelihood,

$$f_2(\boldsymbol{\beta}|c_{k+1}, q_k) = \prod_{i=1}^{n} \left[\frac{\exp\{Z_i'(c_{k+1})\boldsymbol{\beta}\}}{\sum_{j\in R(x_i)} \exp\{Z_j'(c_{k+1})\boldsymbol{\beta}\}} \right]^{\delta_i}. \tag{6}$$

Similar to Step 1, the Metropolis–Hasting algorithm can be applied again to generate $\boldsymbol{\beta}_{k+1}$ from (6). The candidate vector $\boldsymbol{\beta}^*$ is sampled from a multivariate normal distribution with mean $\hat{\boldsymbol{\beta}}_{k+1}$ and variance–covariance matrix $\hat{\Sigma}_{k+1}$,

$$\boldsymbol{\beta}^* \sim N(\hat{\boldsymbol{\beta}}_{k+1}, \hat{\Sigma}_{k+1}),$$

where $\hat{\boldsymbol{\beta}}_{k+1}$ is the maximum likelihood estimate of $\boldsymbol{\beta}$ given $c = c_{k+1}$, and $\hat{\Sigma}_{k+1}$ is the corresponding variance–covariance matrix. Then we let $\boldsymbol{\beta}_{k+1}$ take the value $\boldsymbol{\beta}^*$ with probability

$$\alpha_2 = \frac{f_2(\boldsymbol{\beta}^*|c_{k+1}, q_k)}{f_2(\boldsymbol{\beta}_k|c_{k+1}, q_k)}.$$

Step 3: Let $v = q - 1$ and $\lambda = -\log(1 - c_{k+1})$. We can express the conditional distribution of q for given c_{k+1}, $\boldsymbol{\beta}_{k+1}$ in terms of v and λ in the form of $f_3(q|c_{k+1}, \boldsymbol{\beta}_{k+1}) \propto (q-1)(1-c_{k+1})^{(q-1)} \log(\frac{1}{1-c_{k+1}})$, that is

$$\begin{aligned} f_3(v|c_{k+1}, \boldsymbol{\beta}_{k+1}) &\propto v(1-c_{k+1})^v \log\left(\frac{1}{1-c_{k+1}}\right) \\ &= \lambda v \exp\{-\lambda v\}. \end{aligned}$$

With this formulation, the parameter v has a Gamma distribution with a shape parameter 2 and a scale parameter λ^{-1}. In fact, the choice of the prior distribution (4) for the hyper-parameter q leads to this simple yet flexible conditional distribution for $v = q - 1$ in this step. A random sample of v_{k+1} can easily be obtained using existing software such as the *rgamma*() function in the *R* software package (R Development Core Team, 2012). We then take $q_{k+1} = 1 + v_{k+1}$ to be the $(k+1)^{st}$ MCMC sample for the hyper-parameter q.

For any initial value $(\boldsymbol{\beta}_0, c_0, q_0)$, the Gibbs sampling algorithm can be repeated to obtain a sequence of MCMC samples. Let B be the number of burn-in steps so that the Markov Chain attains a stationary distribution after B iterations. We can take $\{\boldsymbol{\beta}_k, c_k, q_k\}_{k=B+1}^{B+R}$, the R MCMC samples after the B steps of burn-in, as the samples from the marginal posterior distribution of parameters $\boldsymbol{\beta}$, c and q.

3. Statistical inference for c and $\boldsymbol{\beta}$

Based on the standard results for Markov Chain Monte Carlo methods (Gilks et al., 1996), the posterior samples $\{\boldsymbol{\beta}_k, c_k, q_k\}_{k=B+1}^{B+R}$ are used to make statistical inferences such as point and credible interval (C.I.) estimation and hypotheses testing for the threshold parameter c, the regression coefficient $\boldsymbol{\beta}$.

3.1. Estimation

Typically, the sample mean from the posterior MCMC samples

$$\hat{c} = \frac{1}{R}\sum_{k=B+1}^{B+R} c_k, \tag{7}$$

is taken as estimate for the expected value of c.

In addition to the point estimation, it is often of interest to build a $100(1-\alpha)\%$ confidence interval for the threshold parameter c. With the Bayesian method, it is straightforward to construct a $100(1-\alpha)\%$ credible interval (c_L, c_U) based on the posterior MCMC samples

$$\begin{cases} c_L = \max\left(u : \dfrac{1}{R}\displaystyle\sum_{k=B+1}^{B+R} I\{c_k \le u\} < \dfrac{\alpha}{2}\right) \\ c_U = \min\left(u : \dfrac{1}{R}\displaystyle\sum_{k=B+1}^{B+R} I\{c_k \le u\} > 1-\dfrac{\alpha}{2}\right). \end{cases} \tag{8}$$

Note that with the profile likelihood approach, the likelihood function for the threshold parameter c has an irregular form and it is much difficult to obtain the confidence interval for c.

We propose two different methods for deriving point estimates and credible intervals for the regression coefficients $\boldsymbol{\beta}$. The first method is called the marginal method, which is similar to the point estimation (7) and credible interval (8) procedures above for the threshold parameter c, based on the marginal posterior samples $\boldsymbol{\beta}_k$.

In biomedical research, investigators often plug in the point estimate $\hat{c}$ into model (1), which leads to the regular Cox proportional hazards model (1) with

$$h(t|z_{1i}, z_{2i}, \boldsymbol{\beta}; \hat{c}) = h_0(t)\exp\left\{\beta_1 z_{1i} + \beta_2 I(z_{2i} > \hat{c}) + \beta_3 z_{1i} I(z_{2i} > \hat{c})\right\},$$

to obtain the maximum likelihood estimate and the corresponding C.I. for $\boldsymbol{\beta}$. This provides the second method of dealing with $\boldsymbol{\beta}$. We refer to this method as the conditional method since the inference about $\boldsymbol{\beta}$ conditions on $c = \hat{c}$. We are interested in studying the finite sample performance of the conditional method because it is popular in application.

3.2. Hypothesis testing

We will investigate two different approaches, the marginal and the conditional methods, for testing the hypotheses about β. The scenario of $c = 0$ or 1 and $\beta_2 = \beta_3 = 0$ is excluded due to model identifiability problems.

One objective of the marginal inference is to access the significance of the regression coefficients β_1, β_2 and β_3 across all possible values of the threshold parameter $0 < c < 1$. In the marginal method, the two-sided p-value for the hypothesis $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$, can be written as,

$$p = 2 * \min\left(\frac{1}{R}\sum_{k=B+1}^{B+R} I\{\beta_{jk} \le 0\}, \frac{1}{R}\sum_{k=B+1}^{B+R} I\{\beta_{jk} \ge 0\}\right),$$

for $j = 1, 2, 3$. Since the marginal test is based on the marginal posterior distribution of $\boldsymbol{\beta}$, it gives an overall assessment of the subset effect that takes into account of the uncertainty of the threshold parameter c.

In the conditional method, given $c = \hat{c}$, p-values for the hypothesis $H_{0j} : \beta_{j|\hat{c}} = 0$ versus $H_{1j} : \beta_{j|\hat{c}} \neq 0$, $j = 1, 2, 3$, can be derived by plugging in the posterior mean $\hat{c}$ to model (1) as in Section 3.1. Asymptotic results for the partial likelihood can be used to construct the conditional test for H_{0j}.

The finite sample properties of the marginal and the conditional approaches will be evaluated through simulation studies in the following section.

4. Simulation studies

We generate time to event data for $n = 300$ subjects using the proportional hazards model (1). Subjects are randomly assigned to either the treatment ($z_{1i} = 1$) or control ($z_{2i} = 0$) group, each with probability 0.5. Covariate z_{2i} follows a uniform (0, 1) distribution. Survival time T_i has an exponential distribution with hazard function $h_i(t) = h_0(t)\exp(Z_i\boldsymbol{\beta})$, where $h_0(t) = 1$ and $Z_i = (z_{i1}, I\{z_{i2} > c\}, z_{i1}I\{z_{i2} > c\})$. Let C_i be random censoring time generated from a uniform distribution $C_i \sim U(2, 5)$ and $\delta_i = I(T_i < C_i)$ be the censoring indicator. The threshold parameter c takes values from 0.2 to 0.8 for a different proportion of the sub-population with treatment effect. We focus on the subset defined effect and take the main treatment effect $\beta_1 = 0$. The main biomarker factor effect β_2 takes values log(1.0) and log(1.5). The regression coefficient β_3 which describes the treatment effect within the sub-population takes values log(1.0), log(1.5), log(2.5) and log(3.5). For each parameter combination, we replicate the simulation 500 times to assess the finite sample properties of

Table 1
Biases for the point estimates of parameters c, β_1, β_2 and β_3: sample size $n = 300$, $\exp(\beta_1) = 1.0$. The marginal method for $\boldsymbol{\beta}$ are based on marginal posterior samples while the conditional method is based on the asymptotic results from the Cox proportional hazards model, conditioning on $c = \hat{c}$. Results are based on 500 replications.

c	e^{β_2}	e^{β_3}	$\hat{c}$	Marginal method			Conditional method		
				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$(MSE.)	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$(MSE.)
$h_i(t) = h_0(t)\exp(Z_i\boldsymbol{\beta})$[a]									
0.2	1.5	1.0	0.1238	0.0021	−0.0064	−0.0055(0.443)	0.0053	−0.0599	−0.0127(0.402)
0.3	1.5	1.0	0.0686	−0.0007	0.0132	0.0214(0.380)	−0.0006	−0.0178	0.0117(0.350)
0.4	1.5	1.0	0.0347	0.0034	0.0214	−0.0111(0.377)	−0.0007	0.0013	−0.0040(0.343)
0.5	1.5	1.0	0.0080	0.0117	0.0228	−0.0051(0.404)	0.0116	0.0103	−0.0096(0.338)
0.6	1.5	1.0	−0.0222	−0.0001	0.0286	−0.0130(0.421)	0.0033	0.0061	−0.0125(0.354)
0.7	1.5	1.0	−0.0674	0.0046	−0.0047	0.0150(0.471)	−0.0011	−0.0324	0.0120(0.339)
0.8	1.5	1.0	−0.1319	−0.0058	−0.0290	−0.0030(0.512)	−0.0025	−0.0690	−0.0097(0.389)
0.2	1.0	1.5	0.1972	0.0376	−0.0178	−0.0075(0.448)	0.1064	−0.0133	−0.0825(0.414)
0.2	1.0	2.5	0.0384	0.0168	0.0054	0.0035(0.524)	0.0440	−0.0045	−0.0318(0.582)
0.2	1.0	3.5	0.0072	0.0037	0.0121	0.0108(0.614)	0.0093	−0.0018	0.0033(0.687)
0.3	1.0	1.5	0.1280	0.0154	0.0055	0.0144(0.455)	0.0487	−0.0049	−0.0292(0.411)
0.3	1.0	2.5	0.0124	−0.0013	0.0009	0.0263(0.483)	0.0045	−0.0109	0.0178(0.531)
0.3	1.0	3.5	0.0034	−0.0036	−0.0023	0.0278(0.597)	−0.0033	−0.0113	0.0258(0.668)
0.4	1.0	1.5	0.0687	0.0198	0.0199	−0.0025(0.462)	0.0288	0.0140	−0.0187(0.395)
0.4	1.0	2.5	0.0027	−0.0142	−0.0056	0.0216(0.461)	−0.0102	−0.0101	0.0161(0.482)
0.4	1.0	3.5	−0.0008	−0.0139	−0.0072	0.0240(0.574)	−0.0126	−0.0105	0.0199(0.621)
0.5	1.0	1.5	0.0115	0.0048	0.0110	0.0122(0.477)	0.0155	0.0153	−0.0121(0.401)
0.5	1.0	2.5	−0.0036	0.0047	0.0081	0.0040(0.477)	0.0066	0.0060	0.0008(0.525)
0.5	1.0	3.5	−0.0006	0.0105	0.0074	0.0039(0.586)	0.0101	0.0065	0.0030(0.668)
0.6	1.0	1.5	−0.0516	−0.0145	−0.0108	0.0412(0.468)	−0.0042	−0.0023	0.0088(0.418)
0.6	1.0	2.5	−0.0073	−0.0054	−0.0099	0.0200(0.473)	−0.0033	−0.0046	0.0114(0.515)
0.6	1.0	3.5	−0.0032	−0.0023	−0.0087	0.0198(0.579)	−0.0018	−0.0054	0.0138(0.653)
$h_i(t) = h_0(t)\exp(Z_i^*\boldsymbol{\beta})$[a]									
			Mean: $\hat{c}$[c]						
–	1.5	1.0	0.5184	−0.0022	0.0017	−0.0082(0.305)	−0.0041	−0.0084	−0.0138(0.278)
–	1.0	1.5	0.5236	0.0027	−0.0007	−0.0034(0.349)	0.0018	0.0077	−0.0223(0.301)
–	1.0	2.5	0.5142	0.0175	0.0097	−0.0572(0.412)	0.0209	0.0082	−0.0693(0.475)
–	1.0	3.5	0.5187	0.0376	0.0059	−0.0966(0.537)	0.0406	0.0057	−0.1092(0.620)

[a] Data are generated and analyzed based on this model with threshold c.
[b] Data are generated from this model, but analyzed based on (the misspecified) model[a] with threshold c.
[c] Posterior sample mean for c when the threshold model is misspecified.

the proposed Bayesian methods. In each simulation, we set the number of MCMC burn-in to be $B = 2000$ and the number of MCMC samples to be $R^* = 10\,000$. To reduce the auto correlation among MCMC samples, we apply a thinning parameter of 2 to the Gibbs samples, using only the $B+1, B+3, \ldots, B+R^*-1$ Gibbs samples, which reduces the number of MCMC samples for the posterior distribution to $R = 5000$.

The scenario of $\beta_2 = \beta_3 = 0$ is excluded due to the identifiability problem (this is the scenario with no subset treatment effect, or equivalently, with $c = 0$ or 1). The scenarios of $\beta_2 = \log(1.5)$ and $\beta_3 = \log(1.5), \log(2.5), \log(3.5)$ are also excluded because they are very similar to those of $\beta_2 = \log(1.0)$ and $\beta_3 = \log(1.5), \log(2.5), \log(3.5)$. Since the definition of the subset is symmetric, results for the setting with cut-point $c = 0.7$ or 0.8 are quite similar to those with $c = 0.3$ or 0.2, respectively; therefore, only selected results for $c = 0.7$ and 0.8 ($\beta_2 = \log(1.0)$ and $\beta_3 = \log(1.5)$) are presented.

To study the robustness of the Bayesian method, we further evaluate the finite sample properties when the threshold model for the covariate z_2 is misspecified. Specifically, we take

$$f(z_2) = \begin{cases} 0 & \text{if } 0 < z_2 \le 0.4 \\ 4(z_2 - 0.4) & \text{if } 0.4 < z_2 \le 0.65 \\ 1 & \text{if } 0.65 < z_2 < 1 \end{cases}$$

and $Z_i^* = [z_{i1}, f(z_{i2}), z_{i1}f(z_{i2})]$, generate data from hazard function $h_i(t) = h_0(t)\exp(Z_i^*\boldsymbol{\beta})$, but still fit the data using model (1).

The upper panel of Table 1 summarizes the biases of the estimates for parameters c and $\boldsymbol{\beta}$ (both marginal and conditional methods) when the model is correctly specified. For the threshold parameter c, the absolute biases are small in most of the settings that we consider. This confirms that $\hat{c}$ is a consistent estimate of c. In particular, for $\beta_2 = \log(1.0)$, the absolute bias for $\hat{c}$ reduces quickly when β_3 increases. For example, when $c = 0.5$, when β_3 changes from $\log(1.5)$ to $\log(3.5)$, the absolute bias for $\hat{c}$ decreases 0.0115–0.0006. For $\beta_2 = \log(1.5)$ and $\beta_3 = 0$, we observe that $\hat{c}$ has the smallest bias when $c = 0.5$ and the bias increases when c deviates from 0.5.

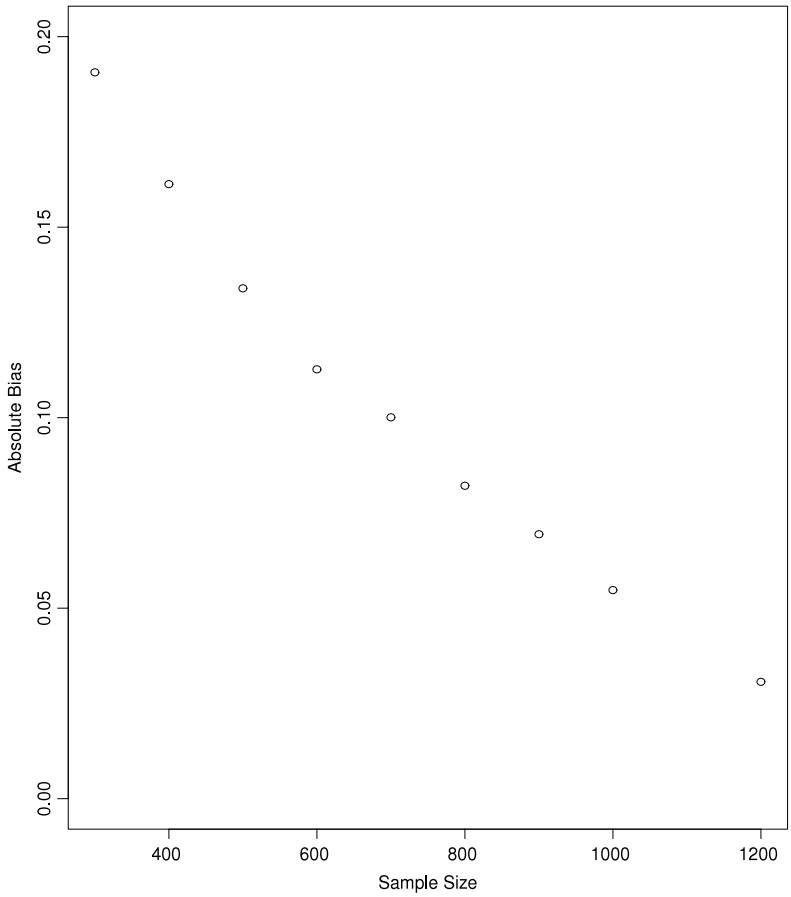

Fig. 1. Empirical absolute bias of the point estimate of the threshold parameter c as a function of sample size n ($c = 0.2$, $\beta_1 = \beta_2 = 0$, $\beta_3 = \log(1.5)$). Results are based on 500 simulation replications.

When $c = 0.2$, $\beta_2 = \log(1.0)$ and $\beta_3 = \log(1.5)$, the bias for $\hat{c}$ is relatively large (bias $= 0.1972$). This is likely due to the relatively small sample size ($n = 300$) and a small subset effect ($\beta_3 = \log(1.5)$). Further simulation shows that for the same parameter specification with different sample size n ranging from 300 to 1200, the bias reduces quickly as the sample size increases (Fig. 1).

When the biomarker z_2 takes effect through the function $f(z_2)$ as specified above and the threshold model is misspecified, instead of reporting the bias for $\hat{c}$ (which does not exist), we report the average posterior mean of $\hat{c}$ from the 500 simulations. These average posterior means for the threshold parameter c are all very close to 0.525, which is the medium point of the function $f(z_2)$ (i.e. $f(0.525) = 0.5$).

In either the marginal method or the conditional method, biases for regression coefficients β_1, β_2 and β_3 are very small, even when the threshold parameter c is not accurately estimated or when the threshold model is misspecified. This indicates that the proposed method provides unbiased estimates for $\boldsymbol{\beta}$, and the method is accurate and robust in the threshold estimation, even with some model misspecifications. Both the marginal method and the conditional method provide comparable mean squared error (MSE) for regression coefficients β_1, β_2 (data not shown) and β_3 (Table 1).

Table 2 shows the empirical coverage probabilities (C.P.) of the 95% credible intervals (C.I.) for threshold parameter c, regression coefficients β_1, β_2 and β_3 and the percentage of rejecting the null hypothesis $\beta_3 = 0$. For the threshold parameter c, the empirical coverage probabilities are close to the nominal level of 95% in all settings of the simulation study, even in the settings where the model has a weak signal to obtain an accurate estimate for the threshold parameter c itself. For example, for $c = 0.2$, $\beta_1 = \log(1.0)$, $\beta_2 = \log(1.0)$ and $\beta_3 = \log(1.5)$, while the posterior mean $\hat{c} = 0.3972$ has a large bias of 0.1972, the 95% C.I. for threshold parameter c has a coverage probability of 93.6%.

The empirical C.P. of the 95% C.I.'s for regression coefficients β using the marginal method are close to the nominal level of 95%. Under the null hypothesis (H_0) situations with $\beta_3 = 0$, test sizes are all close to the nominal level of 5%. Under the alternative hypothesis situations, for the sample size of $n = 300$ with around 10% of censoring, the power of the tests exceeds 80% when $\beta_3 \geq \log 2.5$ in most settings.

For the conditional method, although the regression coefficients $\boldsymbol{\beta}$ can be estimated with a relatively small bias, we notice that for some of the β_j's the empirical coverage probabilities are less than the 95% nominal level. Under the null hypothesis

Table 2
Empirical coverage probabilities for parameters c, β_1, β_2 and β_3 with sample size $n = 300$, $\exp(\beta_1) = 1.0$. The marginal method for $\boldsymbol{\beta}$ is based on marginal posterior samples while the conditional method is based on the asymptotic results from the Cox proportional hazards model, conditioning on $c = \hat{c}$. Results are based on 500 replications.

c	e^{β_2}	e^{β_3}	$\hat{c}$	Marginal method				Conditional method			
				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	REJ%[c]	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	REJ%[c]
$h_i(t) = h_0(t)\exp(Z_i\boldsymbol{\beta})$[a]											
0.2	1.5	1.0	92.0	94.0	93.0	93.6	6.4	91.8	82.8	90.2	9.8
0.3	1.5	1.0	95.2	96.4	95.0	95.4	4.6	93.8	89.2	90.6	9.4
0.4	1.5	1.0	96.4	95.2	95.2	94.8	5.2	94.0	90.2	91.8	8.2
0.5	1.5	1.0	93.8	97.2	94.6	94.0	6.0	95.2	91.2	90.2	9.8
0.6	1.5	1.0	94.4	95.8	96.8	95.4	4.6	94.2	92.4	92.8	7.2
0.7	1.5	1.0	93.0	94.4	95.6	96.2	3.8	92.6	88.2	92.0	8.0
0.8	1.5	1.0	91.4	95.2	94.6	94.2	5.8	93.0	80.0	89.2	10.8
0.2	1.0	1.5	93.6	91.2	95.8	95.6	20.0	78.0	90.2	83.0	28.2
0.2	1.0	2.5	92.2	91.2	94.0	95.8	79.2	83.4	91.2	87.8	81.6
0.2	1.0	3.5	93.6	93.0	94.2	96.0	96.8	89.6	93.6	92.8	97.4
0.3	1.0	1.5	91.6	92.6	93.6	93.0	25.6	82.6	89.4	86.2	33.0
0.3	1.0	2.5	93.2	95.4	94.2	94.0	88.6	91.2	92.0	92.0	91.6
0.3	1.0	3.5	94.0	95.0	93.6	93.2	99.4	93.2	92.6	91.8	99.6
0.4	1.0	1.5	95.0	93.4	94.2	93.8	25.6	90.8	91.0	90.2	37.4
0.4	1.0	2.5	94.4	94.8	94.2	94.4	94.8	92.8	91.6	93.6	95.6
0.4	1.0	3.5	94.8	94.2	93.4	94.8	99.8	94.2	92.8	94.2	99.8
0.5	1.0	1.5	94.2	97.0	94.4	94.0	28.4	92.0	90.2	89.8	38.0
0.5	1.0	2.5	93.4	96.4	94.2	96.0	93.4	94.8	92.6	94.2	95.8
0.5	1.0	3.5	94.2	96.4	94.8	95.6	100	95.4	94.0	95.4	100
0.6	1.0	1.5	94.4	96.0	94.8	95.2	32.6	92.6	90.4	89.4	46.2
0.6	1.0	2.5	94.6	95.2	95.4	96.6	95.2	94.6	94.0	95.2	96.4
0.6	1.0	3.5	96.0	95.2	95.2	95.8	99.6	94.6	94.6	95.6	99.8
$h_i(t) = h_0(t)\exp(Z_i^*\boldsymbol{\beta})$[b]											
			C.I. for c[d]								
–	1.5	1.0	(0.28, 0.75)	95.4	95.0	93.6	6.4	93.2	90.2	90.0	10.0
–	1.0	1.5	(0.21, 0.83)	94.8	94.0	94.6	26.4	91.4	90.0	91.0	39.4
–	1.0	2.5	(0.40, 0.64)	93.2	94.0	95.2	89.4	90.2	93.6	91.4	91.8
–	1.0	3.5	(0.45, 0.59)	92.6	94.4	93.8	99.8	91.4	94.0	91.4	99.8

[a] Data are generated and analyzed based on this model with threshold c.
[b] Data are generated from this model, but analyzed based on (the misspecified) model[a] with threshold c.
[c] Percentage of rejected null hypothesis: $H_0 : \beta_3 = 0$.
[d] Average credible interval for c when the threshold model is misspecified.

of $\beta_3 = 0$, the test sizes are between 7.2% and 10.8% for the settings that we have considered. The empirical C.P. of the 95% C.I. vary for different types of data generated with different model parameter combinations, and are not always close to the 95% nominal level. For example, when $c = 0.2$, $\beta_1 = \beta_2 = 0$ and $\beta_3 = \log 1.5$, the coverage probabilities of 95% C.I. are 78.0%, 90.2% and 83.0% for parameters β_1, β_2 and β_3, respectively.

The profile likelihood method estimates the threshold parameter c by maximizing the profile likelihood function (2). Pons (2003) show that $\frac{1}{n}(\hat{c} - c)$ converges in distribution to a certain probability distribution $F(c)$. However, the explicit distribution function for $F(c)$ is not known and a bootstrap is often necessary to construct the confidence interval for the threshold parameter c. We further conduct numerical simulations to compare the finite sample performance of the proposed MCMC approach and the profile likelihood approach. The 95% C.I. for c is obtained using 500 bootstrap samples for each simulated data. In Table 3, we present results for $c = 0.2$ and 0.5 only, similar results are obtained for $c = 0.3$, 0.4, 0.6.

In terms of the point estimate, the profile likelihood method provides estimates for the threshold parameter c and regression coefficient $\boldsymbol{\beta}$ with small bias when there is a moderate to strong subset effect. The bias is relatively large when there is a small subset treatment effect within a small subgroup of the population. These point estimate performances are all very similar to the proposed MCMC approach.

It is worth noting that the profile likelihood approach using the bootstrap does not provide good coverage probability for the 95% C.I. compared to the proposed Bayesian method. For example, when the subset treatment effect is small ($\beta_3 = \log 1.5$), the empirical C.P. of the 95% C.I. for β_3 is around 80% to 90%, even when half of the study population belong to the sensitive subset ($c = 0.5$). For a strong subset effect of $\beta_3 = \log 3.5$, although the 95% C.I. for β_3 has a proper coverage probability, the empirical C.P. for the 95% C.I. of parameter c goes as low as 74.2% (Table 3).

Additional simulation studies with a uniform prior distribution for threshold parameter c gave similar results (not shown). This suggests that the prior distribution for c may not have a substantial impact on the posterior distribution for both the regression coefficients $\boldsymbol{\beta}$ and the threshold parameter c.

Table 3
Absolute bias and empirical coverage probability (C.P.) for profile likelihood method with a bootstrap 95% credible interval for $\boldsymbol{\beta}$: $n = 300$, $\exp(\beta_1) = 1.0$. Results are based on 500 replications.

c	e^{β_2}	e^{β_3}	Bias				C.P.		
			$\hat{c}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{c}$	$\hat{\beta}_3$[a]	$\hat{\beta}_3$[b]
0.2	1.0	1.5	0.1559	0.0034	0.0400	0.0365	95.8	95.2	79.6
0.2	1.0	2.5	0.0104	−0.0640	−0.0378	0.0927	97.0	92.6	92.0
0.2	1.0	3.5	−0.0052	−0.0530	−0.0285	0.0749	88.2	90.2	94.0
0.2	1.5	1.0	0.0734	0.0054	0.0640	−0.0227	97.8	94.2	78.2
0.2	1.5	1.5	0.0137	0.0014	0.0402	0.0101	94.2	93.2	90.4
0.2	1.5	2.5	−0.0068	−0.0057	0.0079	0.0199	81.4	88.2	92.0
0.2	1.5	3.5	−0.0049	0.0161	0.0210	−0.0023	77.4	89.2	95.6
0.5	1.0	1.5	−0.0043	−0.0218	0.0344	0.0611	98.6	92.0	80.0
0.5	1.0	2.5	−0.0068	−0.0214	0.0025	0.0532	91.6	88.0	92.0
0.5	1.0	3.5	−0.0085	−0.0235	0.0002	0.0380	85.0	89.4	93.8
0.5	1.5	1.0	0.0049	−0.0131	0.0631	0.0089	94.6	93.2	86.2
0.5	1.5	1.5	−0.0069	−0.0069	0.0393	0.0089	86.6	90.8	92.6
0.5	1.5	2.5	−0.0080	−0.0041	0.0185	0.0021	78.2	90.6	95.0
0.5	1.5	3.5	−0.0064	−0.0104	0.0154	0.0210	74.2	89.2	94.8

[a] Based on the asymptotic method.
[b] Based the bootstrap method.

Table 4
Prostate cancer example: the acid phosphatase (AP) biomarker.

Parameter	Estimate	S.E.	95% C. I.		p-value ($\beta_j = 0$)
c	0.803	0.036	0.745	0.871	
Marginal method					
β_1	−0.017	0.136	−0.279	0.258	0.8889
β_2	1.267	0.290	0.692	1.827	<0.0001
β_3	−0.851	0.321	−1.480	0.287	0.007

5. Application

We consider a clinical trial on prostate cancer conducted by the second Veterans Administration Cooperative Urologic Research Group (Byar and Corle, 1977; Andrews and Herzberg, 1985). In this trial, $n = 505$ patients with prostate cancer were randomly assigned to receive either a placebo ($n_0 = 128$ patients) or diethylstilbestrol treatment ($n_1 = 377$ patients). The primary endpoint was overall survival, which is the time from randomization to the time of death from any causes. Participants alive at the end of the study were censored at the last recorded time of being alive. The log-rank test shows no significant difference in survival time distributions between the treatment and placebo groups (p-value = 0.41).

Previous studies (Byar and Corle, 1977; Jiang et al., 2007) suggested that the serum prostatic acid phosphatase (AP) is a predictive biomarker for diethylstilbestrol treatment. In this paper, we will make inferences about the threshold parameter for the AP biomarker using the proposed Bayesian method. We first applied an inverse empirical cumulative distribution function transformation to the AP variable so that it becomes a measurement that is uniformly distributed between 0 and 1.

In the analysis with the proposed method, the threshold parameter c converges quickly in about 500–1000 iterations. After $B = 2000$ burn-in iterations, we further calculate an additional 300 000 Gibbs samples for c, $\boldsymbol{\beta}$ and q. Since the Metropolis–Hasting algorithm is used to sample the threshold parameter c, the MCMC sample may remain unchanged within several iterations (Fig. 2, upper panel). We apply a thinning parameter of 100 to obtain $R = 3000$ samples for the analysis (Fig. 2, lower panel). The first order auto correlation coefficient among the posterior samples of $\{c_k\}$'s reduces from 0.99 to 0.09 after thinning.

The results from the posterior distribution of c and $\boldsymbol{\beta}$ suggest that the acid phosphatase (AP) variable can be used as a biomarker that defines a significant subset treatment effect (Table 4). In the transformed scale of 0 to 1, the threshold parameter is estimated to be $\hat{c} = 0.803$ based on the posterior mean; the corresponding 95% C.I. is (0.684, 0.886). On the original scale of the AP biomarker, the estimated threshold is $\hat{c}* = 46$ and the corresponding 95% C.I. is (27, 107). The estimated cutoff value is similar to the cutoff of 36 obtained by Jiang et al. (2007), while their estimated 95% C.I (9–170) is wider than ours. This suggests that the treatment is effective on a subset of the patients with $AP > 46$ (whose AP values rank among the upper 20%), but not effective on the rest of the patients with AP ≤ 46. According to the marginal method, the treatment main effect is not significant but the main effect of the AP variable is highly significant. When detecting the subset treatment effect, the marginal method gives a smaller p-value of 0.007 compared to the model without the subset effect (p-value = 0.41).

To access the convergence of the MCMC procedure, the algorithm is repeated several times using different random seeds and initial values for c_0, $\boldsymbol{\beta}_0$ and q_0. They all provide similar results for both the threshold parameter c and the regression coefficients $\boldsymbol{\beta}$ (data not shown). This suggests that the proposed Bayesian method is robust to different initial conditions for this example.

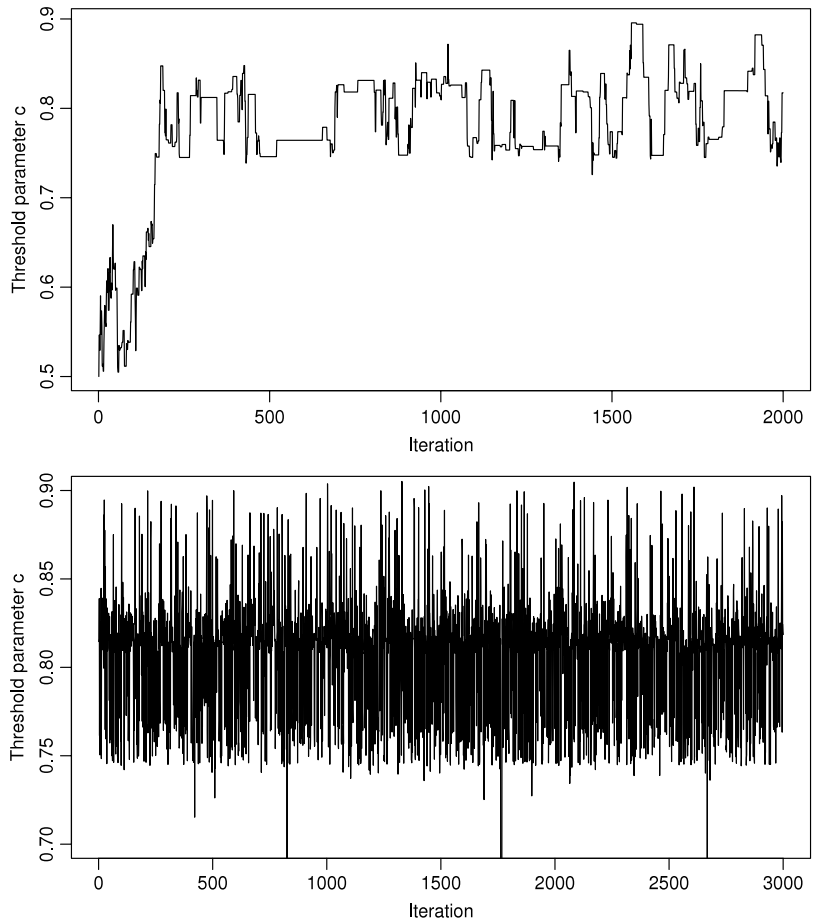

Fig. 2. Prostate cancer data: trace plot for the MCMC samples of threshold parameter c. Upper panel: the first 2000 iterations. Lower panel: the next 3000 iterations with a thinning parameter of 100.

6. Discussion

In this paper, we propose a hierarchical Bayes model for a biomarker defined subset treatment effect for survival time data. We use a Beta distribution Beta$(2, q)$ as the prior for the threshold parameter c and estimate the distribution of the hyper-parameter q through the observed data. We take this approach for the purpose of minimizing the impact of the prior assumption on the posterior distribution for the parameter of interest. When estimating the threshold parameter c, instead of a grid search like those applied in the profile likelihood approach, we apply the Metropolis–Hasting algorithm to sample c from the conditional distribution given $\boldsymbol{\beta}$ and q. This ensures that the threshold parameter c can take any possible values instead of just some pre-specified grid points. Empirical results suggest that the Bayesian method provides excellent finite sample results in terms of bias and coverage probability for the threshold parameter c. The proposed method can be extended to accommodate a continuous interaction effect by replacing $I(z_2 > c)$ with a continuous function $c(z_2)$. It is also possible to take a proper prior distribution for $\boldsymbol{\beta}$, e.g. $\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \Sigma_0)$. The impact of θ_0 and Σ_0 on the posterior distribution of $\boldsymbol{\beta}$ reduces as sample size n becomes larger.

We further provide two different methods for making a statistical inference for the subset treatment effect (that is, the regression coefficient β_3). The first approach, the marginal method, makes an inference about $\boldsymbol{\beta}$ based on the marginal distribution of the posterior distribution of $\hat{\boldsymbol{\beta}}$. By doing this, the method accounts for the uncertainty in the estimation of the threshold parameter c when making a statistical inference about the regression coefficient $\boldsymbol{\beta}$. Numerical simulation indicates that the marginal method provides a robust inference for the regression coefficients $\boldsymbol{\beta}$ in terms of finite sample bias and empirical coverage probability for a 95% credible interval, even when the threshold parameter c is occasionally estimated with relatively large bias.

The second approach, the conditional method, is motivated by the typical strategies used in the profile likelihood literature, which is very popular in biomedical applications. However, numerical simulation shows that the coverage probabilities for the credible intervals are considerable lower than the nominal level in many simulation settings. Overall, the conditional method, though often used in the literature, is not as good as the marginal method.

For comparison, we further conduct numerical simulations for the profile likelihood approach. Numerical results reveal that the corresponding bootstrap method does not have good finite sample properties compared to the hierarchical Bayes model for the C.I estimation. The credible intervals obtained from the Bayesian method tend to have better coverage probabilities than the bootstrap confidence intervals based on the profile likelihood method.

We also investigate the finite sample properties of the proposed methods when the threshold model is misspecified. The simulation study shows that the Bayesian method approximates the true sensitive subset very well. It also provides almost unbiased estimates and credible intervals with good coverage probabilities for the regression coefficients $\boldsymbol{\beta}$.

In this paper, we focus on subset treatment effects in clinical trials with continuous outcomes, which can arise in studies on survival time after treatment, improvement on quality of life measured on a continuous scale, etc. The idea of the hierarchical Bayes model also applies to studies with categorical outcomes such as survival status (patients being alive or dead) at five years after the treatment. The proposed method can be extended to generalized linear models (McCullagh and Nelder, 1989) to handle the categorical outcomes. The proposed methods can also be generalized to accommodate multiple sensitive patient subsets, defined by intervals on biomarker values with a number of threshold parameters. We implement the MCMC algorithm for categorical outcome and survival outcome in the R software (R Development Core Team, 2012), respectively. The source code is available from the first author upon request.

Acknowledgments

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca) and Compute/Calcul Canada. The authors would like to thank two referees and the Associate Editor for their insightful comments and suggestions.

References

Andrews, D., Herzberg, A., 1985. Data. Springer, New York, NY.
Bonetti, M., Gelber, R.D., 2000. A graphical method to assess treatment–covariate interactions using the Cox model on subsets of the data. Statistics in Medicine 19 (19), 2595–2609.
Byar, D., Corle, D., 1977. Selecting optimal treatment in clinical trials using covariate information. Journal of Chronic Diseases 30, 445–459.
Carlin, B.P., Gelfand, A.E., Smith, A.F.M., 1992. Hierarchical Bayesian model for change point problem. Applied Statistics 41, 389–405.
Cox, D.R., 1972. Regression models and life tables (with discussion). Journal of the Royal Statistical Society: Series B 34, 187–220.
Faraggi, D., Simon, R., 1996. A simulation study of cross-validation for selecting an optimal cutpoint in univariate survival analysis. Statistics in Medicine 15 (20), 2203–2213.
Geman, S., Geman, D., 1984. Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.
Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Introducing Markov chain Monte Carlo. In: Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 1–19 (Chapter).
Gray, R.J., 1992. Flexible methods for analyzing survival data using spline, with application to breast cancer prognosis. Journal of the American Statistical Association 87, 942–951.
Jespersen, N.C.B., 1986. Dichotomizing a continuous covariate in the Cox model. Statistical Research Unit, University of Copenhagen 86, 02.
Jiang, W., Freidlin, B., Simon, R., 2007. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. Journal of the National Cancer Institute 99 (13), 1036–1043.
Levine, M.N., Pritchard, K.I., Bramwell, V.H.C., Shepherd, L.E., Tu, D., Paul, N., 2005. Randomized trial comparing cyclophosphamide, epirubicin, and fluorouracil with cyclophosphamide, methotrexate, and fluorouracil in premenopausal women with node-positive breast cancer: update of national cancer institute of Canada clinical trials group trial MA5. Journal of Clinical Oncology 23, 5166–5170.
Luo, X., Boyett, J., 1997. Estimation of a threshold parameter in Cox regression. Communications in Statistics—Theory and Methods 26, 2329–2346.
McCullagh, P., Nelder, J., 1989. Generalized Linear Models, second ed. Chapman and Hall/CRC, New York.
Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., 1953. Equations of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087–1092.
Pons, O., 2003. Estimation in a Cox regression model with a change-point according to a threshold in a covariate. Annals of Statistics 31, 2442–2463.
Pritchard, K.I., Shepherd, L.E., O'Malley, F.P., Andrulis, I.L., Tu, D., Bramwell, V.H., Mark, N., Levine, M.N., 2006. HER2/neu and responsiveness of breast cancer to adjuvant chemotherapy. New England Journal of Medicine 354, 2103–2111.
R Development Core Team, 2012. R: A language and environment for statistical computing.
Royston, P., Altman, D.G., Sauerbrei, W., 2006. Dichotomizing continuous predictors in multiple regression: a bad idea. Statistics in Medicine 25 (1), 127–141.
Royston, P., Sauerbrei, W., 2004. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. Statistics in Medicine 23 (16), 2509–2525.
Wacholder, S., Chen, B.E., Wilcox, A., Macones, G., Gonzalez, P., Befano, B., Hildesheim, A., Rodriguez, A.C., Solomon, D., Herrero, R., Schiffman, M., group, C.V.T., 2010. Risk of miscarriage with bivalent vaccine against human papilloma virus (HPV) types 16 and 18: pooled analysis of two randomised controlled trials. British Medical Journal 340, c712.
Wang, R., Lagakos, S.W., Ware, J.H., Hunter, D.J., Drazen, J.M., 2007. Statistics in medicine–reporting of subgroup analyses in clinical trials. New England Journal of Medicine 357 (21), 2189–2194.
Werft, W., Benner, A., Kopp-Schneider, A., 2012. On the identification of predictive biomarkers: detecting treatment-by-gene interaction in high-dimensional data. Computational Statistics and Data Analysis 56, 1275–1286.