Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Article ID 259474

Research Article

Enhancing the Lasso Approach for Developing a Survival Prediction Model Based on Gene Expression Data

Shuhei Kaneko,[1] Akihiro Hirakawa,[2] and Chikuma Hamada[1]

[1]*Department of Management Science, Graduate School of Engineering, Tokyo University of Science, 1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan*
[2]*Biostatistics and Bioinformatics Section, Center for Advanced Medicine and Clinical Research, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya 466-8560, Japan*

Correspondence should be addressed to Shuhei Kaneko; skaneko.mobile0724@gmail.com

Received 18 September 2014; Accepted 22 December 2014

Academic Editor: Roberto Amato

In the past decade, researchers in oncology have sought to develop survival prediction models using gene expression data. The least absolute shrinkage and selection operator (lasso) has been widely used to select genes that truly correlated with a patient's survival. The lasso selects genes for prediction by shrinking a large number of coefficients of the candidate genes towards zero based on a tuning parameter that is often determined by a cross-validation (CV). However, this method can pass over (or fail to identify) true positive genes (i.e., it identifies false negatives) in certain instances, because the lasso tends to favor the development of a simple prediction model. Here, we attempt to monitor the identification of false negatives by developing a method for estimating the number of true positive (TP) genes for a series of values of a tuning parameter that assumes a mixture distribution for the lasso estimates. Using our developed method, we performed a simulation study to examine its precision in estimating the number of TP genes. Additionally, we applied our method to a real gene expression dataset and found that it was able to identify genes correlated with survival that a CV method was unable to detect.

1. Introduction

In the past decade, researchers have predicted survival in a cancer patient based on gene expression data [1–4]. Revealing the relationship between gene expression profiles and the time to an event of interest (e.g., overall survival, metastasis-free survival) can improve treatment strategies and establish accurate prognostic markers. The Cox proportional hazard model is the most popular method for relating covariates to survival times [5]. However, due to the high dimensionality of gene expression data (i.e., the number of genes expressed exceeds the number of patients), it is not possible to take an estimation approach based on the Cox log partial likelihood. To overcome this problem, a penalized estimation approach, which includes a shrinkage estimation of coefficients, is frequently taken [6–8].

In penalized estimation approaches, the least absolute shrinkage and selection operator (lasso) [9, 10] is often used because of its attractive ability to simultaneously select the genes correlated with survival and estimate the coefficients in the Cox model. The lasso shrinks most of the coefficients towards zero exactly by adding L_1 norm to the Cox log partial likelihood, and the amount of shrinkage is dependent on the tuning parameter. The value of the tuning parameter is often determined by a cross-validation (CV), which maximizes the out-of-data prediction accuracy [11].

Several researchers have investigated the operating characteristics of the lasso. Goeman [12] used the lasso to analyze a publicly available gene expression dataset, obtained from the articles of van't Veer et al. [2] and van de Vijver et al. [3] in which a 70-gene signature for prediction of metastasis-free survival in breast cancer patients had been established. This data included 295 patients with 4919 genes that were prescreened from 24,885 genes based on the quality criteria in van't Veer et al.'s work [2]. The lasso selected 16 genes with which to develop a prediction model of overall survival when using the tuning parameter that was determined using a CV. Goeman [12] also conducted ridge regression using all 4919

genes to develop a model by adding L_2 norm to the Cox log partial likelihood. The prediction accuracy of the lasso and ridge regression were compared, and the ridge regression with 4919 genes slightly outperformed the lasso with 16 genes. Goeman [12] concluded that the lasso potentially passes over genes that are correlated with survival in order to develop a simple prediction model. Bøvelstad et al. [7] reached the same conclusion in a review of the survival prediction methods available for analyzing breast cancer gene expression datasets. Table 1 summarizes a typical result of gene selection by the lasso.

The CV method determines the value of the tuning parameter by considering the trade-off between the number of true positives (TP) and false positives (FP), and so the possibility of identifying false negatives (FN) cannot be eliminated. One solution for identifying more outcome-predictive genes is to monitor the number of TP in several values of the tuning parameter and, subsequently, determine its final value. In this study, we developed a method for estimating the number of TP for a series of values of the tuning parameter. We assumed a mixture distribution with components of TP and FP for the lasso estimates, and these could be used to estimate the number of TP and FP. It is possible to generate the solution path that includes the lasso estimates for a series of values of the tuning parameter using the methods developed by Goeman [12]. Here, we proposed an algorithm to sequentially fit the mixture distribution for this solution path, and we used a simulation study to test the precision of the algorithm when estimating the number of TP. We further demonstrated the proposed algorithm using a well-known diffuse large B-cell lymphoma (DLBCL) dataset comprising overall survival of 240 DLBCL patients and gene expression data of 7399 genes [1].

2. Materials and Methods

2.1. Lasso in the Cox Proportional Hazard Model. The Cox proportional hazard model is the most popular method for evaluating the relationship between gene expression and time to an event of interest [5]. The hazard function of an event at time t for a patient i ($i = 1, \ldots, n$) with the gene expression levels $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ is given by

$$h(t \mid \mathbf{x}_i) = h_0(t) \exp\left(\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}\right), \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is a parameter vector and $h_0(t)$ is the baseline hazard, which is the hazard for the respective individual when all variable values are equal to zero. In the general setting where $n > p$, the coefficients are estimated by maximizing Cox log partial likelihood as follows:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left[\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} - \log \left\{ \sum_{r \in R(t_i)} \exp\left(\mathbf{x}_r^{\mathrm{T}} \boldsymbol{\beta}\right) \right\} \right], \tag{2}$$

where δ_i is an indicator, which is 1, if the survival time is observed, or 0, if censored. $R(t_i)$ is the risk set of the individuals at t_i.

Table 1: Typical results of gene selection by the lasso.

True condition	The lasso	
	Select	No select
Genes that are not correlated with survival (none-outcome-predictive genes)	False positive (FP)	True negative (TN)
Genes that are truly correlated with survival (outcome-predictive genes)	True positive (TP)	False negative (FN)

In the lasso for the high-dimensional setting where $n < p$, the coefficients are estimated by maximizing the following penalized likelihood function [9, 10]:

$$l_p(\boldsymbol{\beta}, \lambda) = l(\boldsymbol{\beta}) - \lambda \sum_{j=1}^{p} \left|\beta_j\right|, \tag{3}$$

where λ is the tuning parameter, which determines the amount of shrinkage.

2.2. Solution Path of the Lasso Estimates. Goeman [12] introduced a method to calculate the solution path of the lasso estimates as a function of λ, $\widehat{\boldsymbol{\beta}}(\lambda)$, which is based on the algorithm developed by Park and Hastie [13]. The method maximizes $l_p(\boldsymbol{\beta}, \lambda)$ at a fixed λ based on a combination of gradient ascent optimization with the Newton-Raphson algorithm. $\widehat{\boldsymbol{\beta}}(\lambda)$ are calculated for $\lambda_0 > \cdots > \lambda_k > \cdots > \lambda_z > 0$ successively, starting from $\lambda_0 = \max_j \partial l / \partial \beta_j|_{\beta_j = 0}$ (which gives $\widehat{\boldsymbol{\beta}}(\lambda_0) = \mathbf{0}$ because the value has zero gradients). λ_z is chosen arbitrarily but is often set to $0.05 \times \lambda_0$ in analyses of gene expression data [14]. The lasso estimates at a current step are set to initial values for calculation of the subsequent step. Step length $\Delta_k = \lambda_k - \lambda_{k+1}$ is the minimum decrement to change the number of selected genes $m^{(k)}$ ($= \#\{j; \widehat{\beta}_j(\lambda_k) \neq 0\}$); that is, only one gene is newly selected or excluded from λ_k to λ_{k+1}.

2.3. Mixture Distribution for Estimating the Number of TP in the Lasso Estimates. To estimate the number of TP in the lasso estimates at a fixed value of λ, we assumed a mixture distribution developed in our previous study [15]. We introduced the mixture distribution based on the two features of the lasso: (i) the lasso selects at most n genes because of the nature of the convex optimization problem when $n < p$ [16, 17] and (ii) in the Bayesian paradigm the lasso estimates are the posterior mode with the independent Laplace prior distribution $f_L(\beta_j; 0, 1/\tau) = (\tau/2)\exp(-\tau|\beta_j|)$, where $f_L(y; a, b) = 1/2b \exp(-|y - a|/b)$ is the probability density function of Laplace distribution with location parameter a

and scale parameter b [9]. Therefore, the mixture distribution assumed for the lasso estimates at λ was as follows:

$$f\left(\widehat{\beta}_j(\lambda);\pi_0,\pi_c,\tau,\mu_c,\sigma_c\right) = \frac{n}{p}\left\{\pi_0 f_L\left(\widehat{\beta}_j(\lambda);0,\frac{1}{\tau}\right)+\sum_{c=1}^{C}\pi_c f_N\left(\widehat{\beta}_j(\lambda);\mu_c,\sigma_c^2\right)\right\} + \left(1-\frac{n}{p}\right) f_L\left(\widehat{\beta}_j(\lambda);0,\epsilon\right), \tag{4}$$

where π_0 and π_c are mixed proportions ($\pi_0 + \sum_{c=1}^{C}\pi_c = 1$); $f_N(\widehat{\beta}_j(\lambda);\mu_c,\sigma_c^2)$ is the probability density function of the normal distribution with mean μ_c ($\neq 0$) and variance σ_c^2 in component c; C is the number of components, which is determined by model selection criteria; and ϵ is the constant value, which is boundlessly close to 0; for example, $\epsilon = 10^{-8}$. The unknown parameters, π_0, π_c, τ, μ_c, and σ_c, are estimated by maximizing the log-likelihood function of (4) by using the Newton-Raphson method.

The mixture distribution defined in (4) is formulated on the basis of the following concepts: since the lasso selects a maximum of n genes when $p > n$, the coefficients for $p-n$ genes are exactly zero; therefore, (4) consists of 2 terms (n/p term and $1-n/p$ term). In the n/p term, the Laplace distribution with location parameter 0 and scale parameter $1/\tau$ was assumed to be the distribution for the FP on the basis of the lasso feature (ii) discussed above, while the C component normal distribution with location parameter μ_c and scale parameter σ_c^2 was assumed as the distribution for the TP. In the $1-n/p$ term, the Laplace distribution with location parameter 0 and scale parameter ϵ was assumed as the distribution of $p-n$ genes based on the aforementioned lasso feature (i).

The f_L with location parameter 0 and scale parameter $1/\tau$ was assumed to be the distribution for the FP on the basis of lasso feature (i), discussed above. The f_N with location parameter μ_c and scale parameter σ_c^2 was assumed as the distribution for the TP. The f_L of the $(1-n/p)$ term was assumed as the distribution of $p-n$ genes based on the aforementioned lasso feature (ii). Given a cut-off value ζ (>0), the estimated proportions of the FP and TP are the area under the estimated Laplace and normal distribution in the n/p term of (4), respectively, and can be written as follows:

$$\widehat{P}_{\mathrm{FP}} = \widehat{\pi}_0\left[\int_{-\infty}^{-\zeta} f_L\left(u;0,\widehat{\tau}^{-1}\right)du + \int_{\zeta}^{+\infty} f_L\left(u;0,\widehat{\tau}^{-1}\right)du\right],$$

$$\widehat{P}_{\mathrm{TP}} = \sum_{c=1}^{C}\widehat{\pi}_c\left[\int_{-\infty}^{-\zeta} f_N\left(u;\widehat{\mu}_c,\widehat{\sigma}_c^2\right)du + \int_{\zeta}^{+\infty} f_N\left(u;\widehat{\mu}_c,\widehat{\sigma}_c^2\right)du\right]. \tag{5}$$

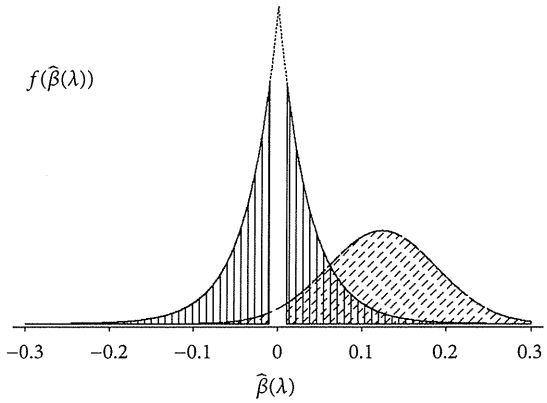

FIGURE 1: Illustration for estimating the number of FP and TP. The areas denoted by the vertical and diagonal lines are the proportion of FP and TP, respectively.

Figure 1 illustrates the calculation in (5) when the number of components, C, is 1. Using (5), the number of TP and FP was estimated by

$$\widehat{\mathrm{FP}} = \frac{\widehat{P}_{\mathrm{FP}}}{\widehat{P}_{\mathrm{TP}}+\widehat{P}_{\mathrm{FP}}}\times m, \tag{6}$$

$$\widehat{\mathrm{TP}} = \frac{\widehat{P}_{\mathrm{TP}}}{\widehat{P}_{\mathrm{TP}}+\widehat{P}_{\mathrm{FP}}}\times m. \tag{7}$$

2.4. Algorithm for Estimating Number of TP in a Series of Values λ. Here, we propose an algorithm to sequentially fit the mixture distribution in (4) to the solution path of the lasso estimates, which was described in Section 2.2. In this algorithm, we assumed that the number of TP changed when the newly selected or excluded gene from λ_k to λ_{k+1} was truly correlated to survival, based on the maximum log-likelihood of (4). First, we approximated $\widehat{P}_{\mathrm{FP}} \approx \widehat{\pi}_0$ and $\widehat{P}_{\mathrm{TP}} \approx \sum_{c=1}^{C}\widehat{\pi}_c$ in (5) by assuming a suitably small cut-off value ζ (≈ 0). We then obtained $\widehat{\pi}_0 = \widehat{\mathrm{FP}}/m$ and $\widehat{\pi}_c = \widehat{\mathrm{TP}}_c/m$ ($c = 1,\ldots,C$) from (6) and (7), respectively, where $\widehat{\mathrm{TP}}_c$ is an estimate of the number of TP in component c. For $k = 1,\ldots,z$, the proposed algorithm was as follows.

Step 1

Step 1.1. In this step, we assumed that the newly selected or excluded gene from λ_k to λ_{k+1} was FP. π_0 denotes the proportion of FP and is set as

$$\pi_0^{(k+1)} = \begin{cases} \dfrac{\widehat{\mathrm{FP}}^{(k)}+1}{m^{(k+1)}}, & \text{if } m^{(k+1)} = m^{(k)}+1, \\ \dfrac{\widehat{\mathrm{FP}}^{(k)}-1}{m^{(k+1)}}, & \text{if } m^{(k+1)} = m^{(k)}-1. \end{cases} \tag{8}$$

For the other components, c ($c = 1,\ldots,C$), set $\pi_c^{(k+1)} = \widehat{\mathrm{TP}}_c^{(k)}/m^{(k+1)}$.

Step 1.2. Given $\widehat{\boldsymbol{\beta}}(\lambda_{k+1})$ and $\pi_0^{(k+1)}, \ldots, \pi_C^{(k+1)}$, calculate the maximum log-likelihood of (4), $L_0^{(k+1)}$.

Step 2

Step 2.1. Set $c = 1$.

Step 2.2. In this step, we assumed that the newly selected or excluded gene from λ_k to λ_{k+1} was TP. For the component c, set

$$\pi_c^{(k+1)} = \begin{cases} \dfrac{\widehat{\mathrm{TP}}_c^{(k)} + 1}{m^{(k+1)}}, & \text{if } m^{(k+1)} = m^{(k)} + 1, \\ \dfrac{\widehat{\mathrm{TP}}_c^{(k)} - 1}{m^{(k+1)}}, & \text{if } m^{(k+1)} = m^{(k)} - 1. \end{cases} \tag{9}$$

For the other components, set $\pi_0^{(k+1)} = \widehat{\mathrm{FP}}^{(k)}/m^{(k+1)}$ and $\pi_d^{(k+1)} = \widehat{\mathrm{TP}}_d^{(k)}/m^{(k+1)}$ $(d = 1, \ldots, C;\ d \neq c)$.

Step 2.3. Given $\widehat{\boldsymbol{\beta}}(\lambda_{k+1})$ and $\pi_0^{(k+1)}, \ldots, \pi_C^{(k+1)}$, calculate the maximum log-likelihood of (4), $L_c^{(k+1)}$.

Step 2.4. Set $c = c + 1$. Repeat Steps 2.2 and 2.3 until $c = C$.

Step 3. In this step, we determined whether the newly selected or excluded gene from λ_k to λ_{k+1} was TP or FP based on the maximum log-likelihood which was calculated in Steps 1.2 and 2.3. If $L_0^{(k+1)}$ was the largest in $L_c^{(k+1)}$ $(c = 0, \ldots, C)$, we assumed that the newly selected or excluded gene was FP; if not, we assumed that it was TP. Therefore, calculate $C_{\max} = \mathrm{argmax}_{c \in \{0,1,\ldots,C\}} L_c^{(k+1)}$. If $C_{\max} = 0$, update $\widehat{\mathrm{FP}}^{(k)}$ as follows:

$$\widehat{\mathrm{FP}}^{(k+1)} = \begin{cases} \widehat{\mathrm{FP}}^{(k)} + 1, & \text{if } m^{(k+1)} = m^{(k)} + 1, \\ \widehat{\mathrm{FP}}^{(k)} - 1, & \text{if } m^{(k+1)} = m^{(k)} - 1. \end{cases} \tag{10}$$

If $C_{\max} > 0$, update $\widehat{\mathrm{TP}}_{C_{\max}}^{(k)}$ as follows:

$$\widehat{\mathrm{TP}}_{C_{\max}}^{(k+1)} = \begin{cases} \widehat{\mathrm{TP}}_{C_{\max}}^{(k)} + 1, & \text{if } m^{(k+1)} = m^{(k)} + 1, \\ \widehat{\mathrm{TP}}_{C_{\max}}^{(k)} - 1, & \text{if } m^{(k+1)} = m^{(k)} - 1. \end{cases} \tag{11}$$

Here, calculate the estimated TP at $k + 1$ by $\widehat{\mathrm{TP}}^{(k+1)} = \sum_{c=1}^{C} \widehat{\mathrm{TP}}_c^{(k+1)}$.

3. Results

3.1. Simulation Study. We performed a simulation study to examine the precision of our estimated TP. In this study, the number of patients, n, was set to 200. The number of genes, p, was set to 1000, which included the p_1 (=5 or 30) outcome-predictive genes that are randomly chosen from p genes in each simulation. The coefficient for gene j $(j = 1, \ldots, p)$, β_j, was set to 1.5 for the p_1 outcome-predictive genes and 0 for the remaining $p - p_1$ none-outcome-predictive genes. We set λ_z to 5 and the number of components, C, to 1 throughout (although C was determined using a model selection criterion in practice). The gene expression levels for patient i, $\mathbf{x}_i$, were generated from the multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ so that the variance was 1 and the correlation $\rho(x_{ik}, x_{il}) = 0$ or $0.5^{|k-l|}$ [18]. The survival time for patient i was generated based on the exponential model $t_i = -\log(U)/\exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$ where U is the uniform random variable between 0 and 1 [19]. In order to evaluate the precision of the estimated TP for various values of λ, we report a number of selected genes, including true TP, and estimated TP and FP, for λ_k $(k = 5, 10, 50, 100, 150)$.

Table 2 shows the average of λ, a number of selected genes, true TP, and estimated TP and FP, through 1000 repeats. We observed that the precision of estimated TP varied depending on the value of both p_1 and k (see Table 2). When $p_1 = 5$, the precision of the estimates was sufficient for $k = 10, 50, 100$, and 150, while TP was slightly underestimated for $k = 5$. However, when $p_1 = 30$, the precision of the estimates was sufficient for $k = 5$, 10, and 150, while TP was overestimated for $k = 50$ and 100. For example, when $p_1 = 30$, $\rho = 0.5$, and $k = 100$, the average number of true and estimated TP was 29.9 and 35.3, respectively. The values of ρ did not greatly affect the accuracy of the estimated TP.

3.2. Real Data Analysis. To illustrate how our proposed algorithm could be used to determine λ, we applied it to the DLBCL dataset, comprising survival of 240 DLBCL patients and gene expression data from 7399 genes [1]. In the gene expression data from the 240 patients, we identified 434 genes with complete sets of gene expression values; all other genes had missing expression values, with an average of 24.7 missing values per gene. Here, we used 0.0 as the missing expression value for descriptive purposes. Similar to Rosenwald et al. [1], we divided the data into two: training data consisting of 160 patients and validation data consisting of 80 patients.

For the training data, we obtained the solution path of the lasso estimates; $\widehat{\boldsymbol{\beta}}(\lambda_k)$ $(k = 0, 1, \ldots, z)$. $\lambda_0 = 72.5$ was calculated as described in Section 2.2. We set $\lambda_z = 3.625$ $(=0.05 \times \lambda_0)$ according to Simon et al. [14].

We applied our proposed algorithm to the obtained solution path. We assumed three mixture distributions on the lasso estimates with $C = 1$, 2, or 3 and compared their goodness of fit for the $\widehat{\boldsymbol{\beta}}(\lambda_k)$ $(k = 0, 1, \ldots, z)$ by the Akaike information criterion (AIC). As a result, we chose $C = 1$ because it had the best AIC for all λ_k $(k = 0, 1, \ldots, z)$.

Figure 2 shows the estimated number of TP in a series of values of λ. We found that the lasso selected at most 42 TP, with the number of selected genes at 96, when $\lambda = 7.19$ (=0.86 as $\log_{10}$). Therefore, we selected $\lambda = 7.19$ as the optimum λ, and the estimated mixture distribution for the value of λ was as follows:

$$\begin{aligned} f\left(\widehat{\beta}_j(7.19)\right) = \frac{160}{7399}\Big\{0.57 \times f_L\left(\widehat{\beta}_j(7.19); 0, 0.11\right) + 0.43 \\ \times f_N\left(\widehat{\beta}_j(7.19); 0.03, 0.11^2\right)\Big\} \\ + \frac{7239}{7399} f_L\left(\widehat{\beta}_j(7.19); 0, 10^{-8}\right). \end{aligned} \tag{12}$$

Table 2: Accuracy of the estimated number of true positives (TP) obtained using the proposed algorithm in the simulation study. Average of a tuning parameter (λ), number of selected genes ($\#\{j; \widehat{\beta}_j(\lambda) \neq 0\}$) in the lasso, true number of true positives (True TP), estimated number of TP ($\widehat{\text{TP}}$), and false positives ($\widehat{\text{FP}}$) are reported at λ_k ($k = 5, 10, 50, 100, 150$) of the solution path.

p_1	ρ	k	λ	$\#\{j; \widehat{\beta}_j(\lambda) \neq 0\}$	True TP	$\widehat{\text{TP}}$	$\widehat{\text{FP}}$
30	0	5	47.0	5.0	4.4	2.9	2.2
		10	40.8	10.1	8.0	5.8	4.3
		50	22.9	48.6	25.6	28.5	20.1
		100	12.6	86.7	29.9	32.1	54.7
		150	8.6	124.5	30.0	30.7	93.9
	0.5	5	48.6	5.0	4.1	2.8	2.2
		10	42.1	10.0	7.5	5.8	4.2
		50	23.5	48.1	25.2	31.9	16.3
		100	12.4	84.9	29.9	35.3	49.6
		150	8.4	121.2	30.0	31.6	89.6
5	0	5	66.9	5.0	5.0	3.0	2.0
		10	26.3	10.4	5.0	5.2	5.2
		50	17.2	50.1	5.0	5.2	44.9
		100	12.7	93.9	5.0	5.0	88.9
		150	9.8	128.4	5.0	5.0	123.4
	0.5	5	66.8	5.0	5.0	3.0	2.0
		10	26.5	10.3	5.0	5.2	5.1
		50	16.9	49.5	5.0	5.1	44.4
		100	12.4	92.1	5.0	5.0	87.1
		150	9.6	125.2	5.0	5.0	120.2

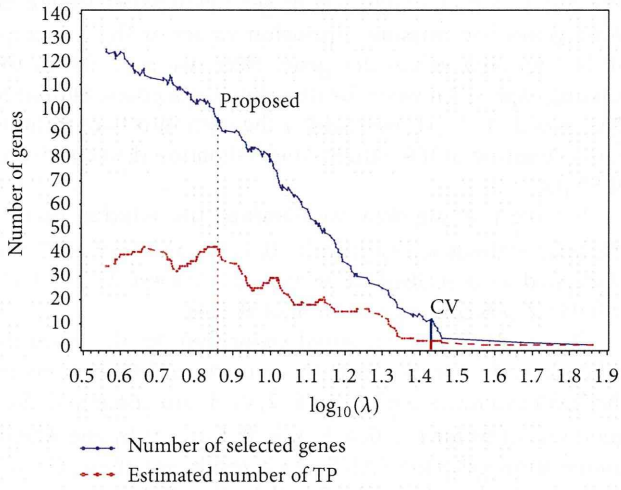

Figure 2: Trace plot of number of selected genes and estimated number of true positives (TP) produced by applying the proposed algorithm to the training data from the diffuse large B-cell lymphoma (DLBCL) dataset. We determined $\lambda = 7.19$ ($\log_{10} = 0.86$) as the optimum λ based on the estimated number of TP. Using cross-validation (CV), we determined $\lambda = 27$ ($\log_{10} = 1.43$) as the optimum λ.

In order to identify the 42 TP from the 96 selected genes, we arranged the 96 in descending order of $|\widehat{\beta}_j|$ and identified the first 42 listed genes with a cut-off value $\zeta = 0.084$. Subsequently, the model that included these 42 genes is identified as the "42 TP-model."

Table 3: GenBank accession numbers and descriptions for 4 genes selected by both CV and the model including the 42 genes identified by the algorithm that we developed.

GenBank accession number	Description
X82240 (AA729003)	T-cell leukemia/lymphoma 1A
AA805575	Thyroxine-binding globulin precursor
LC_29222	—
X59812(H98765)	Cytochrome P450, subfamily XXVIIA polypeptide

In comparison to the 42 TP-model, we performed CV. Briefly, the K-fold CV was given by

$$\text{CV}(\lambda) = \sum_{k=1}^{K} \left\{ l\left(\widehat{\boldsymbol{\beta}}_{(-k)}(\lambda)\right) - l_{(-k)}\left(\widehat{\boldsymbol{\beta}}_{(-k)}(\lambda)\right) \right\}, \quad (13)$$

where $l_{(-k)}(\boldsymbol{\beta})$ and $\widehat{\boldsymbol{\beta}}_{(-k)}$ are the log partial likelihood and the lasso estimate with left kth fold out, respectively. The optimal value of λ was obtained by maximizing $\text{CV}(\lambda)$. On the basis of 5-fold CV, 12 genes were selected with $\lambda = 27$ (=1.43 as $\log_{10}$). Subsequently, the model including these 12 genes is identified as the "CV-model." Notably, both the 42 TP-model with 42 genes and the CV-model with 12 genes selected 4 genes in common. Table 3 shows the GenBank accession number and description for each of the 4 genes selected by both models.

We compared the prediction accuracy of the 42 TP-model and the CV-model using validation data consisting

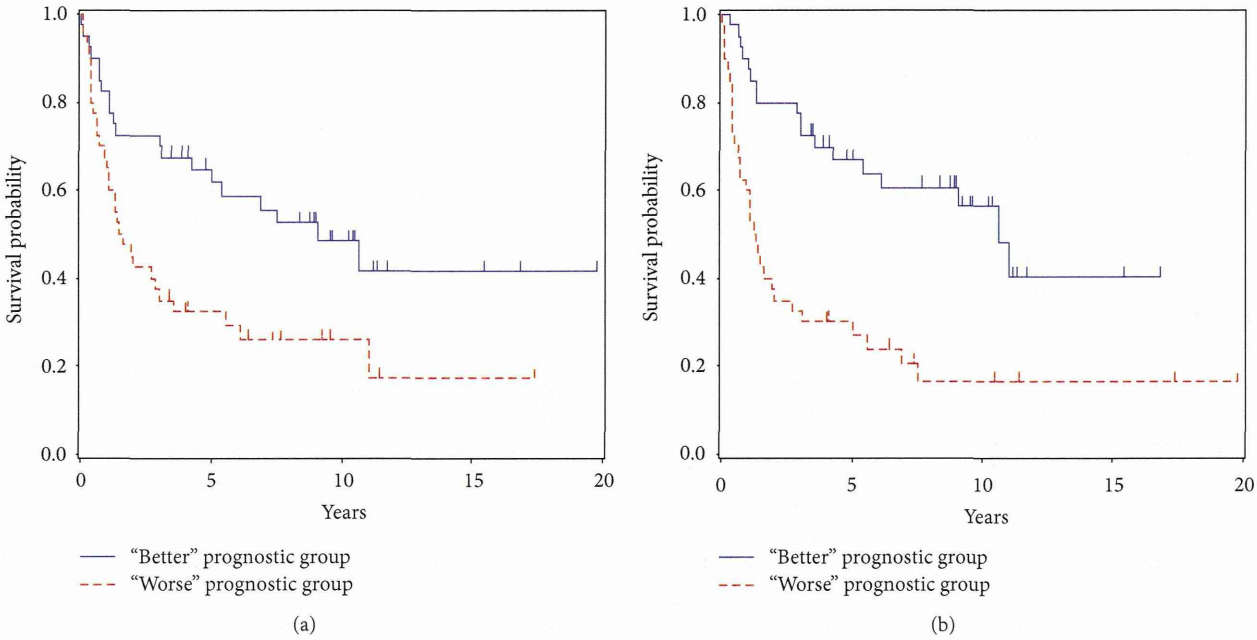

FIGURE 3: Kaplan-Meier curves of overall survival for "better" and "worse" prognostic groups: (a) the model including 12 genes determined by CV (CV-model) and(b) the model including 42 genes identified by the developed method (42 TP-model).

TABLE 4: Values of the comparison criteria for the model including 12 genes determined by CV (CV-model) and the model including the 42 genes identified by our developed algorithm (42 TP-model).

Criteria	CV-model	42 TP-model
P value of the log-rank test	0.007	<0.001
P value for the prognostic index	0.002	<0.001
Deviance	−9.079	−11.297

of 80 patients. For this data, we calculated 3 values that served as comparison criteria: P values for the log-rank test and prognostic index and the deviance. The 80 patients were categorized into 2 groups, the "better" and "worse" prognostic groups, using the boundary of the median of prognostic index $\widehat{\eta}_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}$. The Kaplan-Meier curves between the 2 groups were compared with a log-rank test. Next, we calculated the P value for the parameter α multiplied by the prognostic index $\widehat{\eta}_i$ in the Cox proportional hazard model $h(t_i \mid \mathbf{x}) = h_0(t) \exp(\alpha \widehat{\eta}_i)$. Finally, the deviance was calculated by $-2\{l^{(\text{validation})}(\widehat{\boldsymbol{\beta}}_{\text{training}}) - l^{(\text{validation})}(\mathbf{0})\}$, where $l^{(\text{validation})}(\widehat{\boldsymbol{\beta}}_{\text{training}})$ and $l^{(\text{validation})}(\mathbf{0})$ are the Cox log partial-likelihood function for the estimated coefficients by using the training data and zero vector $\mathbf{0}$, respectively. For each criterion, the lower value suggested better prediction accuracy.

Table 4 shows the values of the 3 criteria for each model. We found that the values of all 3 criteria for the 42 TP-model were lower than those for the CV-model, suggesting that the model based on the proposed method was more accurate (see Table 4). Additionally, Figure 3 shows that the Kaplan-Meier curves for the 42 TP-model distinguished the "better" and "worse" prognostic groups more definitely than those for the CV-model (42 TP-model, $P < 0.001$; CV-model, $P = 0.007$). Therefore, by using our proposed algorithm, we determined λ and were able to select important genes, likely to be correlated with survival, in which the CV was unable to select.

4. Discussions

In this study, we proposed an algorithm for estimating the number of TP on the solution path of lasso estimates. Monitoring and determining the number of TP for a series of values λ are important because they can increase the probability of uncovering all outcome-predictive genes. The number of TP should be estimated with appropriate accuracy. To confirm the accuracy of our TP, we conducted a simulation study using a typical gene expression dataset. We found that the precision of our algorithm for estimating the number of TP was adequate, although an overestimation occurred with some values of λ. However, the overestimation occurred when the true number of TP was saturated, and so it may not cause a problem by passing over genes that truly correlated with survival. In the simulation study where $p_1 = 30$ and $\rho = 0.5$, the maximum average estimated number of TP was 35.3 at $\lambda = 12.4$ (see Table 2). Using this λ to select TP, an average selection of 29.9 TP within 30 outcome-predictive genes can be made, with the number of TP genes that are passed over being negligible in practice.

The data that have been provided in Table 2 showed that the number of false positives increased, while the number of true positives increased and then plateaued as the tuning parameter decreased. To decrease the number of FP identified while maintaining an adequate number of TP, we should determine the value of λ by monitoring both the number of

TP and the false positive rate (=FP/(TP+FP)) in the proposed method.

Additionally, our proposed algorithm was applied to DLBCL data. We determined the value of the tuning parameter based on the maximum number of estimated TP uncovered by the algorithm. We identified 42 TP genes among 96 selected genes based on the ranking of the absolute values of the lasso estimates. We can also identify TP based on model evaluation criteria such as AIC among all possible combinations of 42 genes from 96, that is, ${}_{96}C_{42}(>10^{27})$ combinations in total; however, calculation of AIC for all possible gene combinations is a distant approach. To evaluate the efficiency of the approach using the ranking of the lasso estimates, we calculated the AIC for 10,000 randomly chosen models among all the possible models and subsequently compared it with the AIC of our approach. From 10,000 models, the AIC of 425 models (4.25%) was better than that of our approach. This result indicated that our ranking-based approach has a satisfactory performance in practice with respect to the identification of 42 genes. Although investigation of all possible gene combinations is ideal, our approach is a good alternative.

In the application to DLBCL data, in comparison to a CV method by which 12 genes were identified, we identified 42 TP genes with our algorithm, and we improved the prediction accuracy of the model. In practice, some researchers might be satisfied with identifying a few promising genes and would not be unduly worried about passing over others. In such a situation, the CV would be preferable because it developed the model to uncover a few genes with just a small loss of prediction accuracy. However, genes that are selected by the lasso are often investigated with greater scrutiny by genetic researchers, and so passing over outcome-predictive genes by the lasso could represent a major problem. Indeed, if the lasso passes over outcome-predictive genes, some genetic research may not take place. Therefore, when identifying all outcome-predictive genes is a priority, our proposed algorithm will be most useful.

5. Conclusions

We developed a method for estimating the number of true positives for a series of values of a tuning parameter in the lasso. We demonstrated the utility of the developed method through a simulation study and an application to a real dataset. Our results indicated that our developed method was useful for determining a value for the tuning parameter in the lasso and reducing the probability of passing over genes that are truly correlated with survival.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] M. Rosenwald, G. Wright, W. C. Chan et al., "The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma," *The New England Journal of Medicine*, vol. 346, no. 25, pp. 1937–1947, 2002.

[2] L. J. van't Veer, H. Dai, M. J. van de Vijver et al., "Gene expression profiling predicts clinical outcome of breast cancer," *Nature*, vol. 415, no. 6871, pp. 530–536, 2002.

[3] M. J. van de Vijver, Y. D. He, L. J. van't Veer et al., "A gene-expression signature as a predictor of survival in breast cancer," *The New England Journal of Medicine*, vol. 347, no. 25, pp. 1999–2009, 2002.

[4] Y. Wang, J. G. M. Klijn, Y. Zhang et al., "Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer," *The Lancet*, vol. 365, no. 9460, pp. 671–679, 2005.

[5] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society B: Methodological*, vol. 34, no. 2, pp. 187–220, 1972.

[6] J. Gui and H. Li, "Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data," *Bioinformatics*, vol. 21, no. 13, pp. 3001–3008, 2005.

[7] H. M. Bøvelstad, S. Nygård, H. L. Størvold et al., "Predicting survival from microarray data—a comparative study," *Bioinformatics*, vol. 23, no. 16, pp. 2080–2087, 2007.

[8] W. N. van Wieringen, D. Kun, R. Hampel, and A.-L. Boulesteix, "Survival prediction using gene expression data: a review and comparison," *Computational Statistics & Data Analysis*, vol. 53, no. 5, pp. 1590–1603, 2009.

[9] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 58, no. 1, pp. 267–288, 1996.

[10] R. Tibshirani, "The lasso method for variable selection in the cox model," *Statistics in Medicine*, vol. 16, no. 4, pp. 385–395, 1997.

[11] P. J. M. Verweij and H. C. van Houwelingen, "Cross-validation in survival analysis," *Statistics in Medicine*, vol. 12, no. 24, pp. 2305–2314, 1993.

[12] J. J. Goeman, "L_1 penalized estimation in the Cox proportional hazards model," *Biometrical Journal*, vol. 52, no. 1, pp. 70–84, 2010.

[13] M. Y. Park and T. Hastie, "L_1-regularization path algorithm for generalized linear models," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 69, no. 4, pp. 659–677, 2007.

[14] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for Cox's proportional hazards model via coordinate descent," *Journal of Statistical Software*, vol. 39, no. 5, pp. 1–13, 2011.

[15] S. Kaneko, A. Hirakawa, and C. Hamada, "Gene selection using a high-dimensional regression model with microarrays in cancer prognostic studies," *Cancer Informatics*, vol. 11, pp. 29–39, 2012.

[16] B. Efron, T. Hastie, I. Johnstone et al., "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[17] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.

[18] R. J. Tibshirani, "Univariate shrinkage in the Cox model for high dimensional data," *Statistical Applications in Genetics and Molecular Biology*, vol. 8, no. 1, pp. 3498–3528, 2009.

[19] R. Bender, T. Augustin, and M. Blettner, "Generating survival times to simulate Cox proportional hazards models," *Statistics in Medicine*, vol. 24, no. 11, pp. 1713–1723, 2005.

Journal of Biopharmaceutical Statistics, 24: 310–325, 2013

ISSN: 1054-3406 print/1520-5711 online
DOI: 10.1080/10543406.2013.863779

Taylor & Francis
Taylor & Francis Group

BAYESIAN MODEL AVERAGING CONTINUAL REASSESSMENT METHOD FOR BIVARIATE BINARY EFFICACY AND TOXICITY OUTCOMES IN PHASE I ONCOLOGY TRIALS

Takashi Asakawa[1,2], Akihiro Hirakawa[3], and Chikuma Hamada[2]
[1]*Chugai Pharmaceutical Co., Ltd., Tokyo, Japan*
[2]*Faculty of Engineering, Tokyo University of Science, Tokyo, Japan*
[3]*Center for Advanced Medicine and Clinical Research, Nagoya University Graduate School of Medicine, Aichi, Japan*

Many dose-finding approaches that could evaluate bivariate binary efficacy and toxicity outcomes have been proposed in recent years. In such designs, the operating characteristics with finite sample size can be greatly affected by the assumed dose–toxicity andlor dose–efficacy relationship. However, we do not have much information about a new agent we investigated at the planning stage of Phase I trials and so always face to the risk of misspecifying the true dose–toxicity andlor dose–efficacy relationship by arbitrarily and subjectively choosing skeletons. In this article, we proposed the Bayesian model averaging bivariate continual reassessment method to cope with above risk.

Key Words: Bayesian model averaging; Continual reassessment method; Dose-finding; Oncology; Phase I.

1. INTRODUCTION

The primary goal of a Phase I oncology clinical trial is to determine the maximum tolerated dose (MTD) that should be used during further development of a new agent. The MTD is defined as the dose of a new agent with toxicity probability closest to the investigators' target. As an alternative to the traditional algorithm-based design such as the 3 + 3 design, the continual reassessment method (CRM) was developed by O'Quigley et al. (1990). Given the investigators' prespecified toxicity probability at each dose, CRM updates the estimates of these dose toxicity probabilities based on the Bayes theorem. The operating characteristics of the CRM with respect to the estimation of MTD are more favorable than those of the algorithm-based designs in several works (Chevret, 1993; Faries, 1993; O'Quigley and Chevret, 1991).

The assumption of a monotonically increasing dose–toxicity curve is almost always appropriate from a biological perspective; however, a monotonically

Received April 20, 2012; Accepted October 21, 2012
Address correspondence to Takashi Asakawa, 2-1-1 Nihonbashi-Muromachi, Chuo-ku, Tokyo 103-8324, Japan; E-mail: asakawtks@chugai-pharm.co.jp

increasing relationship between dose and efficacy has been challenged by the recent development of targeted agents (e.g., molecular targeted agents, therapeutic vaccines, and immunotherapy). To meet such a requirement, many dose-finding approaches that could evaluate bivariate binary efficacy and toxicity outcomes have been proposed. As described in Mandrekar et al. (2010), most of them are extended versions of the CRM and are broadly categorized into two types. In the first type of approach, the joint distribution of binary efficacy and toxicity can be collapsed into an ordinal trinary variable, such as no toxicity and no efficacy, no toxicity but with efficacy, and toxicity (Ivanova, 2003; O'Quigley et al., 2001; Thall and Russell, 1998; Whitehead et al., 2006; Zhang et al., 2006). In the second approach, the bivariate structure of outcomes can be maintained in a joint probability distribution (Braun, 2002; Thall and Cook, 2004). In this article, we entirely focus on the latter approach.

In dose-finding approaches based on toxicity and efficacy, operating characteristics with finite sample size can be greatly affected by the assumed dose–toxicity and/or dose–efficacy relationship (which we refer to in this work as skeletons). In practice, we do not have much information about a new agent we investigated at the planning stage of a Phase I oncology trial. Moreover, it is more challenging in a situation in which the joint evaluation of efficacy and toxicity is required. Thus, we always face the risk of misspecifying the true dose–toxicity and/or dose–efficacy relationship by arbitrarily and subjectively choosing skeletons.

One idea to accommodate the uncertainty in the skeleton specification is to introduce Bayesian model averaging (BMA), which estimates the posterior probability for toxicity and efficacy by averaging posterior probabilities (Raftery et al., 1997). Yin and Yuan (2009) proposed the Bayesian model averaging CRM (BMA-CRM) that solves the arbitrariness and subjectivity of the prespecified toxicity probabilities in the CRM. Similar to the method of Yin and Yuan, we propose the BMA bivariate CRM (BMA-bCRM) approach for dose finding, which uses bivariate toxicity and efficacy outcomes in single-agent Phase I trials. Specifically, we prespecify some sets of skeletons for both dose–efficacy and dose–toxicity relationships. For each skeleton, we subsequently estimate the posterior probabilities of toxicity and efficacy at each dose level. We obtain BMA estimates for the toxicity and efficacy probabilities by averaging the posterior probabilities of two outcomes for each skeleton. The dose escalation/deescalation decision rules are defined based on these estimates. Thus, instead of using a single skeleton for each outcome, we use multiple combinations of skeletons for both dose–efficacy and dose–toxicity relationships in parallel and rely on the BMA estimator for decision making. The proposed BMA approach could incorporate multiple assumptions in terms of a skeleton for efficacy and toxicity. In addition, the proposed BMA approach allows us to tune the weights of each model adaptively on the basis of the observed data. Therefore, the proposed BMA approach would provide estimates of efficacy and toxicity probability that are close to the estimates of the best-fitting set of skeletons for the observed data among the assumed sets of skeletons.

To examine the utility of the BMA-bCRM approach, we compared the operating characteristics of the BMA-bCRM approach with those of the ordinal bivariate continual reassessment method (bCRM) approach, which assumes specific skeletons for toxicity and efficacy, through simulation studies under various scenarios.

We organize the remainder of this study as follows. In section 2, we introduce the proposed BMA-bCRM approach for dose-finding using bivariate toxicity and efficacy outcomes. In sections 3 and 4, we compare the operating characteristics of the proposed BMA-bCRM and bCRM approaches through simulation studies and examine the operating characteristics of the proposed BMA-bCRM approach under various scenarios. Finally, in sections 5, we discuss the characteristics of the proposed approach in further detail.

2. METHOD

2.1. Bayesian Model Averaging Bivariate Continual Reassessment Method

In general, limited information for a new agent is available at the planning stage. The investigators might also have different opinions of the skeletons. As a consequence, our choice of skeletons might be largely arbitrary, subjective, and of a wide variety. Furthermore, there is more uncertainty than with CRM based on toxicity when we jointly evaluate the efficacy and toxicity. Thus, the major issue is that we always face the nonnegligible possibility of misspecification of the true dose efficacy and/or dose toxicity. To address these difficulties, we adopt the idea of BMA-CRM, proposed by Yin and Yuan (2009), which could mitigate the uncertainty of the assumed skeletons for efficacy and toxicity. We prespecify multiple combinations of skeletons for efficacy and toxicity. Each combination represents one of the assumptions for efficacy and toxicity probabilities. For each combination (which we term a working model), we implement the bCRM. To obtain the BMA estimate of efficacy and toxicity probabilities for dose level j, we estimate the weighted average of them for the prespecified working models. A weight of each working model is adaptively estimated according to the fitness to the observed outcomes. The BMA provides the robust property in terms of misspecifying the true dose efficacy and/or dose toxicity. The BMA is also known to provide a better prediction than does any single specific model (Hoeting et al., 1999; Raftery et al., 1997).

Let $WM_k(\pi_{ekj}(\beta_{ek}), \pi_{tkj}(\beta_{tk}))$, $(k = 1, 2, \cdots, K)$ be the kth working model, which consists of combination of arbitrary skeletons, where π_{ekj} and π_{tkj} are the estimators of efficacy and toxicity probabilities at jth dose, respectively, and β_{ek} and β_{tk} are unknown model parameters for kth working model, respectively. We assume the power model for dose–efficacy and dose–toxicity relationships in this study; thus, the working model WM_k is given by

$$\pi_{ekj}(\beta_{ek}) = p_{ekj}^{\exp(\beta_{ek})} \tag{1}$$

$$\pi_{tkj}(\beta_{tk}) = p_{tkj}^{\exp(\beta_{tk})}, \tag{2}$$

which consist of the kth skeletons for efficacy probability p_{ekj} and toxicity probability p_{tkj}. Suppose that n_j patients have been treated and y_{ej} and y_{tj} patients have experienced efficacy and toxicity, respectively. Under the observed data $D = \{(n_j, y_{ej}, y_{tj}), j = 1, \cdots, J\}$ and z_j defined as the number of patients whose response is $y_{ej} = 1$ and $y_{tj} = 1$ at dose j, the likelihood function for WM_k is given by

$$L(D|\beta_{ek}, \beta_{tk}, \psi_k, WM_k) \propto \prod_{j=1}^{J} \pi_{ekj}^{y_{ej}}(1-\pi_{ekj})^{(n_j-y_{ej})}\pi_{tkj}^{y_{tj}}(1-\pi_{tkj})^{(n_j-y_{tj})}\psi_k^{z_j}(1-\psi_k)^{n_j-z_j}, \tag{3}$$

which is identical to the bivariate structure of the underlying probability model that Braun (2002) proposed. Then we can estimate the posterior means of π_{ekj} and π_{tkj} at dose level j by the kth working model using Bayes's theorem along with bCRM. Specifically,

$$\hat{\pi}_{ekj} = \int_{\beta_{ek}} \pi_{ekj}(\beta_{ek}) \int_{\beta_{tk}} \int_{\psi_k} \frac{L(D|\beta_{ek},\beta_{tk},\psi_k,WM_k)f(\beta_{ek})f(\beta_{tk})f(\psi_k)}{\int_{\beta_{ek}}\int_{\beta_{tk}}\int_{\psi_k} L(D|\beta_{ek},\beta_{tk},\psi_k,WM_k)f(\beta_{ek})f(\beta_{tk})f(\psi_k)d\beta_{ek}d\beta_{tk}d\psi_k} d\beta_{tk}d\psi_k d\beta_{ek} \tag{4}$$

$$\hat{\pi}_{tkj} = \int_{\beta_{tk}} \pi_{tkj}(\beta_{tk}) \int_{\beta_{ek}} \int_{\psi_k} \frac{L(D|\beta_{ek},\beta_{tk},\psi_k,WM_k)f(\beta_{ek})f(\beta_{tk})f(\psi_k)}{\int_{\beta_{ek}}\int_{\beta_{tk}}\int_{\psi_k} L(D|\beta_{ek},\beta_{tk},\psi_k,WM_k)f(\beta_{ek})f(\beta_{tk})f(\psi_k)d\beta_{ek}d\beta_{tk}d\psi_k} d\beta_{ek}d\psi_k d\beta_{tk}, \tag{5}$$

where $f(\beta_{ek})$, $f(\beta_{tk})$, and $f(\psi_k)$, are the independent prior distributions for each parameter under kth working model, while they are abbreviated the condition WM_k for simplicity. We assume the normal prior distribution $N(0, 4^2)$ for gradient parameters β_{ek} and β_{tk}, and we assume the beta distribution $Be(2, 2)$ for association parameter ψ_k to have a prior mean value of 0.5 with sufficiently vague information.

Let $\Pr(WM_k)$ be the prior probability that represents the prior relative certainty (or importance) for kth working model with the restriction $\sum_k \Pr(WM_k) = 1$. In this article, we assume that each working model has equal prior probability. As are the posterior means of efficacy and toxicity probabilities, these probabilities for each working model are also adaptively updated as posterior probabilities. The posterior probability is given by

$$\begin{aligned}
\Pr(WM_k|D) &= \frac{L(D|WM_k)\Pr(WM_k)}{\sum_{i=1}^{K} L(D|WM_i)\Pr(WM_i)} \\
&= \frac{\eta_{k0}B_{k0}}{\sum_{i=1}^{K}\eta_{i0}B_{i0}}, \qquad (6)\\
\eta_{k0} &= \Pr(WM_k)/\Pr(WM_0), \qquad (7)\\
B_{k0} &= \Pr(D|WM_k)/\Pr(D|WM_0) \\
&= L(D|WM_k)/L(D|WM_0), \qquad (8)
\end{aligned}$$

where $L(D|WM_k)$ is a marginal likelihood of kth working model, η_{k0} is a prior odds for WM_k against the reference working model WM_0, and B_{k0} is a Bayes factor for WM_k against WM_0. Using $\Pr(WM_k|D)$ as a weight for the kth working model, the BMA estimates for efficacy and toxicity probabilities at the jth dose level are obtained simply by a weighted average of $\hat{\pi}_{ekj}$ and $\hat{\pi}_{tkj}$ across the K working models:

$$\bar{\pi}_{ej} = \sum_{k=1}^{K} \hat{\pi}_{ekj}\Pr(WM_k|D) \tag{9}$$

$$\bar{\pi}_{tj} = \sum_{k=1}^{K} \hat{\pi}_{tkj}\Pr(WM_k|D). \tag{10}$$

As mentioned before, by weighting the posterior means $\hat{\pi}_{ekj}$ and $\hat{\pi}_{tkj}$ with $\Pr(WM_k|D)$, which reflects the relative degree of fitness to the observed outcomes, the proposed BMA-bCRM approach could adaptively maximize the influence of well-fitting working models and simultaneously minimize the influence of poorly fitting working models. Consequently, the proposed BMA-bCRM approach could manage the misspecification of the true dose-toxicity and/or dose-efficacy better than the ordinal bCRM approach, which assumes only a set of skeleton combinations for efficacy and toxicity.

The parameter estimation is easy to compute based on the Markov-chain Monte Carlo (MCMC) method. We estimated the posterior distribution of model parameters using a random-walk Metropolis algorithm to generate the sample for generating recursive draws from a particular Markov chain, the stationary distribution of which is the same as the posterior joint distribution of parameters using PROC MCMC in SAS, version 9.2 (SAS Institute Inc., Cary, NC). In this article, we set a burn-in period of 5,000 iterations with a chain length of 50,000, retaining every fifth sample; therefore, the inference about the posterior distribution is based on the 10,000 effective samples.

2.2. Dose-Finding Algorithm

Patients are allocated a specific dose level in cohort, which consists of three patients. In the proposed approach, the skipping dose level in the escalation or de-escalation is not allowed. We define the minimum requirement criteria for the recommended dose (RD) to ensure at least minimum efficacy and maximum allowable toxicity with high probability. To obtain the probability that satisfies the minimum efficacy or maximum allowable toxicity, we apply the BMA approach using $\Pr(WM_k|D)$. The criteria are given by

$$\sum_{k=1}^{K} \Pr(\hat{\pi}_{ekj} \geq c_e) \Pr(WM_k|D) \geq 0.9 \tag{11}$$

$$\sum_{k=1}^{K} \Pr(\hat{\pi}_{tkj} \leq c_t) \Pr(WM_k|D) \geq 0.9, \tag{12}$$

where c_e and c_t are set to 0.2 and 0.3, respectively. The trial is terminated when no dose levels satisfy these criteria. To avoid an inappropriate termination due to unstable estimates of the probabilities just defined, the criteria are activated after the outcomes in some cohort of patients become available. Among dose levels satisfying the criteria, the dose level assigned to the next cohort of patients is that which minimizes the weighted Euclidean distance from the target (ϕ_e, ϕ_t), using BMA estimates of efficacy and toxicity probabilities, such that

$$ED_j = \sqrt{w_e(\phi_e - \bar{\pi}_{ej})^2 + (1 - w_e)(\phi_t - \bar{\pi}_{tj})^2}. \tag{13}$$

We assume $\phi_e = 1$ and $\phi_t = 0$ as the target values and $w_e = 0.5$. When the planned maximum number of patients is reached, then the RD is determined by the BMA estimates of efficacy and toxicity probability based on all accumulated outcomes.

3. SIMULATION STUDIES

3.1. Simulation Setting

We investigate the operating characteristics of the proposed BMA-bCRM approach through simulation studies in eight true scenarios. We assume five dose levels and four sets of working models for efficacy and toxicity probabilities. Table 1 shows the working models we assumed. The first working model is for the case in which efficacy increases proportionally in increments of 10% from 20% at the initial dose level and toxicity increases slowly at the low dose levels but increases quickly from fourth dose level. The second working model is for the case in which efficacy increases proportionally in increments of 5% from 25% and toxicity increases quickly with relatively high toxicity at the initial dose level. The third working model is for the case in which efficacy increases slowly at the low dose levels but increases quickly at the fourth dose level and toxicity increases slowly with no intolerable dose levels. The fourth working model is for the case in which efficacy increases monotonically at the low dose levels but becomes constant from the third dose level and toxicity increases proportionally in increments of 10% from 10%.

We expect that the proposed BMA-bCRM approach can effectively capture these features and appropriately weight each working model based on the observed outcomes. We refer to the ordinal bCRM using each working model as WM_1, WM_2, WM_3, and WM_4, and compare the operating characteristics of them with those of the proposed BMA-bCRM approach. Because the ordinal bCRM could be considered a special case of BMA-bCRM, that is $K = 1$, we could implement it by the strategy introduced in the previous chapter.

Table 2 shows the true efficacy and toxicity probabilities under each scenario in the first row and the dose selection probabilities for RD at the end of trial and the average number of patients treated at each dose in the subsequent rows. These results are displayed according to the employed design, specifically, WM_1, WM_2, WM_3, WM_4, and BMA-bCRM. We also show the probability of no appropriate dose level at the end of trial, denoted as "None," the average percentage of patients experienced efficacy and toxicity, and the average total number of patients. For the proposed BMA-bCRM approach, we assume no preference for any specific working model and assign equal prior probability $\Pr(WM_k) = 0.25$ for $k = 1, \cdots, 4$.

We derived the correlated Bernoulli efficacy and toxicity outcomes using uniformly distributed correlated outcome between 0 and 1, which is derived by

Table 1 Assumed efficacy and toxicity probabilities (p_{ej}, p_{tj}) (%)

	Dose level				
Working model	1	2	3	4	5
WM_1	(20, 5)	(30, 10)	(40, 15)	(50, 30)	(60, 45)
WM_2	(25, 15)	(30, 25)	(35, 45)	(40, 55)	(45, 65)
WM_3	(15, 05)	(20, 10)	(25, 15)	(40, 20)	(45, 25)
WM_4	(20, 10)	(30, 20)	(50, 30)	(50, 40)	(50, 50)

Note. (p_{ej}, p_{tj}) means efficacy and toxicity probabilities in percentage, respectively, for jth dose level in each working model (WM).

Table 2 Results of simulation studies comparing the proposed BMA-bCRM approach to the bCRMs with efficacy and toxicity target $(\Phi_e, \Phi_t) = (100\%, 0\%)$

		Selection probabilities (%) for RD at the end of trial								
		Dose level								
Scenario (p_{ej}, p_{tj})	Design	1	2	3	4	5	None	Average percentage of efficacy	Average percentage of toxicity	Average number of patients
1		(20, 5)	(30, 10)	(40, 15)	(50, 30)	(60, 45)				
	WM_1	0	0	15.9	**66.4**	16	1.7	20.3	11.7	44.4
	Number of patients	3.5	3.2	10.5	20.1	16				
	WM_2	0	5.9	63.3	**25.6**	2.7	2.5	17.3	8	44.2
	Number of patients	4.2	8.7	20.6	13.5	9.4				
	WM_3	0	0	5.6	**50**	42.2	2.2	21.8	14	44.3
	Number of patients	3.6	3.3	4.3	14.7	20.7				
	WM_4	0	0	98.3	**0**	0	1.7	16.5	6.2	44.4
	Number of patients	3.5	3.2	38.1	0	0				
	BMA-bCRM	0	0	16.3	**69.9**	11.4	2.4	19.8	11.2	44.2
	Number of patients	3.8	3.3	10.3	21.4	13.2				
2		(25, 15)	(30, 25)	(35, 45)	(40, 55)	(45, 65)				
	WM_1	2.7	**32.2**	58.5	1.2	0.2	5.2	14.3	15.3	43.6
	Number of patients	5.5	10.8	23.8	10.6	6				
	WM_2	5.7	**80.3**	10.1	0.3	0	3.6	12.9	11.1	44
	Number of patients	11.3	26.3	13.8	5.5	5				
	WM_3	3.5	**39.8**	42.8	6	0.5	7.4	14.5	15.9	43.2
	Number of patients	6.3	11.9	14.5	9.5	8.1				
	WM_4	2.7	**38.4**	51.7	0	0	7.2	13.9	14	43.4
	Number of patients	5.7	13	26	0	0				
	BMA-bCRM	2	**41.2**	50.6	2.1	0.3	3.8	14.2	14.9	43.9
	Number of patients	6.5	12.9	20.9	9.5	5.5				
3		(15, 5)	(20, 10)	(25, 15)	(40, 20)	(45, 25)				
	WM_1	0	0	3.5	21.8	**69.6**	5.1	15.9	8.6	43.3
	Number of patients	3.6	3.3	5.8	10.8	27.4				
	WM_2	0	2.5	29.2	22.7	**38.7**	6.9	13.1	7.1	42.8
	Number of patients	4.3	7.2	13.9	10.2	23.6				
	WM_3	0	0	0	2.6	**91.8**	5.6	16.6	9.1	43.2
	Number of patients	3.7	3.2	3.2	4.6	30.7				
	WM_4	0	0	89.9	0	**0**	10.1	9.9	5.8	42
	Number of patients	3.5	3.3	36.4	0	0				
	BMA-bCRM	0	0	1.6	21	**71.2**	6.2	15.7	8.5	43
	Number of patients	4	3.3	5	11.8	25.4				
4		(20, 10)	(30, 20)	(50, 30)	(50, 40)	(50, 50)				
	WM_1	0.1	5.2	**67.2**	23.3	1.8	2.4	19.8	13.4	44.2
	Number of patients	4.1	5.3	23.5	17	12.7				
	WM_2	0.4	63.6	**30.4**	1.7	0.3	3.6	15.2	9.8	43.8
	Number of patients	6.8	23.2	17.3	8.5	8.3				
	WM_3	0.1	8.3	**35**	37.9	15.4	3.3	19.4	15.1	43.9
	Number of patients	4.6	5.7	9.9	14.9	15.8				
	WM_4	0	9.5	**88.3**	0	0	2.2	19.5	11.8	44.3
	Number of patients	4.2	6.5	34.5	0	0				
	BMA-bCRM	0.1	7.7	**65.3**	22	1.5	3.4	19.2	12.9	43.9
	Number of patients	4.8	6.1	22.2	15.7	10.2				

(Continued)

Table 2 continued

Scenario (p_{ej}, p_{ij})	Design	Selection probabilities (%) for RD at the end of trial: Dose level 1	2	3	4	5	None	Average percentage of efficacy	Average percentage of toxicity	Average number of patients
5		(20, 20)	(30, 30)	(40, 50)	(50, 55)	(60, 60)				
	WM_1	11.4	**47.7**	23.8	0.3	0	16.8	13.1	15	40.3
	Number of patients	8	15.3	17.1	8.3	4.4				
	WM_2	14.4	**72.3**	1.4	0	0	11.9	11.1	11.5	41.5
	Number of patients	15.3	25.9	8.3	4.3	4				
	WM_3	11.7	**50.1**	21.6	1.1	0.2	15.3	13.8	15.3	40.6
	Number of patients	8.8	14.9	11.4	7.6	7				
	WM_4	11.5	**57.1**	14	0	0	17.4	12.7	14.3	40.7
	Number of patients	8.5	18.6	16.3	0	0				
	BMA-bCRM	10.2	**59.6**	18	0.5	0.1	11.6	13.1	14.7	41.5
	Number of patients	8.8	18.4	15.1	8.2	5.3				
6		(5, 5)	(10, 10)	(20, 15)	(30, 25)	(35, 45)				
	WM_1	0	0	3.9	**52.7**	23.1	20.3	9.9	10.2	38.6
	Number of patients	4.3	3.4	5.4	17	19.3				
	WM_2	0	1.1	26	**28.4**	13.8	30.7	7.5	7.2	35.6
	Number of patients	4.9	5.9	13.4	13	14.5				
	WM_3	0	0	2.2	**27.8**	46.1	23.9	10.1	11.6	37.4
	Number of patients	4.5	3.2	3.8	9.2	25				
	WM_4	0	0	64.7	**0**	0	35.3	6.1	4.7	35.3
	Number of patients	4	3.3	32	0	0				
	BMA-bCRM	0	0	2.4	**45.3**	25.8	26.5	9.2	9.5	36.4
	Number of patients	4.7	3.3	5.3	16.8	18.1				
7		(20, 70)	(30, 75)	(40, 85)	(50, 90)	(55, 95)				
	WM_1	0	0	0	0	0	**100**	3.5	9.3	12.2
	Number of patients	5.6	3.3	3.9	3	0				
	WM_2	0	0	0	0	0	**100**	2.7	8.7	12.2
	Number of patients	10.6	3.3	3	3	0				
	WM_3	0	0	0	0	0	**100**	3.6	9.3	12.2
	Number of patients	5.7	3.3	3.1	3	0				
	WM_4	0	0	0	0	0	**100**	3.4	9.3	12.3
	Number of patients	5.7	3.4	5.2	0	0				
	BMA-bCRM	0	0	0	0	0	**100**	3.3	9.3	12.3
	Number of patients	6.9	3.2	3.2	3	3.4				
8		(5, 20)	(8, 30)	(9, 40)	(12, 50)	(20, 60)				
	WM_1	0.1	0.9	5.4	0.6	0	**93**	1.6	6.7	20.7
	Number of patients	7.3	5.6	11.7	10.6	7.1				
	WM_2	0.8	4.1	1.5	0	0	**93.6**	1.3	5.2	19.7
	Number of patients	10.6	9.6	7.5	5.2	5.1				
	WM_3	0.2	1.1	1.6	2.2	0.2	**94.7**	1.7	6.8	19.9
	Number of patients	7.6	5.1	6.2	7.2	10.1				
	WM_4	0.1	0.5	6.3	0	0	**93.1**	1.5	6.1	19.8
	Number of patients	7	5.8	14.8	0	0				
	BMA-bCRM	0.3	1.4	4.3	0.8	0.2	**93**	1.6	6.4	20.5
	Number of patients	8	6.5	10	8.6	6.7				

an inverse function method against correlated bivariate normal outcomes. In our simulation studies, the true correlation coefficient between these outcomes is assumed to be 0.5 in the scale of bivariate normal outcomes. The maximum number of patients is set to 45, and we performed 1,000 simulated trials for each scenario. These simulation studies were performed by using SAS version 9.2. The SAS code of the proposed BMA-bCRM approach used in these simulation studies is available on our website (http://www.rs.kagu.tus.ac.jp/hamada/lab.html).

3.2. Simulation Results

In scenario 1 (WM_1 was true and the fourth dose level was the true RD), the selection probabilities at fourth dose level were considerably different among the four working model using bCRM under different set of skeletons. As expected, WM_1 had the highest selection probability, 66.4%. WM_4 had the lowest probability (0%) but selected the third dose level with a probability of 98.3% because the assumed skeleton for efficacy was constant from the third dose level to the fifth dose level. Thus, the assumed skeletons may dominate the operating characteristics and induce such an extreme result. WM_2 also had a poor probability (25.6%) but selected a third dose level with a probability of 63.3%. In contrast, the proposed BMA-bCRM approach had a selection probability of 69.7%, which was considerably similar to the probability under the true working model WM_1. The number of patients treated at each dose level also differed greatly among the four working models. Except for WM_1, the most frequent dose patients were assigned was not the fourth dose level, which was the correct RD. Thus, if we had employed bCRM by WM_2 or WM_4 as a trial design, the incorrect RD was the likely result, which could not achieve the maximum benefit-risk balance.

In scenario 2 (WM_2 was true and the second dose level was the true RD), the correct RD selection probability using the proposed BMA-bCRM approach was approximately 40%, which was second best among the five designs. The worst working model was WM_1, which had a correct RD selection probability of approximately 30%. In scenario 3 (WM_3 was true and the fifth dose level was the true RD), the worst working model was WM_4, which had a correct RD selection probability of 0%. WM_2 was also the second worst working model, which had less than 40% probability. On the other hand, the proposed BMA-bCRM approach had a probability of greater than 70%, which was the second best among the five designs. In scenario 4 (WM_4 was true and the third dose level was the true RD), the proposed BMA-bCRM approach is the third best among the five designs, having a correct selection probability of approximately 65%. However, in case of WM_2 or WM_3, the probability reduced to less than 40%.

In scenarios 5 and 6, the second dose level and the fourth dose level were the true RD, respectively. None of working models was true in these scenarios. Nevertheless, the proposed BMA-CRM approach was robust with a correct RD selection probability close to the best-fitting working model. Scenarios 7 and 8 were the cases in which none of the dose levels were appropriate for RD. The probabilities of selecting the decision "no appropriate dose levels" using different designs were close and had a high probability among the five designs. Also, the average number of total patients did not differ among designs.

According to the results of simulation studies, as expected, the proposed BMA-bCRM approach is robust in terms of misspecifying the true dose efficacy

and/or dose toxicity. In scenarios 1 to 4, in which one of the working models was true, the differences in the correct RD selection probabilities between the most poorly fitting working model and the proposed BMA-bCRM approach were approximately 10% to 70%. Even in scenarios 5 and 6, in which none of the working models were true, the differences were approximately 15% to 45%. The correct RD selection probability using the proposed BMA-bCRM approach was not typically better than that of the best-fitting working model, but the second best among the five designs in most of cases. The average number of patients treated at RD was also almost second best among the five designs, while it was always larger than that of the poorly fitting working model. The average number of patients treated at toxic dose levels, which have true toxicity probability over 30%, was not that much larger than that of other working models. The average number of patients for whom efficacy was observed and the average number of patients for whom toxicity was observed were also similar to that of best-fitting working model.

These robust properties of the proposed BMA-bCRM approach came from appropriate weights, which reflected the fitness of each working model to observed outcomes. Figure 1 showed the transition of the average values of the estimated $\Pr(WM_k|D)$ for the proposed BMA-bCRM approach against the accumulating number of patients in scenarios 1 to 6. In any case, the average weights of well-

Table 3 Sensitivity analysis of the proposed BMA-bCRM approach with different prior distributions of model parameters under scenario 1

Scenario (p_{ej}, p_{tj})	Design	Selection probabilities (%) for RD at the end of trial: Dose level 1	2	3	4	5	None	Average percentage of efficacy	Average percentage of toxicity	Average number of patients
1		(20, 5)(30, 10)(40, 15)(50, 30)(60, 45)								
		$\sigma = 1$, $\psi \sim$Be(2, 2)								
	BMA-bCRM	0	0	16.6	**71.7**	11.7	0	20.5	11.5	45
	Number of patients	3.1	3.1	10.6	22.7	12.9				
		$\sigma = 4$, $\psi \sim$Be(2, 2)								
	BMA-bCRM	0	0	16.3	**69.9**	11.4	2.4	19.8	11.2	44.2
	Number of patients	3.8	3.3	10.3	21.4	13.2				
		$\sigma = 10$, $\psi \sim$Be(2, 2)								
	BMA-bCRM	0	0	17.1	**68.6**	12.3	2	19.9	11.1	44.3
	Number of patients	3.8	3.3	10.4	21.3	13.3				
		$\sigma = 4$, $\psi \sim$Be(0.5, 0.5)								
	BMA-bCRM	0	0	16.8	**70**	11.2	2	19.9	11.2	44.3
	Number of patients	3.8	3.3	10.4	21.4	13				
		$\sigma = 4$, $\psi \sim$Be(10, 10)								
	BMA-bCRM	0	0	16.4	**69.2**	12	2.4	19.8	11.1	44.2
	Number of patients	3.8	3.3	10.3	21.4	13.1				

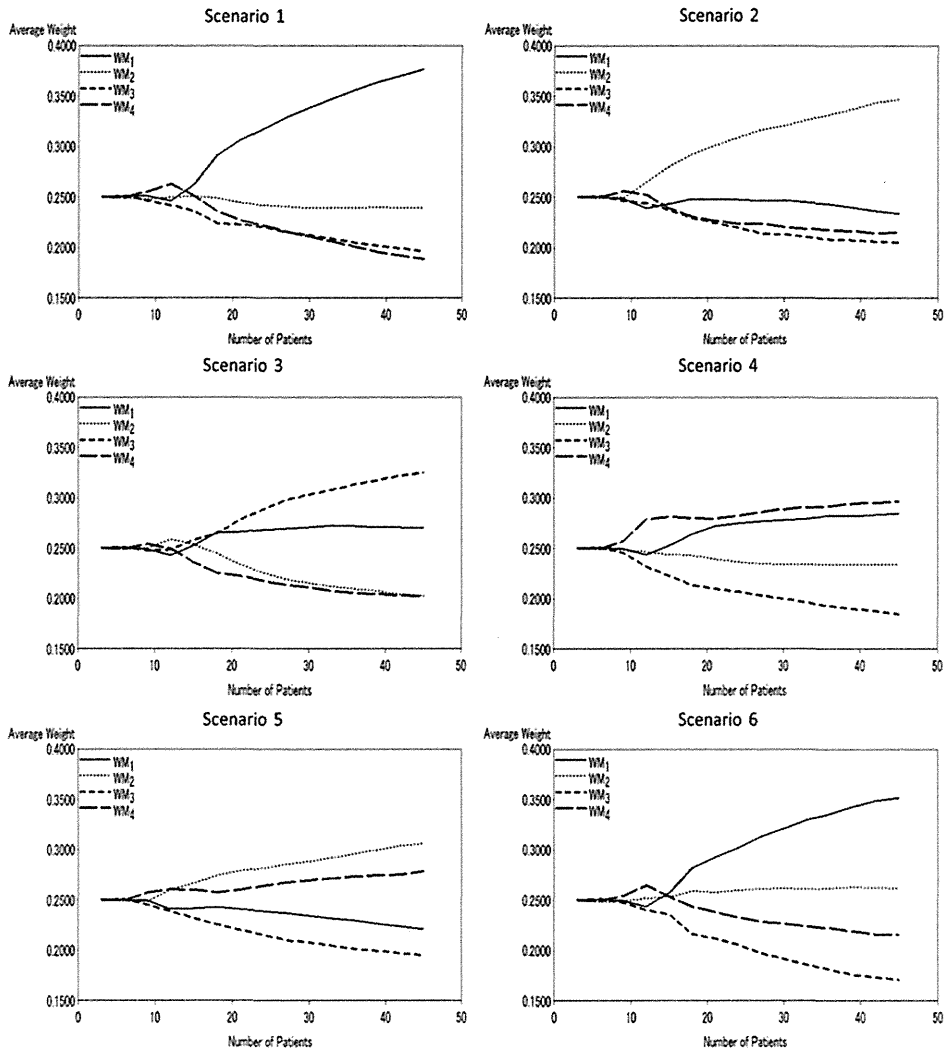

Figure 1 Transition of average weights for each working model.

fitting working models were indeed larger than those of poorly fitting working models in our simulation setting. Although the proposed BMA-bCRM approach showed a uniformly robust property under various scenarios, it was not only affected by the well-fitting working model but also by the other working models because of its strategic nature. In scenario 1, the proposed BMA-bCRM approach showed a larger correct RD selection probability than that of WM_1, which was true (the fourth dose level was the true RD) because of other working models, WM_2, WM_3, and WM_4, assumed RD as second, fifth, and third dose levels, respectively. In this case, the BMA estimate of RD was more likely to shrink the upper dose level. On the contrary, this shrinkage somewhat harms the operating characteristics in scenario 2, in which WM_2 was true. Therefore, there were some

differences in the operating characteristics despite the fact that the average weights for the well-fitting working model and for other working models were almost the same.

4. SENSITIVITY ANALYSES

To further evaluate the robustness of the proposed BMA-bCRM approach, we conducted some sensitivity analyses. At first, we evaluated the impact of the prior distribution for model parameters and association parameter. Instead of $\sigma = 4$, which is the deviation parameter of the employed normal prior for the gradient parameter, we employed either a more informative parameter ($\sigma = 1$) or a more vague one ($\sigma = 10$). To evaluate the impacts of the prior distribution for the association parameter, we used $Be(0.5, 0.5)$ and $Be(10, 10)$ instead of $Be(2, 2)$ under $\sigma = 4$. Furthermore, we evaluated the impact of the correlation coefficient between efficacy and toxicity outcomes, which was assumed to be $\rho = 0.5$ in the scale of bivariate normal outcomes. In this sensitivity analysis, we evaluated the cases in which the true correlation coefficient was assumed to be $\rho = 0.1$ or $\rho = 0.9$. The results of these sensitivity analyses in scenario 1, shown in Table 3 and Table 4, indicated that the prespecified prior distributions for the model parameters or the strength of underlying correlation between efficacy and toxicity outcomes did not have a major impact on operating characteristics.

Next, we evaluated the performance of the proposed BMA-bCRM approach using a different number of working models. Under scenario 6, we increased or decreased the number of working models from two to six. For the two working

Table 4 Sensitivity analysis of the proposed BMA-bCRM approach with different values of assumed strength of correlation between efficacy and toxicity outcomes under scenario 1

		Selection probabilities (%) for RD at the end of trial								
		Dose level								
Scenario (p_{ej}, p_{tj})	Design	1	2	3	4	5	None	Average percentage of efficacy	Average percentage of toxicity	Average number of patients
1		(20, 5)(30, 10)(40, 15)(50, 30)(60, 45)								
		$\rho = 0.1$								
	BMA-bCRM	0	0.1	17.2	**71.7**	7.7	3.3	19.5	10.8	43.9
	Number of patients	3.9	3.3	10.6	22.4	11.5				
		$\rho = 0.5$								
	BMA-bCRM	0	0	16.3	**69.9**	11.4	2.4	19.8	11.2	44.2
	Number of patients	3.8	3.3	10.3	21.4	13.2				
		$\rho = 0.9$								
	BMA-bCRM	0	0	15.2	**70**	14.1	0.7	20.3	11.4	44.8
	Number of patients	3.6	3.3	10	20.9	14				

Table 5 Comparison of operating characteristics of the proposed BMA-bCRM approach using two, three, four, five, and six working models under scenario 6

Scenario (p_{ej}, p_{tj})	Design	Selection probabilities (%) for RD at the end of trial: Dose level 1	2	3	4	5	None	Average percentage of efficacy	Average percentage of toxicity	Average number of patients
6		(5, 5)(10, 10)(20, 15)(30, 25)(35, 45)								
		Two working models								
	BMA-bCRM	0	0	56.8	**15.7**	0	27.5	6.9	5.5	36.2
	Number of patients	4.7	3.4	25.9	14.5	0				
		Three working models								
	BMA-bCRM	0	0	28	**45**	0	27	7.8	6.4	36.3
	Number of patients	4.7	3.3	15.1	21.8	0				
		Four working models								
	BMA-bCRM	0	0	2.4	**45.3**	25.8	26.5	9.2	9.5	36.4
	Number of patients	4.7	3.3	5.3	16.8	18.1				
		Five working models								
	BMA-bCRM	0	0	3.9	**47.1**	29.3	19.7	10	10.4	39
	Number of patients	4.5	3.7	4.9	16.2	18.7				
		Six working models								
	BMA-bCRM	0	0	3.1	**43.2**	30	23.7	9.6	10.4	37.7
	Number of patients	4.5	3.6	4.9	14.4	19.8				

models approach, we employed the WM_1 and WM_2; for the three working models approach, we employed the WM_1, WM_2, and WM_3. For the five working models approach and six working models approach, we add the fifth working model WM_5: $\{(1, 5), (20, 20), (40, 45), (60, 60), (65, 65)\}$ and sixth working model WM_6: $\{(5, 20), (8, 30), (9, 40), (12, 50), (20, 60)\}$ in this order. Table 5 shows the results of these sensitivity analyses. In terms of the correct RD selection probability, there was no substantial difference among the setting except for the two working models setting. These results indicate that as long as a sufficiently reasonable working model is selected, we do not have to prepare many working models; three or four working models may be sufficient.

Furthermore, we investigated the impact of selected working models for the proposed BMA-bCRM approach. We prepared three different sets of working models for the proposed BMA-bCRM approach and compared the operating characteristics of the proposed BMA-bCRM approaches using different sets of working models under scenario 2. From the simulation results summarized in Table 6, there was no critical difference in operating characteristics among the proposed BMA-bCRM approach using different sets of working models, but there was some divergence in terms of the correct RD selection probability, approximately